Scholarly Quality Measurements: A Systematic Literature Review

Rand Alchokr[1][0000-0003-0112-5430], Abhishek Gopalrao[1], Gunter Saake[1][0000-0001-9576-8474], Thomas Leich[2][0000-0001-9580-7728], and Jacob Krüger[3][0000-0002-0283-248X]

[1] Otto-von-Guericke University, Magdeburg, Germany
{rand.alchokr,Abhishek.Gopalrao,saake}@ovgu.de
[2] Harz University & METOP GmbH, Wernigerode, Germany
tleich@hs-harz.de
[3] Eindhoven University of Technology, Eindhoven, The Netherlands
j.kruger@tue.nl

Abstract. The academic publishing landscape is rapidly evolving, making quality assessments and impact evaluations of scientific papers increasingly challenging. Understanding the respective methods is crucial for maintaining the integrity, quality, and relevance of academic publishing in such a changing environment. In this paper, we investigate existing quality-assessment methods for scientific papers, as well as their advantages and disadvantages. For this purpose, we conducted a systematic literature review to capture a comprehensive overview of existing methods, which led to 43 papers and 14 methods. Specifically, we analyze their usage, strengths, and weaknesses, in addition to potential avenues for enhancements. The results can support researchers by providing the knowledge to navigate through quality-assessment methods to make evaluations concerning the reliability and suitability of diverse methods within a specific scientific context.

Keywords: Academic publishing · Publications · Assessment methods · Quality

1 Introduction

The continuous surge in the number of research papers necessitates the implementation of rigorous evaluation procedures to uphold the integrity and credibility of scientific research. Consequently, it becomes imperative to establish and employ robust methods for assessing the quality of scientific papers. These methods play a pivotal role in ensuring that only publications adhering to the highest standards of integrity and reliability are disseminated among the scientific community, which, in turn, contributes to the advancement of knowledge. By employing robust quality-assessment methods, researchers can identify papers that meet rigorous standards and provide valuable insights or contributions to their respective disciplines. The ongoing assessment of paper quality, both pre- and post-publication, plays a vital role in upholding the standards of scientific research and in ensuring the credibility and validity of scholarly knowledge. Praus [84] states that assessing research performance involves the application of objective metrics that gauge productivity and citation influence [73]. These metrics can be classified into two main categories. The first category comprises quantitative measures related to research outputs, such as papers, books, reports, book chapters, patent [77, 83], and various economic and commercial parameters [3]. The second category revolves around the evaluation of citations, either independently [112, 120] or in combination with the volume of published papers [2]. Unfortunately, an overview of existing quality-assessment methods (also going beyond metrics) is missing.

In this paper, we report a systematic literature review (SLR) with which we elicited an overview of existing quality-assessment methods for scientific papers, with a specific focus on their mechanisms, advantages, disadvantages, and applicability. The primary objective of our research is to provide researchers with a thorough understanding of different quality-assessment methods. By meticulously analyzing and synthesizing the findings from numerous studies, this paper aims to furnish researchers with a deeper overview of the mechanisms and techniques employed in determining the quality of scientific papers.

Please note that quality assessment varies by context, and researchers may lack awareness of applicable metrics and methods. Particularly, specializations on a particular field can limit their ability to assess papers beyond that field. Also, evolving assessment criteria can pose challenges, especially for long-time researchers. So, quality assessments can be resource-intensive, with few available tools or guidelines, leaving researchers unsure where to start. Ultimately, our research aims to promote a culture of robust and trustworthy scientific practices, fostering the advancement and credibility of scientific research as a whole. For this purpose, we contribute the following:

- We review 43 papers to provide an overview of quality-assessment methods employed to evaluate the impact and quality of scientific papers.
- We extract and discuss key issues that point out the advantages and disadvantages regarding the quality-assessment methods.
- We define research directions to improve existing quality-assessment methods.

We hope that our contributions can help researchers improve quality-assessment methods for scientific papers and make these fairer.

2 Background and Related Work

With the continuous advancement of technology (e.g., large language models) and the emergence of new research areas, scientists are sharing an ever-growing amount of research each year, resulting in an overwhelming abundance of published works [5]. This increase poses challenges, especially for evaluating the quality of publications and research. Some common challenges are:

- The steadily increasing volume of publications makes it challenging for researchers to keep up with the latest advancements, potentially leading to crucial insights being missed [63].
- The pressure to publish can lead to an increase in low-quality or poorly reviewed papers, including those in predatory journals [41].
- With an ever-increasing number of papers, peer reviewers face difficulties in providing thorough and timely reviews, potentially leading to oversights in the assessment processes [107].
- Publishing in various different venues, including conferences, journals, and pre-print servers, makes it harder to track and assess the latest research developments comprehensively [24].

Quality is a multifaceted concept that varies depending on its context. It typically pertains to the attributes defining excellence or standards of something [57]. Josiam et al. [51] stress the significance of high-quality research for reasons like enhancing participant engagement, reducing biases, and ensuring trustworthiness. They propose eight universal criteria for assessing research quality, encompassing topic relevance, rigor, authenticity, credibility, contribution, ethics, and coherence [110]. Key methods for assessing paper quality include peer reviews, bibliometrics like citation rates, impact factors, circulation, manuscript acceptance rates, and more [15, 16, 101]. Research on knowledge quality has surged, with a consensus that knowledge quality differs from typical information and data quality [81, 85, 117]. However, the literature remains fragmented due to the subjective nature of knowledge [71, 95], contextual factors [30, 62, 72], and challenges in defining knowledge quality [85].

In education, Bridges et al. [23] suggest three potential quality indicators for research publications: publication venue, citation frequency, and number of downloads. However, these are not ideal substitutes for reading-based assessments. Datasets can be assessed for quality using various research data indicators. Konkiel et al. [57] classified these indicators into five categories: quality, citation-based, altmetrics, usage statistics, and reuse indicators. The study of Koya et al. [60] also addresses dataset quality, dividing it into internal view dimensions (design and operational aspects) and external view dimensions (utilization and value). Various researchers have studied perspectives for assessing research quality like peer review, publication count, citations, or research grants [88, 97, 101] and defined research quality through plausibility, originality, and scientific value [8, 82]. The work of Margherita et al. [70] compiles 77 quality dimensions into five categories: research vision, research process, research description, research diffusion, and research impact. In research, "impact" denotes the significance or consequences associated with a particular study or paper within its field, the scientific community, or society. This impact can materialize in various forms, including citation frequency, practical applicability, policy implications, paradigm-shifting insights, or educational contributions [90].

Confusion between a paper's impact and its inherent quality arises from several factors. First, the common use of citation counts as an impact measure may mistakenly link high citations with high quality, even when citations result from critiques or debates. Second, a paper's impact varies across research fields, making universal quality assessments difficult. Lastly, the timing of effects can be ambiguous, as influential papers may take time to gain recognition and citations, potentially influencing perceptions of their quality. Evaluating paper quality can be subjective and impacted by individual biases. Impact metrics provide quantifiable data, but may not capture the full range of quality. Balancing short-term practicality and long-term significance can be challenging [88]. In conclusion, impact and quality are connected but represent separate aspects of research assessments, with both being important for evaluating papers.

3 Methodology

We aimed to understand contemporary methods used for assessing the impact and quality of scientific papers. For this purpose, we employed a systematic literature review following the guidelines for software engineering research proposed by Kitchenham [54]. Proceeding from these guidelines, our methodology for the systematic search involves the steps we explain in this section.

3.1 Research Questions

We defined three research questions (RQs) to guide our work:

RQ_1 *What are current state-of-the-art quality-assessment methods employed to evaluate the impact and quality of scientific papers?*
This question explores temporary methods used to assess the quality and impact of scientific publications. Understanding these methods helps researchers make informed decisions for choosing and evaluating papers, promoting robust research practices.

RQ_2 *What are the pros and cons of the quality-assessment methods?*
This question aims to compare and analyze current quality-assessment methods to uncover their strengths and limitations. Investigating these pros and cons provides valuable insights for researchers and practitioners, enhancing research evaluations. This analysis advances quality-assessment practices in the scientific domain.

RQ_3 *How can the existing quality-assessment methods be improved to better ascertain the quality and impact of scientific papers?*
This question aims to explore enhancements in quality-assessment methods for a more accurate evaluation of scientific papers. By proposing and analyzing potential improvements, we seek to contribute to more effective and reliable methods for evaluating research publications, thereby benefiting researchers and advancing research-assessment practices.

Answering these questions provides an overview for future research to build upon.

3.2 Search Strategy

According to Kitchenham [54], selecting search terms as well as defining search resources are key for a systematic and reliable search strategy. To generate search strings, we defined key terms. We combined these terms with OR operators to derive different search strings. Based on the individual search strings, we built the following search string by connecting each string with an AND operator:

```
"(TITLE((‘quality’ AND (‘indicator*’ OR ‘evaluation*’ OR
‘assessment*’ OR ‘measurement*’ OR ‘metric*’ OR ‘method*’) AND
(‘research’ OR ‘publication*’ OR ‘scholarly’ OR ‘paper*’))) AND
(LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND
(LIMIT-TO (LANGUAGE , "English"))"
```

We deployed our final search string on Scopus[4] as our primary digital library, which offers a wide range of high-quality studies.

3.3 Selection Criteria

We defined the following inclusion criteria (ICs) for identifying relevant papers based on our research questions:

IC_1 The papers underwent a formal review process and are officially published in a journal, workshop, or conference. This requirement ensures that the selected papers have been evaluated and fulfill a certain level of quality.

IC_2 The papers are available in PDF format, while we ignored audio, video, or HTML formats.

IC_3 The focus of our SLR is on papers belonging to any domain. So, we did not narrow our selection to papers conducted within a certain domain.

IC_4 The primary aim of our study is to identify the current state-of-the-art in quality-assessment methods for scientific papers, in both pre- and post-publication. Thus, we aimed to retrieve papers that directly address this objective, exploring the mechanisms and techniques used.

IC_5 Only papers with accessible full-texts are included. Due to limitations in digital libraries, some papers may be inaccessible to us. Consequently, we can focus on papers with fully accessible text only.

Additionally, we defined the following exclusion criteria (ECs):

EC_1 Papers written in languages other than English are excluded. We set English as the standard search language in our search string.

EC_2 Papers that contain only abstracts or similar short summaries are not included. These types of papers offer limited and unsatisfactory information.

EC_3 Thesis papers, scientific transcripts, non-traditional publications, non-commercial publications, and non-academic publications are excluded. These papers may lack systematic or meaningful information.

EC_4 Papers without publisher information or publication type are excluded. The credibility of papers lacking this information is questionable, and thus they are not considered.

Using these criteria, we aimed to ensure the quality of the primary studies we selected for conducting our SLR.

[4] https://www.scopus.com/

3.4 Data Extraction

During an SLR, the step of extracting data is vital to systematically gather data from the selected primary studies and to address the defined research questions. Proper data extraction ensures the reliability of the data synthesis, standardization, quality assessment, gap identification, and transparency. To collect bibliographic information, we use standardized forms with predetermined fields for each primary study: name of the authors, date of publication, publication ID, title of publication, publication details, and publication venue.

To ensure the systematic extraction of information from each primary study with respect to our research questions, we established specific extraction forms that capture key sections. These sections included the study's *summary, objective, proposed method, method description, results and findings, methodology evaluation, as well as limitations of the method.* We designed these forms to facilitate the consistent collection of these essential details from each study.

3.5 Conduct

In Figure 1, we display the conduct of our literature search and the number of papers we ended up with after each step. We executed the search in August 2023, retrieving a total of **2,667** papers. The first and second authors independently screened all publications to identify those that provided direct evidence relevant to our research questions by applying part of the ICs and ECs, such as language, type, and format. This way, the two authors narrowed down the number of papers to **1,806**. Disagreements between the two authors were resolved through discussions until they found a common denominator. After analyzing titles and abstracts in detail, the number of papers decreased to **541** and subsequently **118**. Then, we applied our ICs and ECs by skimming the full texts of the remaining papers, leading to a conclusive count of **81** papers. Afterwards, we thoroughly scrutinized the full text of these papers reaching **40** we deemed relevant. Finally, we performed backwards snowballing on these 40 papers, which led to the identification of **3** new papers (43 total). We performed a cross-validation of the selected papers and extracted the relevant data.

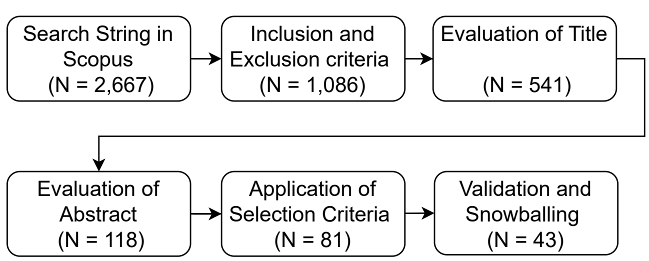

Fig. 1: Phase wise paper collection

Table 1: Summary of the quality-assessment methods and the corresponding primary studies (PS).

ID	Method	PSs	Quality	Impact	First
1	Journal Impact Factor	[6,25,32,42,48–50,55, 86, 102, 103, 107, 111, 118, 121]	√	√	1955
2	Citation Count	[6, 8, 23, 32, 37, 38, 48, 55,57,88,102,103,107, 113, 118]	√	√	1960
3	Peer Review	[8, 23, 34, 48, 49, 76, 88, 89, 107, 113, 121]	√		1970
4	Crown Indicator	[50]	√	√	1995
5	Novel Methodology	[55]	√	√	2000
6	SIGAPS	[111]	√		2002
7	H-index	[8, 50, 84, 102, 107],		√	2005
8	Journal Rankings	[48, 78]	√	√	2007
9	EERQI	[23, 42, 43]	√		2008
10	Altmetrics	[57]	√	√	2010
11	Peer Review: post-publication	[9]	√		2018
12	HCP	[84]		√	2019
13	SRQAM	[37]	√	√	2021
14	RipetaScore	[106]	√		2022

3.6 Results and Discussion

In this section, we present the results of our SLR. We identified a total of 43 papers, which we analyzed according to the criteria we defined for the data extraction.

RQ_1 What are current state-of-the-art quality-assessment methods employed to evaluate the impact and quality of scientific papers?

We started by constructing a table summarizing the diverse quality-assessment methods found in the selected papers. Notably, 30 of these papers discussed one or more assessment methods. The remaining papers either explored quality in different contexts or discussed the pros and cons of specific methods. In Table 1, we present all quality-assessment methods we gathered. Several of them belong to the category of bibliometrics.

Bibliometric indicators employ statistical and mathematical metrics to measure both quality and quantity [45, 50]. Bibliometrics includes employing various indicators like citation counts, journal impact factors, or the h-index. Researchers use these indicators to measure their impact, while organizations employ them for appointments, promotions, funding, and assessing research quality [33, 50].

1. Journal Impact Factor (JIF) was introduced by Eugene Garfield in 1955 to measure the average paper's citations within a specific year [39]. It is calculated as the ratio of year 3 citations to items published in years 1 and 2, relative to the number of substantive papers in those years [39]. It is used for evaluating scientists, research groups, academic promotions, and funding allocations based on the assumption that a journal represents its papers [48, 69, 96]. While JIFs are a widely accepted indicator for journal quality, they may not reliably represent individual paper quality [107].

2. Citation Count reflects the attention a paper receives from peers, effectively serving as a form of peer review [23]. It was first introduced during the development of bibliometrics which was tied to Eugene Garfield's introduction of the Science Citation Index (SCI) in 1961 [7]. Initially, SCI served as a bibliographic database for information retrieval but also enabled quantitative analysis of scientific literature [115]. Databases like SCI, now part of Web of Science, became popular for citation analysis. They record paper references, allowing citation counts for research assessments. In the early 2000s, competing databases like Scopus and Google Scholar also incorporated citation statistics [8]. The simplicity of using citation counts makes it the top choice for many evaluation tasks [113]. In theory, citation counts measure research productivity compared more reliably compared to paper counts, as they reflect a paper's utilization and its contribution to science [113]. Citation data is used for research quality assessments [107] to quantify how often research impacts and possibly influences the work of researchers, akin to an extended form of peer review. Citations indicate peer recognition and are a vital quality indicator; their absence suggest either exceptionally complex or low-quality work [102].

3. Peer Review is a primary criterion for assessing research quality, prioritizing quality over quantity [88]. This method involves expert referees providing feedback, and funding agencies have increasingly detailed assessments that consider factors like originality, alignment with research programs, priority areas, and social relevance. Seven vital criteria in the peer review process have been identified: importance, usefulness, relevance, methodological soundness, ethical integrity, completeness, and accuracy [89]. Peer reviewing improves research quality according to multiple studies [34]. There are more than 35,000 peer-reviewed journals worldwide publishing millions of papers annually, resulting in varying quality and hierarchies among them [114].

4. Crown Indicator can be categorized as a performance indicator, specifically falling under the domain of research-performance indicators [50]. The frequency with which a paper is cited by others is used as a metric for assessing the author's performance. Calculating the average number of citations per year involves dividing the total number of citations by the number of years within a specific time frame. It is alternatively named the field-normalized citation score and was formulated by the Centre for Science and Technology Studies at Leiden University [74]. The Crown indicator normalizes citation rates at a higher level, comparing the average citation rate of a researcher, research group, or department to the average rate within the fields where they have published [50].

5. Novel Methodology utilizes statistical sampling, bootstrapping, and classification techniques; and was introduced to assess the ranking of conferences and journals that were not featured in Information Systems' listings due to limitations in time and resources [55]. It leverages citations to gauge the quality of academic sources and involves a sampling error. To measure this error, bootstrapping or resampling techniques are applied, which provide the respective confidence intervals. These intervals are utilized to group journals and proceedings into quality classes that share similarities, with book publishers being clustered separately [55].

6. Systeme d'interrogation, de gestionet d'analyse des publications scientifiques (SIGAPS) was created at the French University Hospital, Lille (CHU) [29], to evaluate a medical institution's capacity for conducting clinical research. While the quantity of scientific papers produced by an institution is a commonly recognized measure of research output, it is crucial to consider not only the quantity but also the quality of these papers [111]. The paper's quality was assessed by defining it as the ratio of the total score of a journal's impact factor relative to the number of papers [40]. The primary limitation of this metric lies in its inability to compare the impact factor of journals across different scientific domains [111]. SIGAPS incorporates various elements, including the ranking of the journal and the author's position within the paper.

7. H-index was originally introduced by Hirsch in 2005 [47], and swiftly gained popularity as a widely used bibliometric measure [8]. It serves as a straightforward indicator of citations and research productivity [84]. To calculate it, a researcher's papers are ranked by citation count, and the h-value is the point at which the number of citations matches or exceeds the number of papers [47, 50]. The h-index serves as a metric for evaluating a researcher's overall impact. It overcomes limitations of other measures, such as total papers, total citations, citations per paper, and "significant papers." Hirsch argues that when two researchers share a similar h-index, their overall scientific influence is comparable, irrespective of differences in total publications or citations. The h-index remains comparably unaffected by exceptionally rarely or highly cited papers [18, 50].

8. Journal Rankings are employed as proxy measures to evaluate research quality when assessing research outputs [48]. Despite the reliability of published journal quality rankings, they have influenced authors' submission decisions. This has placed lower-ranked journals in a struggle with fewer and lower-quality submissions. Within accounting, it is evident that this has not significantly enhanced the overall quality of papers, but instead has limited the diversity, originality, and practical applicability of research [78].

9. European Educational Research Quality Indicators (EERQI) aim to improve the assessment of educational research quality, since traditional methods based on citation rankings and journal impact factors fall short of adequately covering European social sciences and humanities papers. This oversight affected researchers, institutions, subject domains, and languages within European science [43]. This methods aims to develop practical tools to enhance quality assessment, transparency, and quality of the detection process [43]. In turn, it should simplify and make quality assessments less time-consuming.

10. Altmetrics short for "alternative metrics," started in 2010 to track online interactions with research, shedding light on its broader impact. This data is collected from various online sources and provides real-time insights into how research is disseminated, discussed, and utilized. These metrics complement traditional citation counts, offering a more comprehensive view of research impact [98–100]. Altmetrics are links to research content, reflecting various forms of engagement, including social media interactions, peer reviews, and more [57]. They are typically interpreted as attention or "buzz" [17, 36, 105, 109], reach or readership [31, 58], and quality [20, 79]. Altmetric[5] and PlumX Metrics[6] are widely recognized altmetrics services designed to monitor altmetrics associated with research data.

11. Post-Publication Peer Review, unlike traditional metrics and pre-publication peer reviews, offers more detailed insights and can shed light on research that may not align with the reader's primary focus [9, 104]. Such reviews come in different forms, such as open reviews, comments, recommendations, and discussions—and appear on different platforms like Peeriodicals and FacultyOpinions. On such platforms, communities assess the quality and significance of papers to mitigate the limitations of traditional metrics and pre-publication reviews.

12. High-ranked Citations Percentage (HCP) was introduced to address limitations of the h-index [84]. It identifies highly-cited papers by utilizing the h-index as a reference point and computes the ratio of their citations to the total count. The researchers tested this approach among various scholars to differentiate h-core papers and their citations from the overall dataset. HCP and h-index are independent and complementary, intended for unbiased evaluations regardless of career length or field. However, HCP was negatively linked to excessive self-citations and rarely cited papers. The most successful authors had HCP above 70 % and an h-index above 15. Studies suggest that using HCP alongside the h-index is feasible for fair assessments of research quality based on publications and citations [84].

13. Sustainable Research Quality Assessment Model (SRQAM) was introduced to assess research quality indicators like publication rates and citations within higher education institutions [37]. It helps evaluate research trends and impact in organizational contexts. These assessments inform research policies and fund allocation, leading to impactful outcomes [37]. Traditional research-assessment methods often aim for national and international comparisons. However, the assessment of research quality is closely linked to the evaluation methods across institutions [11, 14, 66]. Although internal evaluation methods within research institutions have been proposed for comparing outcomes at both national and international levels, the literature has yet to document the creation of systematic inter-institutional models for assessing research quality and trends [13, 46].

14. RipetaScore is a new quality-assessment method proposed due to the disadvantages of existing metrics in assessing the quality, transparency, and trustworthiness of research papers [106]. This method advocates three crucial elements in establishing trust: trust in the paper, author, and data. It is a means to systematically gauge a paper's research practices, professionalism, and reproducibility. The trust in reproducibility score assesses a paper's elements

[5] https://www.altmetric.com/
[6] https://plumanalytics.com/

aiding replication. The professionalism score evaluates the author's credibility and transparency regarding external influences on their work [106]. In contrast, the component "Trust in Research" assesses if a paper meets established research standards. This is vital for automated analysis, since some publishers do not differentiate between editorials, communications, and research papers in their metadata, which is why the RipetaScore includes this component [106].

RQ_2 What are the pros and cons of the quality-assessment methods?

Exploring the methods' advantages and shortcomings offers valuable insights for both researchers and practitioners, ultimately improving research evaluation practices in the academic field. We gathered the respective information and summarize every method's pros and cons in Table 2. We list the primary studies underpinning the problems points in the column disadvantages.

RQ_3 How can the existing quality-assessment methods be improved to better ascertain the quality and impact of scientific papers?

Based on our findings, we identified potential enhancements for each method.
1. Journal Impact Factor. Creating field-specific metrics to account for different publication and citation patterns across disciplines and extending the period for calculating impact factors to capture the broader impact, particularly in fields with long publication delays. Promoting inclusiveness and diversity as indicators to mitigate biases against papers by non-native English speakers and underrepresented regions or groups. Considering a combination of multiple metrics, including citations, altmetrics, and peer review assessments for a more comprehensive evaluation of research impact.
2. Citation Count. Working towards limiting the impact of "self-citations," enhancing citation database accuracy, adjusting for field-specific citation patterns, and accounting for the time factor [113]. Combining citation metrics with altmetrics to provide a more comprehensive assessment of research impact and integrate citations with peer review to offer a more holistic method for evaluating research quality.
3. Peer Review. Diversifying the pool of reviewers based on gender, ethnicity, and geographical location to reduce biases and enhance perspective diversity. Training reviewers to improve their skills and understanding of the process, and establishing specialized panels for interdisciplinary research to acknowledge the challenges of such work. Implementing and strengthening ethical guidelines for authors, reviewers, and editors to prevent misconduct and uphold research integrity. Check peer-reviewing related research that investigates authors' concerns of bias [10].
4. Crown indicator. Using average citations is valuable, but may not capture highly-cited papers or recent research impact. Therefore, incorporating diverse metrics, like altmetrics or recognizing groundbreaking work, can enhance the evaluation. Categorizing papers by journal subject categories has limitations and a more precise approach based on content or keywords can provide more accurate field attribution. The method gives more weight to older publications, which may

Table 2: Pros and cons of each quality method.

	Pros	**Cons**
1. Journal Impact Factor	• Quantifies research quality and impact in a standardized manner • Assesses research at all levels, from papers to entire journals, institutions, and nations • Has a historical basis dating back to the 1960s and 1970s, making it a well-established method • Benchmarks and compares journal impact, aiding research assessment • Is globally accepted and used	• Can be manipulated by publishing provocative papers and seeking commentaries that generate citations but aren't citable • Significant variation among scientific fields • Different fields have unique citation patterns, which are not accounted for; without normalization, cross-discipline comparisons are not possible • A high value reflects a journal's impact, but it does not guarantee a paper's quality; it may come from a few highly cited papers [6, 50, 86, 96, 111, 121]
2. Citation Count	• No personal assessment required • Comparison relies on quantitative analysis using raw data • International scale comparison • Quick, easy, and affordable • Has multiple uses, like assessing research importance, impact, and gauging research reuse	• Varies enormously from one discipline to another • Manipulation by "citation clubs" and self-citation • Does not directly indicate a paper's value • Influenced by paper topics and researcher reputations • Mistakes in names or indexing can confuse citation analyses [8, 23, 26, 32, 44, 65, 75, 80, 113, 118]
3. Peer Review	• A dependable way to ensure research quality, vital for academic and scientific credibility • Involves field experts to enhance the credibility, and ensure the best evaluation • Offers objective quality assessment by independent, unbiased reviewers • Emphasizing peer review motivates researchers to produce high-quality, well-researched, and well-presented work	• A primary concern is selecting impartial, expert reviewers • Peer-reviewed journals establish hierarchies, driving researchers to target prestigious ones only • Potential presence of implicit bias within panels • There is a need for clear evaluation criteria in advance • High workload, time demands, and substantial costs [23, 52, 88, 107, 108, 113]
4. Crown indicator	• Research performance measure for assessing individual researchers or groups, quantifying research impact and performance • Uses paper citation frequency, an objective, and widely accepted academic measure • Normalizes citation rates by accounting for variables like research field and publication year, making it a fairer assessment method • Overcomes impact-factor limitations, like paper type and research field variations, for a more comprehensive assessment by considering various factors	• Using Thomson Reuters Subject categories to classify papers can be limited due to the interdisciplinary nature of modern research • Normalizing citation rates compares researchers to their field's average rate but may not address paper-level variations • Can favor older papers, which may not reflect the current impact of research and can put recent work at a disadvantage • Depends on the journal's field categorization, overlooking papers from one field that may appear in journals from other categories [50]
5. Novel Methodology	• It goes beyond traditional indexes providing a comprehensive view of the academic impact • Simplifying certain elements, like not adjusting for journal size and extending the time frame, makes this methodology a faster and cost-effective way compared to traditional methods • Extending to various academic fields beyond Information Systems, including accounting, finance, and marketing • Bootstrapping and confidence intervals provide a statistical measure of ranking accuracy, bolstering assessment robustness	• Has a sampling error that affects accuracy. • Primarily targets Information Systems field, and its applicability to other fields may need validation and adaptation • Data collection and storage for this method can be resource-intensive, making it better for organizations or institutions than individual researchers • Comparing this methodology's rankings with ISI's impact values may not be valid due to their distinct features and limitations [55]
6. SIGAPS	• Objectively assesses research paper quality based on the journal's impact factor • Its transparency, with well-defined formulas, ensures reproducible and understandable for others • Aids evidence-based resource allocation • Using SAS for data aggregation and analysis streamlines and manages large volumes of data efficiently	• Calculating the JIF score may assign identical scores to papers with the highest and lowest IF in the same percentile • Inconsistency in assessing institutions' research track records, with differing priorities in paper quality, quantity, and citation counts • Impact of trendy research fields, which affect the budget allocation and create challenges for countering these shifts [111]
7. H-index	• User-friendly, easy to calculate and understand • Supports both the number of papers produced and the impact of their citations • Assesses research productivity at the individual, department, university, institute levels, and academic journals • Measures the quality as well as the quantity of research productivity	• Influenced by researcher age, career consistency, field specificity, insensitivity to highly-cited papers, and underestimation of those with a few impactful papers • Lacks field-specific adjustments and doesn't account for career length, disadvantaging younger researchers • A researcher's h-index remains high even if they stop publishing after an active 15-20 year career, without late-career publications [8, 19, 22, 27, 50, 64, 84, 94, 107]

Method	**Pros**	**Cons**
8. Journal Rankings	• Standardize research quality assessment that offers a common framework for comparison • Objective benchmarks for researchers, institutions, and policymakers, aiding decision-making • Top journals have more visibility and readership • Funding agencies and institutions use journal rankings to allocate resources to areas with a history of high-quality research • Researchers use journal rankings to choose the right publication outlet for their work	• Researchers may lean towards topics accepted by top journals, discouraging innovative or unconventional research • Seeking high-ranking journal publications can foster fierce competition, emphasizing quantity over quality • Based on limited indicators, might miss contributions that don't align with ranking criteria • Striving for prestigious journal publications may lead to unethical practices like salami slicing and data manipulation • Emphasize metrics like citations and favoring established fields over emerging ones [78]
9. EERQI	• Addresses traditional methods limitations • Creates language-flexible tools to accommodate diverse research publications • Automated semantic analysis speeds up research text review and assessment • Combining intrinsic and extrinsic quality indicators improves the research quality review process • Its Peer Review Questionnaire gained acceptance in education research, showing its practicality and reliability	• Comprises complex tools and techniques that some users might struggle to grasp • Implementing EERQI tools may demand resources for multiple databases, search engines, and linguistic technologies • Aims to reduce biases, but indicator and tool selection can introduce bias based on their design and implementation • Focused on education research but aims to expand to other fields, necessitating further testing and adaptation for broader use [43]
10. Altmetrics	• Record engagement faster than citations, with aggregators tracking dataset mentions within hours, while citations take months to materialize in peer-reviewed literature • Highlight interaction across a diverse audience, from the public to researchers • Evaluate diverse sources, valuable for non-standard areas like datasets and software • Show how research is used in interdisciplinary contexts, offering a comprehensive view • Enhance traditional citation metrics, offering a nuanced impact assessment	• Limited use leading to a lack of understanding and responsible use • Could be easily manipulated • The use of automated "bots" can artificially inflate altmetrics, including those for research data, concealing genuine engagement • lack discipline-specific benchmarks for research data due to their limited repository coverage, unlike the comprehensive Data Citation Index • Lacks a universal standard, making it challenging to assess research impact across diverse fields, unlike traditional citation metrics [4,28,56,57,61,68,87]
11. Post-publ review	• Valuable tools for researchers, aiding in publication discovery and research assessment • Complement traditional metrics • Leads to higher-quality, comprehensive, and constructive reviews	• They are recent with limited adoption in various domains • Adds to researchers' workload beyond their regular responsibilities • Sharing research ideas can be challenging due to research competition • A centralized hub is needed to encourage researchers to contribute [9,21,116]
12. HCP	• Serves as a complementary metric to the h-index, addressing some of its limitations • It values quality with highly cited papers for a nuanced evaluation beyond the quantity • Fairer in assessing a broader range of researchers by not penalizing shorter or interrupted research careers, unlike the h-index • Applicable across diverse fields, unaffected by researchers' specific scientific domains • Combining HCP and the h-index provides a balanced assessment of research quality and quantity	• Calculating and interpreting HCP involves several variables and constants, which can be labor-intensive • Sensitivity to citation rates and papers per year can introduce score variability with rate changes • Assigning individual researchers their constants may introduce subjectivity or uncertainty • Partially addresses selectivity but lacks in-depth insights into a researcher's publication strategy • Relatively new and may not be as widely adopted or recognized as the traditional h-index [84]
13. SQRAM	• Provides a holistic evaluation of research quality in higher education, considering citations, publications, staff, and enrollment • Adaptable to unique circumstances and priorities • Uses a systematic approach to reduce biases and subjectivity, enhancing assessment credibility and objectivity • Uses reliable data sources • Facilitates benchmarking by comparing universities, helping institutions recognize research strengths and areas for improvement	• Relying on external data sources can lead to data availability and accuracy variations, potentially impacting assessment reliability • Choosing criteria and indicators, such as language and author affiliation thresholds, can introduce subjectivity and biases into the assessment • Substantial resources, including databases, software, and data analysis expertise may be needed • Complex approach can be challenging for institutions lacking the needed resources or expertise in research evaluation [11,14,66]
14. RipetaScore	• Evaluates research papers, considering quality, reproducibility, and professionalism, providing a nuanced trustworthiness assessment • Automated tools for evaluating scientific quality enhance objectivity and consistency by reducing subjectivity in assessments • Promotes research transparency by focusing on methodology, data/code availability, and ethics, elevating reporting standards • Automated tools save time and scales trustworthiness evaluation	• It might not cover all research aspects or fit all research types, and its effectiveness depends on the research's nature • Inputs can change with new research, impacting consistency in assessing older and newer work • May encounter complexities and challenges in ongoing updates, potentially affecting reliability • Combining with other methods, may add complexity to result interpretation [106]

not reflect the current impact, so adjusting the weighting to consider recent and older papers equally could enhance the accuracy.

5. Novel Methodology. Considering shorter periods for data analysis can provide insights into more recent research impact and changes in citation patterns. Accounting for publication-medium size is essential to account for variations in impact based on where a paper is published.

6. SIGAPS. Developing metric indicators for assessing the quality of the entire research process, extending beyond publications. This index should encompass other factors, such as protocol deviations and protocol amendments [111]. Additionally, metrics like citation index and altmetrics should be considered, while also accounting for publication year and research field [40].

7. H-index. Various alternatives to the h-index have been proposed, such as the g-index, h(2)-index, contemporary h-index, and Zhang's e-index, but they lack widespread use and empirical support [35, 53, 59, 119]. Therefore, working towards enhancing their usage in the scientific community is essential.

8. Journal Rankings. Incorporating qualitative assessments including peer reviews, editorial reputation, and societal impact into journal rankings. Ensuring inclusiveness by considering journals from various regions and backgrounds to prevent biases and recognize journals that facilitate interdisciplinary research. Adapting ranking methodologies to accommodate evolving research practices and challenges, while also ensuring regular updates.

9. EERQI. The pilot study involved a small number of documents and reviewers. Therefore, future research should include more documents, use automatic semantic analysis, and improve indicator assessment [67, 91, 92]. Additionally, more work should be done to address reliability, consider language differences, and explore the EERQI framework's impact on paper scoring.

10. Altmetrics. Promoting cross-disciplinary engagement by ensuring that the data sources cover a wide range of scientific disciplines. More work towards integration with citation-based metrics and peer review assessments. Researchers and evaluators should have access to information about how altmetrics are derived, which would improve transparency and their understanding.

11. Peer Review: Post-Publication. Carrying out the process in different research fields to figure out the pros and cons of each of them. Performing thorough examinations of the validity of this method and its alignment with other quality indicators. Encouraging a centralized platform to provide these reviews and encourage reviewers to engage in such activity is essential for the success of this method [9].

12. HCP. Defining HCP more broadly by using alternative bibliometric indicators, such as the g-index that reflects "visibility" and "lifetime achievements" more than h-ranked citations [35]. Exploring other variants of the h-index, including the A-index, R-index, m-index, for a comprehensive study on researcher evaluation across different fields [1, 12, 93].

13. SQRAM. Ensuring data quality by validating the accuracy and consistency of data collected from different sources and consider strategies to handle potential inaccuracies and biases. Conducting sensitivity analysis to assess the robustness of the model's findings, which involves testing the impact of variations in model parameters and data sources.

14. RipetaScore. Improving authorship identification and assessment methods will lead to increased accuracy of the RipetaScore in establishing authorship trust. As research reproducibility standards evolve and the paper corpus grows, the RipetaScore will have to adapt to these changes [106].

4 Conclusion

In this paper, we presented an SLR of existing methods that consider assessment techniques of scientific papers. We identified and analyzed 43 papers. Based on these papers, we identified 14 methods. Not surprisingly, each of these methods has its advantages and disadvantages. Within our analysis, we identified several issues and challenges related to each method and enhancements and ideas to improve them. We conclude that the future of assessing research quality involves the construction of interdisciplinary and multifaceted models. There is a need to focus on crafting comprehensive models encompassing a broader array of quality attributes and dimensions, considering the intricate intersections between various research domains. In doing so, we hope to enhance the transparency, rigor, and reproducibility of published research and foster a culture of quality and integrity within the scientific community

To pursue this line of research, we plan to continue investigating quality-assessment methods utilized nowadays by conducting in-depth research to examine the validity and reliability of each method and shed light on hidden gaps concerning these methods. As our primary source for this study selection was Scopus database, a first future step would be to search other sources, such as the Springer and Google Scholar databases. This expansion has the potential to reveal additional tools and guidelines that might not have been considered during the selection of our primary studies.

References

1. Abdi, H., Williams, L.J.: Principal component analysis. Wiley interdisciplinary reviews: computational statistics **2**(4), 433–459 (2010)
2. Abramo, G., Cicero, T., D'Angelo, C.A.: Individual research performance: A proposal for comparing apples to oranges. Journal of informetrics **7**(2), 528–539 (2013)
3. Abramo, G., D'Angelo, C.A.: How do you define and measure research productivity?. Scientometrics **101**(2), 1129–1144 (2014)
4. Adie, E.: Gaming altmetrics. Altmetric blog (2013)
5. Agarwal, S., Agrawal, A.: Assessment of the impact of publication year and keywords in research manuscript quality using fuzzy inference system. In: Proceedings of 6th International Conference on Recent Trends in Computing: ICRTC 2020. pp. 75–82. Springer (2021)
6. Ahlgren, P., Waltman, L.: The correlation between citation-based and expert-based assessments of publication channels: Snip and sjr vs. norwegian quality assessments. Journal of informetrics **8**(4), 985–996 (2014)

7. Aksnes, D.: Citations and their use as indicators in science policy : studies of validity and applicability issues with a particular focus on highly cited papers. Creativity and innovation management (2005)
8. Aksnes, D.W., Langfeldt, L., Wouters, P.: Citations, citation indicators, and research quality: An overview of basic concepts and theories. Sage Open **9**(1) (2019)
9. Alchokr, R., Krüger, J., Saake, G., Leich, T.: A comparative analysis of article recommendation platforms. In: 2021 ACM/IEEE joint conference on digital libraries (JCDL). pp. 1–10. IEEE (2021)
10. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: Peer-Reviewing and submission dynamics around top software-engineering venues: a juniors' perspective. In: International Conference on Evaluation and Assessment in Software Engineering (EASE). ACM (2022)
11. Allik, J.: Factors affecting bibliometric indicators of scientific quality. Trames **17**(3), 199–214 (2013)
12. Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., Herrera, F.: H-index: A review focused in its variants, computation and standardization for different scientific fields. Journal of informetrics **3**(4), 273–289 (2009)
13. Bhatti, U.A., Huang, M., Wu, D., Zhang, Y., Mehmood, A., Han, H.: Recommendation system using feature extraction and pattern recognition in clinical care systems. Enterprise information systems **13**(3), 329–351 (2019)
14. Bhatti, U.A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yu, Z., Yuan, L.: Time series analysis and forecasting of air pollution particulate matter (pm 2.5): an sarima and factor analysis approach. IEEE Access **9**, 41019–41031 (2021)
15. Birken, C.S., Parkin, P.C.: In which journals will pediatricians find the best evidence for clinical practice?. Pediatrics **103**(5), 941–947 (1999)
16. Bloom, B.S., Retbi, A., Dahan, S., Jonsson, E.: Evaluation of randomized controlled trials on complementary and alternative medicine. International journal of technology assessment in health care **16**(1), 13–21 (2000)
17. Bornmann, L.: Do altmetrics point to the broader impact of research? an overview of benefits and disadvantages of altmetrics. Journal of informetrics **8**(4), 895–903 (2014)
18. Bornmann, L., Daniel, H.D.: Does the h-index for ranking of scientists really work?. Scientometrics **65**, 391–392 (2005)
19. Bornmann, L., Daniel, H.D.: The state of h index research: Is the h index the ideal way to measure research performance?. EMBO reports **10**(1), 2–6 (2009)
20. Bornmann, L., Haunschild, R.: Do altmetrics correlate with the quality of papers? a large-scale empirical study based on f1000prime data. PloS one **13**(5) (2018)
21. Bornmann, L., Marx, W.: Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts?. Journal of informetrics **9**(2), 408–418 (2015)
22. Bornmann, L., Mutz, R., Daniel, H.D.: Are there better indices for evaluation purposes than the h index? a comparison of nine different variants of the h index using data from biomedicine. Journal of the american society for information science and technology **59**(5), 830–837 (2008)
23. Bridges, D.: Research quality assessment in education: impossible science, possible art?. British educational research Journal **35**(4), 497–517 (2009)
24. Cabanac, G., Oikonomidi, T., Boutron, I.: Day-to-day discovery of preprint–publication links. Scientometrics **126**(6), 5285–5304 (2021)

25. Coelho, P., Antunes, C., Costa, H., Kroon, E., Lima, S., Linardi, P.: The use and misuse of the" impact factor" as a parameter for evaluation of scientific publication quality: a proposal to rationalize its application. Brazilian journal of medical and biological research **36**, 1605–1612 (2003)
26. Corbyn, Z.: An easy way to boost a paper's citations. Nature pp. 1476–4687 (2010)
27. Da Luz, M.P., Marques-Portella, C., Mendlowicz, M., Gleiser, S., Silva Freire Coutinho, E., Figueira, I.: Institutional h-index: the performance of a new metric in the evaluation of brazilian psychiatric post-graduation programs. Scientometrics **77**, 361–368 (2008)
28. DeSanto, D., Nichols, A.: Scholarly metrics baseline: a survey of faculty knowledge, use, and opinion about scholarly metrics. College & research libraries **78**(2) (2017)
29. Devos, P., Dufresne, E., Renard, J.M., Beuscart, R.: Sigaps a prototype of bibliographic tool for medical research evaluation. In: The new navigators: from professionals to patients, pp. 721–726. IOS Press (2003)
30. Dey, A.K.: Understanding and using context. Personal and ubiquitous computing **5**, 4–7 (2001)
31. Dinsmore, A., Allen, L., Dolby, K.: Alternative perspectives on impact: the potential of alms and altmetrics to inform funders about research impact. PLoS biology **12**(11) (2014)
32. Dougherty, M.R., Horne, Z.: Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. Royal society open science **9**(8) (2022)
33. Durieux, V., Gevenois, P.A.: Bibliometric indicators: quality measurements of scientific publication. Radiology **255**(2), 342–351 (2010)
34. Ebadifar, A., Eftekhari, M., Owlia, P., Habibi, E., Ghalenoee, E., Bagheri, M., Falahat, K., Eltemasi, M., Sobhani, Z., Akhondzadeh, S.: How to assess quality of research in iran, from input to impact? introduction of peer-based research evaluation model in iran. Archives of Iranian medicine **20**(11) (2017)
35. Egghe, L.: Theory and practise of the g-index. Scientometrics **69**(1), 131–152 (2006)
36. Erdt, M., Nagarajan, A., Sin, S.C.J., Theng, Y.L.: Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. Scientometrics **109**, 1117–1166 (2016)
37. Fahim, A., Tan, Q., Sahabuddin, M.: Sustainable research quality assessment model (srqam) for higher education using regression. In: 2021 IEEE International Conference on Educational Technology (ICET). pp. 268–272. IEEE (2021)
38. Franceschini, F., Maisano, D., Mastrogiacomo, L.: Research quality evaluation: comparing citation counts considering bibliometric database errors. Quality & Quantity **49**, 155–165 (2015)
39. Garfield, E.: Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. Science **178**(4060), 471–479 (1972)
40. Garfield, E.: Citation indexes for science. a new dimension in documentation through association of ideas. International journal of epidemiology **35**(5), 1123–1127 (2006)
41. Gasparyan, A.Y., Nurmashev, B., Voronov, A.A., Gerasimov, A.N., Koroleva, A.M., Kitas, G.D.: The pressure to publish more and the scope of predatory publishing activities. Journal of Korean medical science **31**(12), 1874–1878 (2016)
42. Glujovsky, D., Riestra, B., Coscia, A., Boggino, C., Comande, D., Ciapponi, A.: Assessment of research quality in major infertility journals. Fertility and sterility **98**(6), 1539–1543 (2012)

43. Gogolin, I.: European educational research quality indicators (eerqi): An experiment. Research assessment in the humanities: Towards criteria and procedures pp. 103–111 (2016)
44. Goldstein, H., Maier, G.: The use and valuation of journals in planning scholarship: Peer assessment versus impact factors. Journal of planning education and research **30**(1), 66–75 (2010)
45. Haeffner-Cavaillon, N., Graillot-Gak, C.: The use of bibliometric indicators to help peer-review assessment. Archivum immunologiae et therapiae experimentalis **57**, 33–38 (2009)
46. Hassanain, M., Anil, S., Abdo, A.: Institutional research evaluation model (irem): A framework for measuring organizational research trends and impact and its application in medical academia in saudi arabia. Journal of epidemiology and global health **6**(4), 249–256 (2016)
47. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences **102**(46), 16569–16572 (2005)
48. Jarwal, S.D., Brion, A.M., King, M.L.: Measuring research quality using the journal impact factor, citations and 'ranked journals': Blunt instruments or inspired metrics?. Journal of Higher Education Policy and Management **31**(4), 289–300 (2009)
49. Jones, M.R.H.: Can research quality be measured quantitatively? on quality of scholarship, numerical research indicators and academic publishing-experiences from norway. Fennia-International journal of geography **195**(2), 164–174 (2017)
50. Joshi, M.A.: Bibliometric indicators for evaluating the quality of scientific publications. The journal of contemporary dental practice **15**(2), 258 (2014)
51. Josiam, M., Lee, W., Johnson, T., Pee, C., Hall, J.: Beyond selecting a methodology: Discussing research quality, ethical, and equity considerations in qualitative engineering education research. In: 2022 ASEE annual conference & exposition (2022)
52. Jost, J.T., Rudman, L.A., Blair, I.V., Carney, D.R., Dasgupta, N., Glaser, J., Hardin, C.D.: The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. Research in organizational behavior **29**, 39–69 (2009)
53. Khan, N.R., Thompson, C.J., Taylor, D.R., Gabrick, K.S., Choudhri, A.F., Boop, F.R., Klimo Jr, P.: Part ii: Should the h-index be modified? an analysis of the m-quotient, contemporary h-index, authorship value, and impact factor. World neurosurgery **80**(6), 766–774 (2013)
54. Kitchenham, B.A., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE-2007-01, Keele University and University of Durham (2007)
55. Kleijnen, J.P., Van Groenendaal, W.: Measuring the quality of publications: new methodology and case study. Information Processing & Management **36**(4), 551–570 (2000)
56. Konkiel, S.: Altmetrics: diversifying the understanding of influential scholarship. Palgrave communications **2**(1), 1–7 (2016)
57. Konkiel, S.: Assessing the impact and quality of research data using altmetrics and other indicators. Scholarly assessment reports **2**(1) (2020)
58. Konkiel, S., Scherer, D.: New opportunities for repositories in the age of altmetrics. Bulletin of the american society for information science and technology **39**(4), 22–26 (2013)

59. Kosmulski, M., et al.: A new hirsch-type index saves time and works equally well as the original h-index. ISSI newsletter **2**(3), 4–6 (2006)
60. Koya, K., Chowdhury, G.: A quality and popularity based ranking method for research datasets. In: Proceedings of the 2022 4th Asia Pacific information technology conference. pp. 103–110 (2022)
61. Kratz, J.E., Strasser, C.: Making data count. Scientific data **2**(1), 1–5 (2015)
62. Kwan, M.M., Balasubramanian, P.: Knowledgescope: managing knowledge in context. Decision support systems **35**(4), 467–486 (2003)
63. Larsen, P., Von Ins, M.: The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics **84**(3), 575–603 (2010)
64. Lazaridis, T.: Ranking university departments using the mean h-index. Scientometrics **82**(2), 211–216 (2010)
65. Leydesdorff, L., Wouters, P., Bornmann, L.: Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators—a state-of-the-art report. Scientometrics **109**, 2129–2150 (2016)
66. Li, Z., Ho, Y.S.: Use of citation per publication as an indicator to evaluate contingent valuation research. Scientometrics **75**, 97–110 (2008)
67. Locoro, A., David, J., Euzenat, J.: Context-based matching: design of a flexible framework and experiment. Journal on data semantics **3**, 25–46 (2014)
68. Lowenberg, D., Chodacki, J., Fenner, M., Kemp, J., Jones, M.: Open data metrics: Lighting the fire. Zenodo (2019)
69. Lowy, C.: Impact factor limits funding. The Lancet **350**(9083) (1997)
70. Margherita, A., Elia, G., Petti, C.: What is quality in research? building a framework of design, process and impact attributes and evaluation perspectives. Sustainability **14**(5), 30–34 (2022)
71. Melkas, H., Harmaakorpi, V.: Data, information and knowledge in regional innovation networks: Quality considerations and brokerage functions. European journal of innovation management **11**(1), 103–124 (2008)
72. Memmi, D.: The social context of knowledge pp. 189–208 (2008)
73. Mingers, J., Leydesdorff, L.: A review of theory and practice in scientometrics. European journal of operational research **246**(1), 1–19 (2015)
74. Moed, H., De Bruin, R., Van Leeuwen, T.: New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. Scientometrics **33**(3), 381–422 (1995)
75. Moed, H.F.: The impact-factors debate: the isi's uses and limits. Nature **415**(6873), 731–732 (2002)
76. Moldoveanu, B., Cuciureanu, G.: Publishing as an indicator of scientific research quality and ethics: the case of law journals from moldova. Science and engineering ethics **26**(2), 1039–1052 (2020)
77. Nagpaul, P., Roy, S.: Constructing a multi-objective measure of research performance. Scientometrics **56**(3), 383–402 (2003)
78. Northcott, D., Linacre, S.: Producing spaces for academic discourse: The impact of research assessment exercises and journal quality rankings. Australian accounting review **20**(1), 38–54 (2010)
79. Nuzzolese, A.G., Ciancarini, P., Gangemi, A., Peroni, S., Poggi, F., Presutti, V.: Do altmetrics work for assessing research quality?. Scientometrics **118**(2), 539–562 (2019)
80. Petersen, A.M., Fortunato, S., Pan, R.K., Kaski, K., Penner, O., Rungi, A., Riccaboni, M., Stanley, H.E., Pammolli, F.: Reputation and impact in academic

careers. Proceedings of the national academy of sciences **111**(43), 15316–15321 (2014)
81. Pierce, E., Kahn, B., Melkas, H.: A comparison of quality issues for data, information, and knowledge. In: 17th Annual information resources management association international conference (2006)
82. Polanyi, M.: The republic of science: Its political and economic theory (1962). Knowing and being; essays (1969)
83. Praus, P.: Statistical evaluation of research performance of young university scholars: a case study. Transinformação **30**, 167–177 (2018)
84. Praus, P.: High-ranked citations percentage as an indicator of publications quality. Scientometrics **120**, 319–329 (2019)
85. Rao, L., Osei-Bryson, K.M.: Towards defining dimensions of knowledge systems quality. Expert systems with applications **33**(2), 368–378 (2007)
86. Rezaei, G.N., Azizi, F.: The impact factor-based quality assessment of biomedical research institutes in iran: Effect of impact factor normalization by subject (2007)
87. Roemer, R.C., Borchardt, R.: Issues, controversies, and opportunities for altmetrics. Library technology reports **51**(5), 20–30 (2015)
88. Rudd, E.: The evaluation of the quality of research. Studies in higher education **13**(1), 45–57 (1988)
89. Sabetzadeh, F., Tsui, E., Lee, W.B.: Assessment of uncertainty in the quality of knowledge in the research publication review process. In: 2013 10th international conference on fuzzy systems and knowledge discovery (FSKD). pp. 946–950. IEEE (2013)
90. Saha, S., Saint, S., Christakis, D.A.: Impact factor: a valid measure of journal quality? Journal of the medical library association **91**(1), 42 (2003)
91. Sándor, Á., Vorndran, A.: Enhancing relevance ranking of the eerqi search engine. In: Assessing Quality in European Educational Research: Indicators and Approaches, pp. 56–59. Springer (2014)
92. Sándor, Á., Vorndran, A.: Highlighting salient sentences for reading assistance. In: Assessing quality in European educational research: indicators and approaches, pp. 43–55. Springer (2014)
93. Schreiber, M., Malesios, C., Psarakis, S.: Exploratory factor analysis for the hirsch index, 17 h-type variants, and some traditional bibliometric indicators. Journal of informetrics **6**(3), 347–358 (2012)
94. Schubert, A., Glänzel, W.: A systematic analysis of hirsch-type indices for journals. Journal of informetrics **1**(3), 179–184 (2007)
95. Seawright, K.W., Young, S.T.: A quality definition continuum. Interfaces **26**(3), 107–113 (1996)
96. Seglen, P.O.: Why the impact factor of journals should not be used for evaluating research. Bmj **314**(7079), 497 (1997)
97. Shakeel, Y., Alchokr, R., Krüger, J., Leich, T., Saake, G.: Are altmetrics useful for assessing scientific impact? a survey. In: International conference on management of digital ecoSystems (MEDES). ACM (2022)
98. Shakeel, Y., Alchokr, R., Krüger, J., Leich, T., Saake, G.: Incorporating altmetrics to support selection and assessment of publications during literature analyses. In: International conference on vvaluation and assessment in software engineering (EASE)). p. 180–189. ACM (2022)
99. Shakeel, Y., Alchokr, R., Krüger, J., Leich, T., Saake, G.: Altmetrics and citation counts: An empirical analysis of the computer science domain. In: Joint Conference on Digital Libraries (JCDL). pp. 1–11. ACM (2022)

100. Shakeel, Y., Alchokr, R., Krüger, J., Saake, G., Leich, T.: Are altmetrics proxies or complements to citations for assessing impact in computer science? In: Joint conference on digital libraries (JCDL). pp. 284–286. ACM/IEEE (2021)
101. Shakeel, Y., Krüger, J., Saake, G., Leich, T.: Indicating studies' quality based on open data in digital libraries. In: International conference on business information systems. pp. 579–590. BIS, Springer (2018)
102. Sharma, O.P.: Quality indicators of scientific research (2012)
103. Sombatsompop, N., Kositchaiyong, A., Markpin, T., Inrit, S.: Scientific evaluations of citation quality of international research articles in the sci database: Thailand case study. Scientometrics **66**(3), 521–535 (2006)
104. Spezi, V., Wakeling, S., Pinfield, S., Fry, J., Creaser, C., Willett, P.: "let the community decide"? the vision and reality of soundness-only peer review in open-access mega-journals. Journal of documentation **74**(1), 137–161 (2018)
105. Sugimoto, C.: Attention is not impact" and other challenges for altmetrics. Discover the future of research: Wiley exchanges (2015)
106. Sumner, J.Q., Vitale, C.H., McIntosh, L.D.: Ripetascore: measuring the quality, transparency, and trustworthiness of a scientific work. Frontiers in research metrics and analytics **6**, 751–734 (2022)
107. Taylor, J.: The assessment of research quality in uk universities: Peer review or metrics? British journal of management **22**(2), 202–217 (2011)
108. Tetlock, P.E., Mitchell, G.: Implicit bias and accountability systems: What must organizations do to prevent discrimination? Research in organizational behavior **29**, 3–38 (2009)
109. Thelwall, M., Haustein, S., Larivière, V., Sugimoto, C.R.: Do altmetrics work? twitter and ten other social web services. PLoS one **8**(5) (2013)
110. Tracy, S.J.: Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. Qualitative inquiry **16**(10), 837–851 (2010)
111. Ueda, R., Nishizaki, Y., Homma, Y., Sanada, S., Otsuka, T., Yasuno, S., Matsuyama, K., Yanagisawa, N., Nagao, M., Fujibayashi, K., et al.: Importance of quality assessment in clinical research in japan. Frontiers in Pharmacology **10**, 1228 (2019)
112. Vieira, E.S., Gomes, J.A.: Citations to scientific articles: Its distribution and dependence on the article features. Journal of informetrics **4**(1), 1–13 (2010)
113. Wallmark, J.T., Sedig, K.G.: Quality of research measured by citation method and by peer review, a comparison. IEEE transactions on engineering management (4), 218–222 (1986)
114. Ware, M., Mabe, M.: The stm report: An overview of scientific and scholarly journal publishing (2015)
115. Welljams-Dorof, A.: Quantitative citation data as indicators in science evaluations: A primer on their appropriate use. Evaluating science and scientists: An East-West dialogue on research evaluation in postcommunist Europe pp. 202–211 (1997)
116. Wouters, P., Costas, R.: Users, narcissism and control: tracking the impact of scholarly publications in the 21st century (2012)
117. Yoo, D.K., Vonderembse, M.A., Ragu-Nathan, T.: Knowledge quality: antecedents and consequence in project teams. Journal of knowledge management **15**(2), 329–343 (2011)
118. Zanon, B.: Research quality assessment and planning journals. Italian journal of planning practice **2**(2), 96–123 (2014)
119. Zhang, C.T.: The e-index, complementing the h-index for excess citations. PLoS One **4**(5) (2009)

120. Zhou, J., Zeng, A., Fan, Y., Di, Z.: The representative works of scientists. Scientometrics **117**(3), 1721–1732 (2018)
121. Zwahlen, M., Junker, C., Egger, M.: Commentary ii-the journal impact factor in the evaluation of research quality: villain, scapegoat or innocent bystander? Sozial-und Präventivmedizin **49**(1), 19–22 (2004)