



Are papers published in predatory journals worthless? A geopolitical dimension revealed by content-based analysis of citations

Zehra Taşkın^{1,2}, Franciszek Krawczyk², and Emanuel Kulczycki²

¹Department of Information Management, Hacettepe University, Ankara, Turkey

²Scholarly Communication Research Group, Adam Mickiewicz University, Poznań, Poland

an open access  journal



Citation: Taşkın, Z., Krawczyk, F., & Kulczycki, E. (2023). Are papers published in predatory journals worthless? A geopolitical dimension revealed by content-based analysis of citations. *Quantitative Science Studies*, 4(1), 44–67. https://doi.org/10.1162/qss_a_00242

DOI:
https://doi.org/10.1162/qss_a_00242

Peer Review:
https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00242

Received: 28 June 2022
Accepted: 19 January 2023

Corresponding Author:
Emanuel Kulczycki
emek@amu.edu.pl

Handling Editor:
Ludo Waltman

Copyright: © 2023 Zehra Taşkın, Franciszek Krawczyk, and Emanuel Kulczycki. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: content analysis of citations, content-based citation analysis, impact factor journals, predatory journals, questionable journals

ABSTRACT

This study uses content-based citation analysis to move beyond the simplified classification of predatory journals. We present that, when we analyze papers not only in terms of the quantity of their citations but also the content of these citations, we are able to show the various roles played by papers published in journals accused of being predatory. To accomplish this, we analyzed the content of 9,995 citances (i.e., citation sentences) from 6,706 papers indexed in the Web of Science Core Collection, which cites papers published in so-called “predatory” (or questionable) journals. The analysis revealed that the vast majority of such citances are neutral (97.3%), and negative citations of articles published in the analyzed journals are almost completely nonexistent (0.8%). Moreover, the analysis revealed that the most frequently mentioned countries in the citances are India, Pakistan, and Iran, with mentions of Western countries being rare. This highlights a geopolitical bias and shows the usefulness of looking at such journals as mislocated centers of scholarly communication. The analyzed journals provide regional data prevalent for mainstream scholarly discussions, and the idea of predatory publishing hides geopolitical inequalities in global scholarly publishing. Our findings also contribute to the further development of content-based citation analysis.

1. INTRODUCTION

The term *predatory journals* hides complex geopolitical inequalities, various motivations for scholarly publishing, and the local contexts in which these journals proliferate (Krawczyk & Kulczycki, 2021b). Similarly, the practice of citation-counting hides the role played by a given citation in developing the argument of a paper and the motivation for citing. In this study, we argue that the hidden phenomena are strongly related and that revealing this relation might deepen the understanding of transformations currently taking place in academia affecting scholarly communication. We prefer using the term *questionable journals* instead of *predatory journals*, as we argued in our previous study (Kulczycki, Hołowiecki et al., 2021), because the former term does not imply a predatory intent of the publisher.

Previous studies that counted the number of citations referring to articles in questionable journals (Frandsen, 2017; Moussa, 2021) have been unable to show the more complex nature of the phenomenon described as predatory publishing due to limitations of the method—that

is, citation-counting. This study goes beyond this limitation and aims to examine the content of citations referring to questionable journals in journals that are widely accepted as legitimate (i.e., indexed in Web of Science Core Collection [WoS]).

The main research question is twofold. First, as a follow-up study to the previous study that examined the number of articles in questionable journals that are cited in WoS-indexed journals (Kulczycki et al., 2021), we investigate the context of citations of questionable journals in legitimate journals. Second, we address the question of whether the content of citations (i.e., citation sentences) is specific to peripheral or semiperipheral countries (i.e., refers to local affairs). In terms of knowledge production, we understand there is a strongly one-sided influence of knowledge produced in the center compared to knowledge production in peripheries. Moreover, we reflect on what it could mean that questionable journals take on the role of *mislocated centers of scholarly communication*, which is the term we coined to describe and criticize the role of some publication channels in peripheral or semiperipheral countries without condemning scholars who publish in them or accusing publishers of bad intentions (Krawczyk & Kulczycki, 2021b).

1.1. Questioning the Concept of Predatory Journals

Over the past decade, predatory publishing has been one of the most discussed topics not only in the science of science but also among policymakers. Since Jeffrey Beall (2012) created the first list of so-called predatory journals in 2012, many papers have warned against such publication channels, as well as against predatory conferences or fake metrics (Krawczyk & Kulczycki, 2021a). The term *predatory journals* coined by Beall refers to journals that dishonestly use the open-access model and deceive scholars in favor of their own financial interests. Beall (2018) also argued that, because of predatory journals, pseudoscientific articles can leak into mainstream scholarly literature. Grudniewicz, Moher et al.'s (2019) recent definition does not link predatory publishing to the open access concept nor does it focus on the review process, as the authors consider it difficult to assess. They highlighted that such journals prioritize their self-interest at the expense of scholarship and are characterized by false or misleading information, poor editorial practices, and a lack of transparency. However, with definitions focusing on the journals, the quality of the *articles* in these predatory journals is not often considered. When citations referring to predatory journals are considered, a primary suggested solution has been that researchers simply stop citing such journals altogether (Oermann, Nicoll et al., 2020).

The term *predatory journals* is a simple label for complex and multidimensional practices in scholarly communication. The debate over predatory publishing focuses almost entirely on journals published in English in non-English-speaking countries (Eykens, Guns et al., 2019; Grudniewicz et al., 2019; Moussa, 2021). Various lists of predatory journals, such as the discontinued Beall's List or the more complex and transparent Cabell's Predatory Reports, are perceived as useful tools for indicating undesirable journals; however, they provide a relatively simplistic point of view: "Good" journals are published mostly in central countries in English while "bad" journals are published mostly in (semi)peripheral countries in English. Such a dichotomy is not valid: There are many bad journals with aggressive business models in central countries and many good journals published in English and, primarily, local languages in (semi)peripheral countries.

Moreover, many editorially reputable journals from large commercial publishers possess business models that could be accused of being predatory or questionable (Siler, 2020). The term *predatory journal* evokes many negative connotations; however, researchers from

peripheral or semiperipheral countries often publish in such journals because they are counted in research evaluation regimes in semiperipheral countries (Rochmyaningsih, 2019; Teixeira da Silva, Moradzadeh et al., 2022). Previous studies have revealed that peripheral or semiperipheral countries (sometimes called *developing countries*), such as India, Iran, and Turkey, are more profoundly affected by predatory practices than central countries (especially the United States and Western Europe; Demir, 2018; Eve & Priego, 2017; Kulczycki, Hołowiecki et al., 2022). As described in our previous paper (Krawczyk & Kulczycki, 2021b), if a journal starts to be viewed as prestigious in semiperipheral countries (e.g., when it is indexed in Scopus) while it is still seen as questionable in central countries, it becomes a mislocated center of scholarly communication.

In our daily work as researchers and policy advisors, we observe that many scholars and policymakers assume that all articles published in questionable journals could not be published elsewhere and thus are of low quality. With this in mind, in this study, we aim to address the question of whether one can transfer the assessment of a journal (i.e., as a questionable channel of communication) to the evaluation of a single article published in the journal. The results show that going beyond citation-counting allows us to reveal the more complex phenomena behind the simplified notion of a predatory journal and undermines possible assumptions regarding the predation of articles.

1.2. Beyond Simply Counting: Content-Based Citation Analysis

Researchers are expected by their institutions and policymakers at various national and global levels to publish in journals with high impact factors and receive a large number of citations of these publications. Although it has been reported in many studies that this research evaluation practice is problematic (Hicks, Wouters et al., 2015; “Read the Declaration,” n.d.; Wilsdon, Allen et al., 2015), throughout most of the world, academic success is still determined according to these criteria. However, a system based solely on citation quantity faces various significant challenges. For example, policymakers attempt to draw a clear line between “good” and “bad” journals by using journal impact factors as an indicator of journal quality. This creates a scholarly environment in which articles in high-impact-factor journals are considered legitimate or of good quality, while articles in questionable journals are deemed worthless. Interestingly, our previous study revealed that questionable journals are often cited by legitimate ones (Kulczycki et al., 2021). This questions the meaning of citation quantity and reveals new methods for differentiating citations in terms of their content.

Our argument regarding questionable publishing is similar to the more nuanced approaches to predatory publishing, such as the campaign “Think. Check. Submit,” which does not as substantially rely on lists of good and bad journals (www.thinkchecksubmit.org). Additionally, Cabell’s Predatory Reports attempts to evaluate journals based on several categories, such as whether they provide misleading information, send spam, or have a website that seems too focused on collecting publishing fees (Siler, 2020). In this regard, content-based citation analysis can help provide new understanding about these journals as well as the papers published in them, which is important as valuable papers can sometimes be published in journals with questionable publishing practices, causing them to be overlooked.

Content-based citation analysis is not a new approach to citation analysis. Most citation analysis studies, assuming that not all citations are equal, have started with the questions “Why do authors cite?” or “What are the motivations of authors to cite?” (Bonzi & Snyder, 1991; Brooks, 1986; Cano, 1989; Chubin & Moitra, 1975; Cronin, 1981; Garfield, 1970). Various classification schemes have been developed to date, and citations have been classified

according to these schemes using natural language processing tools and machine learning techniques (Iqbal, Hassan et al., 2021). Content-based citation analysis methods manually carried out on small samples in older studies have gained momentum today with the diversification of computerized processing methods and the increase in access to full scientific texts, as predicted by Teufel (1999).

The first results of content-based analysis in practice have already started to be reported. A deep learning tool called *scite* was launched for classifying citation contexts (Nicholson, Mordaunt et al., 2021). *Scite* obtains documents, mines citation contexts, matches references, and classifies citations in terms of their meanings (supporting, contrasting, and mentioning citations). In addition to *scite*, WoS added a new service to its citation indexes called Enriched Cited References (Clarivate, 2021), which provides information regarding the location of citations in the text in terms of the Introduction, Methodology, Results, and Discussion (IMRaD) structure and their purpose (support, differ, basis, background, and discuss). Two of the services use citation extraction, the mining of full texts, and automatic classification. These developments show that the future of citation analysis has started to be reshaped by content-based citation analysis systems. Our study contributes not only to the content-based citation analysis literature by providing a new corpus but also defines some of its present challenges and proposes solutions.

2. MATERIALS AND METHODS

2.1. Data

In this paper, we investigate citances in WoS-indexed articles referring to questionable journal articles to understand the contexts of the citations. A *citance* is a neology created by Nakov, Schwartz, and Hearst (2004) to define the sentence(s) surrounding the citation within a document. To achieve the aim of this study, all cited and citing articles and their metadata were downloaded as PDFs and stored in a MySQL database. Inaccessible articles ($n = 44$) were removed from the data set. A description of the data set is shown in Figure 1.

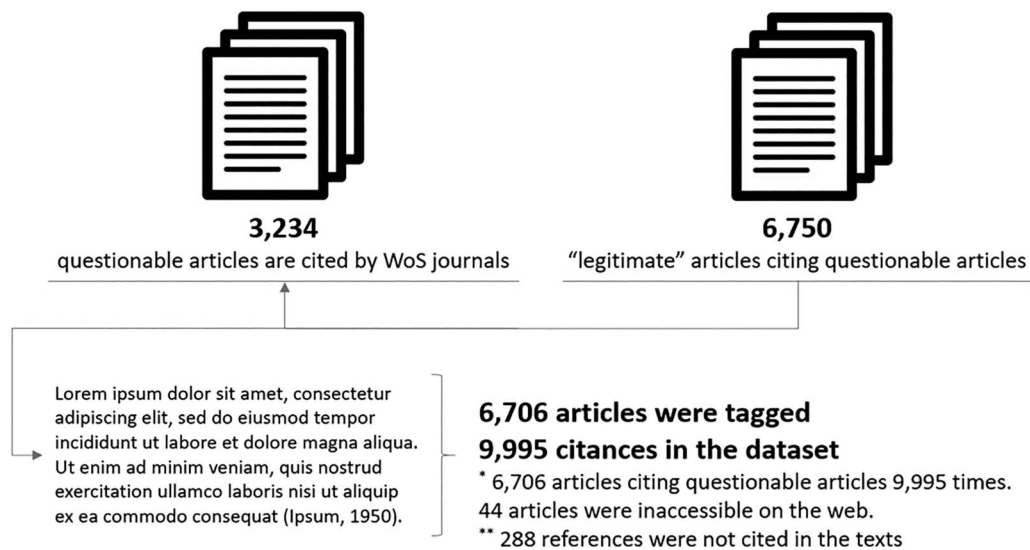


Figure 1. Descriptive statistics of the data set.

2.2. Classification Scheme

To understand and classify the content of citations, we conducted content-based analysis. For the effective content-based analysis of citations, both supervised and unsupervised methods have been suggested in the literature (Athar, 2011; Taşkın & Al, 2018). In this study, to classify citations in terms of their content, we chose expert tagging. The citation classification scheme developed by Taşkın and Al (2018) was used in the expert tagging process (see Figure 2).

In the tagging process, all citations are classified according to four main categories: meaning, purpose, shape, and array. The meaning class defines the authors' interpretation of the work they have cited (positive, negative, both, or neutral). In the purpose class, the classification is made by considering the author's objective for the citation, such as providing literature examples, giving a definition, using a methodology, or validating research results or data. Citations are sometimes accompanied by the author's name or direct quotes from their work; additionally, at times, there may be many works cited within one sentence. These factors are assessed by the shape class. Finally, the array class is used to understand in which sections, how many times, and in how many different sections each study is cited.

2.3. Collection of Citances and Tagging Process

For the expert tagging procedure, a database was created with a custom tagging interface written in the PHP and JavaScript programming languages (see Figure 3). To provide an accurate tagging process, 20 citances were tagged by all the authors (referred to in the document as *taggers*) before they began the tagging process; then, the results were discussed. We called this process *calibration*. The main aim of calibration was to develop a common understanding among all taggers. Then, all citances were tagged by the authors.

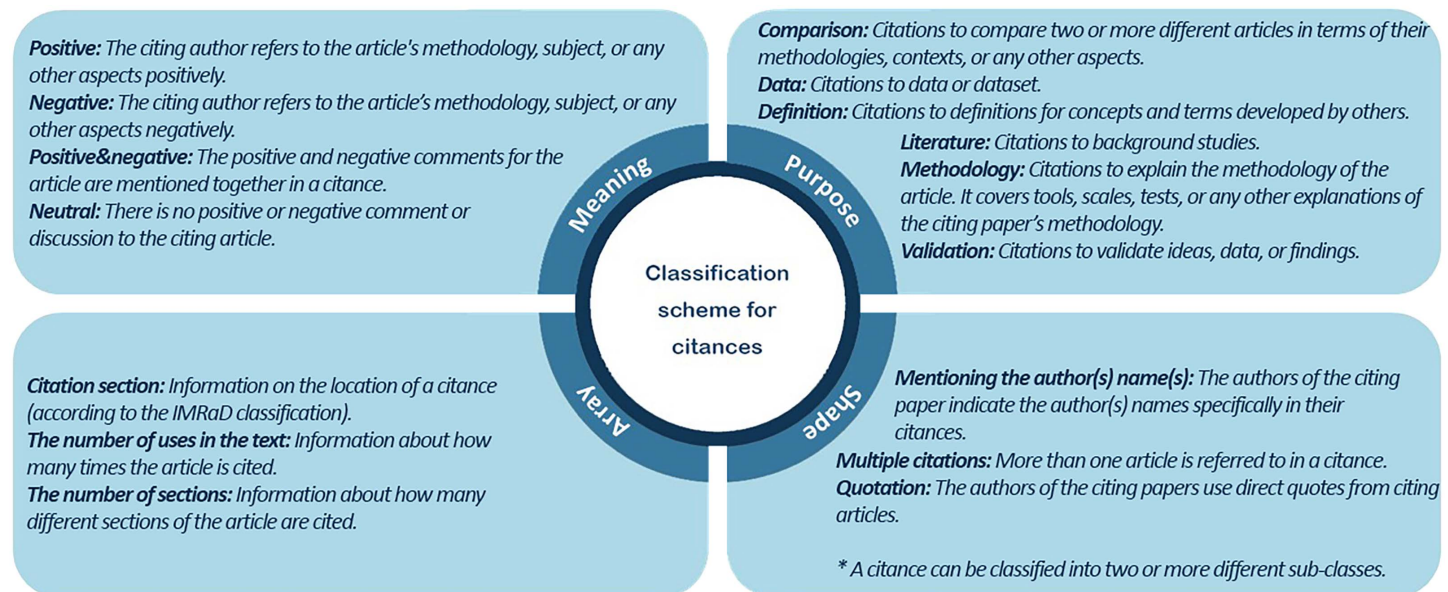


Figure 2. Four classes of the classification scheme for citances. An example of each citation class is presented in Appendix A.

(a) Page 1: Tagging interface

Citation Tagging

Copy ID

Citing article ID: Not Cited

Nothing selected

Citation sentence:

Citation section: Nothing selected Shape: Nothing selected Meaning: Nothing selected Purpose: Nothing selected

Language:

Save

The Last:
 Citing article ID: C55_88_1
 Citation sentence: Understanding of the environmental issues makes them sensitive towards the nature-related problems and lead them to act in an environmentally friendly manner (Pickett-Baker & Ozaki, 2008; Zuraidah et al., 2012) and show sustainable consumption behaviour.

(b) Page 2: Editing/exporting interface

Cite Sentences

[Export Citation List](#)

Search Reset

2096 records displaying

ArticleID	Citation sentence	Meaning	Purpose	Cit.section	Shape	Lang	User	Date	Update Delete
C55_88_1	Understanding of the environmental issues makes them sensitive towards the nature-related problems and lead them to act in an environmentally friendly manner (Pickett-Baker & Ozaki, 2008; Zuraidah et al., 2012) and show sustainable consumption behaviour.	neutral	literature	introduction	multi-cite		zehra	2021-03-28 16:49:42	
C55_87_2	Many researchers have been extending the technology acceptance model (TAM) by incorporating additional factors that can facilitate its generalisability (Ahmad et al., 2010; Al-Ruz and Khasawneh, 2011; Alenezi et al., 2010; Chitungo and Munongo, 2013; Hanafizadeh et al., 2014; Lee et al., 2011; Rasimah et al., 2011; Shin, 2012; Shittu et al., 2013; Stone and Baker-Eveleth, 2013; Sun and Zhang, 2006; Terzis et al., 2013; Tezci, 2011; Thompson et al., 2006; Wong et al., 2012; Zejno and Islam, 2012). However, the literature indicated that very few studies have been conducted by adding SAT into the original TAM (Davis et al., 1989) or diffusion of innovation theory (Devaraj et al., 2002; Kumar and Ravindran, 2012; Lee and Lehto, 2013; Lee and Park, 2008; Rogers, 1995; Shipps and Philline, 2013).	neutral	literature	introduction	multi-cite		zehra	2021-03-28 16:48:55	

Figure 3. Tagging interface.

In the tagging interface, a password-protected account was identified for each tagger. The process for tagging one citation followed the steps listed below:

1. The tagger opens the PDFs of the citing and cited papers.
2. The tagger finds citations using author names or titles and copies the citations (which can be one sentence or more). Taggers must check the whole paragraph and decide which parts refer to the citing paper.
 - a. If the reference is not mentioned in the text, the tagger selects the “Not Cited” (yellow) option. When this option is selected, the other dropdown menus are deactivated.
 - b. If the citation is not written in English or other languages in which the taggers are fluent, the taggers use Google Translate. Then, they tag the citation using the translated version.
3. The tagger chooses the citation classes.
4. If the language of a questionable paper is not English, the tagger writes language information about the paper.
5. To make desired changes for the tagged citations, the tagger can use the editing screen, which is shown in Figure 3(b).

After the tagging process, all citations were classified by a tagger regarding the four main citation classes. However, the need for validation of citations arose for the Meaning class, which is based on the interpretation of the taggers and is relatively more subjective. To meet this need, all positive, negative, positive/negative citations, and 273 randomly selected neutral citations tagged by a tagger were retagged by all three taggers. The interannotator agreement scores are presented in detail in Section 5.

2.4. Visualizations, Analysis, and Statistical Tests

A chi-square test was used to compare the groups, and Cramér’s V was used for the effect size of the test. IBM SPSS version 26 was used to conduct the statistical tests and analyses. To calculate interannotator agreement for the positive, negative, and neutral citations, Tinsley and Weiss’ (2000) percentage calculation was used. The calculation is shown in Eq. 1:

$$\rho = \frac{N_a}{N_a + N_d} \quad (1)$$

where N_a = the number of agreements and N_d = the number of disagreements.

We conducted a content analysis of the citations by counting the word frequency. After identifying country names in the citations, we used VOSviewer to analyze the occurrences of keywords. All keywords were unified and standardized before the analyses were performed. The full counting method was chosen for the visualization and content analysis of the citations. In Figure 7, the co-occurrences of keywords of 1,474 citations with country names that appeared at least 10 times are shown. Country self-citations, which are provided in Table 2, indicate the number of citations covering country names made by authors affiliated with an institution from the same country. Only corresponding authors are considered.

The findings of the study are presented in two sections. First, we describe the general characteristics of the analyzed citations by reporting the number of citations per text, the sections in which the citations were made, and the purpose and shape of the citations. In the second section, we describe the content analysis of the citations, which revealed non-Western

countries were the most frequently mentioned countries. Via the analysis of the word co-occurrences, we describe the contexts in which these countries were mentioned.

3. FINDINGS

3.1. General Characteristics of Citances

3.1.1. Descriptive statistics

We examined the full texts of 3,221 questionable articles and their 6,706 WoS-indexed citers. We tagged the citances by conducting 10,283 transactions in the tagging process. However, 288 citations of legitimate articles (2.8%) were not referred to or mentioned in the article bodies despite being listed in the reference sections. Seventy-eight per cent of the missing citations were published in articles indexed in the Emerging Sources Citation Index (ESCI), 22% in Journal Citation Reports (JCR), and only one journal was indexed in the Arts & Humanities Citation Index (AHCI). After removing the noncited references, we had 9,995 citances in the initial data set. The descriptive statistics for the cited and citing papers are shown in Table 1.

Table 1 shows that 67% of the legitimate articles cited questionable articles one time in the text, and 90% of the legitimate articles cited questionable articles one to three times. These

Table 1. Distribution of the number of citances referring to questionable articles in legitimate articles

Number of uses in the text	N of citing articles	Cumulative %	Cumulative total
1	4,509	67.2	4,509
2	1,144	84.3	2,288
3	384	90.0	1,152
4	169	92.5	676
5	91	93.9	455
6	62	94.8	372
7	23	95.1	161
8	8	95.2	64
9	8	95.4	72
10	8	95.5	80
11	3	95.5	33
12	3	95.6	36
13	2	95.6	26
14	2	95.6	28
15	1	95.6	15
28	1	95.7	28
Not cited in the text	288	100.0	
Total	6,706		9,995

statistics indicate the need to understand the author motivations behind citations, a task for which the content-based analysis of citances can help.

3.1.2. Sections and purposes of citations

According to the results, 66.9% of the citances were found in the introduction section, followed by the discussion (17.1%) and findings (9%) sections (see Figure 4). The distribution of the citances in the IMRaD categories differed from one of our previous studies. Taşkın and Al (2018) found 85% of citances in the introduction section in the Turkish library and information science literature. However, studies in the literature have suggested that citations in the methodology, findings, or discussion sections are more important than citations in the introduction section (Maričić, Spaventi et al., 1998; Voos & Dagaev, 1976). Therefore, it is important to investigate citances in different sections.

When the purpose of the citances in each section was investigated in this study, we found that 90% of citances in the introduction section were literature citations. However, the distribution of the classes in the other sections was quite different compared to that in the introduction section. For example, unsurprisingly, almost 70% of citances in the methodology section were intended to explain the methods of the study. Overall, the purpose class of citances differed according to the IMRaD sections ($\chi^2(10) = 8227.559, p < 0.001, V = 0.454$).

It should be noted that more than 20% of citances in the findings section and almost 50% of citances in the discussion section had the objective of comparing or validating the articles

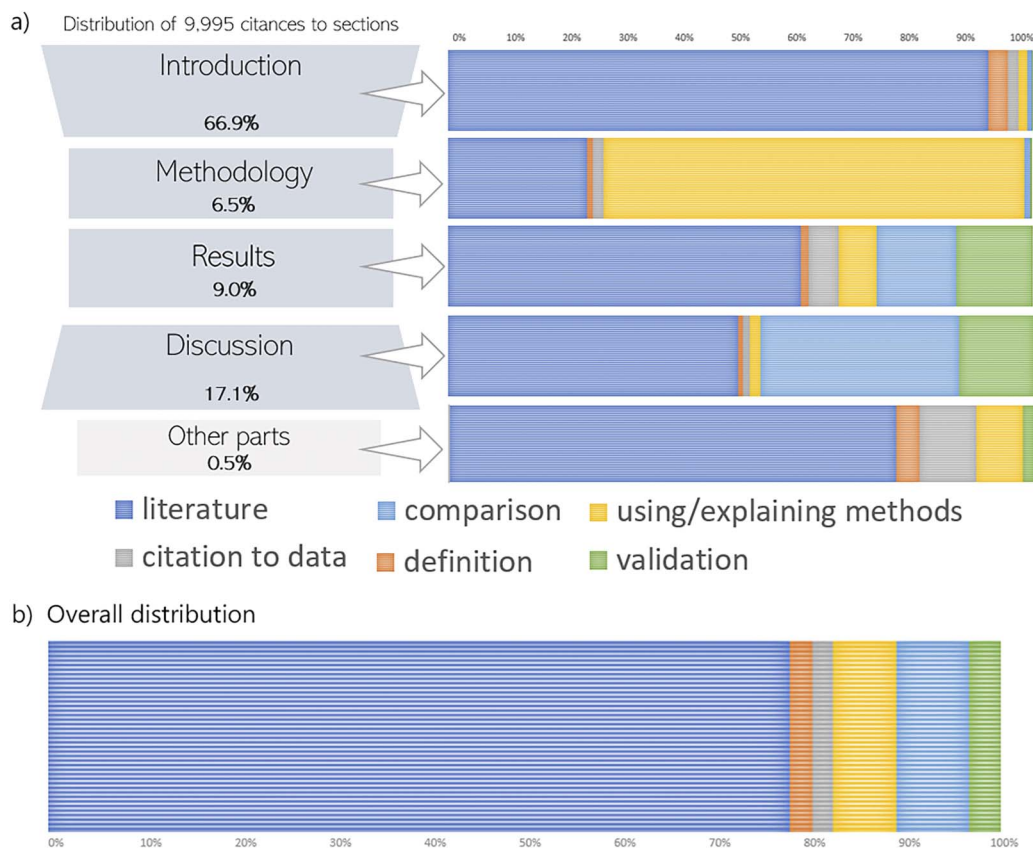


Figure 4. Distribution of citances in the purpose and array classes.

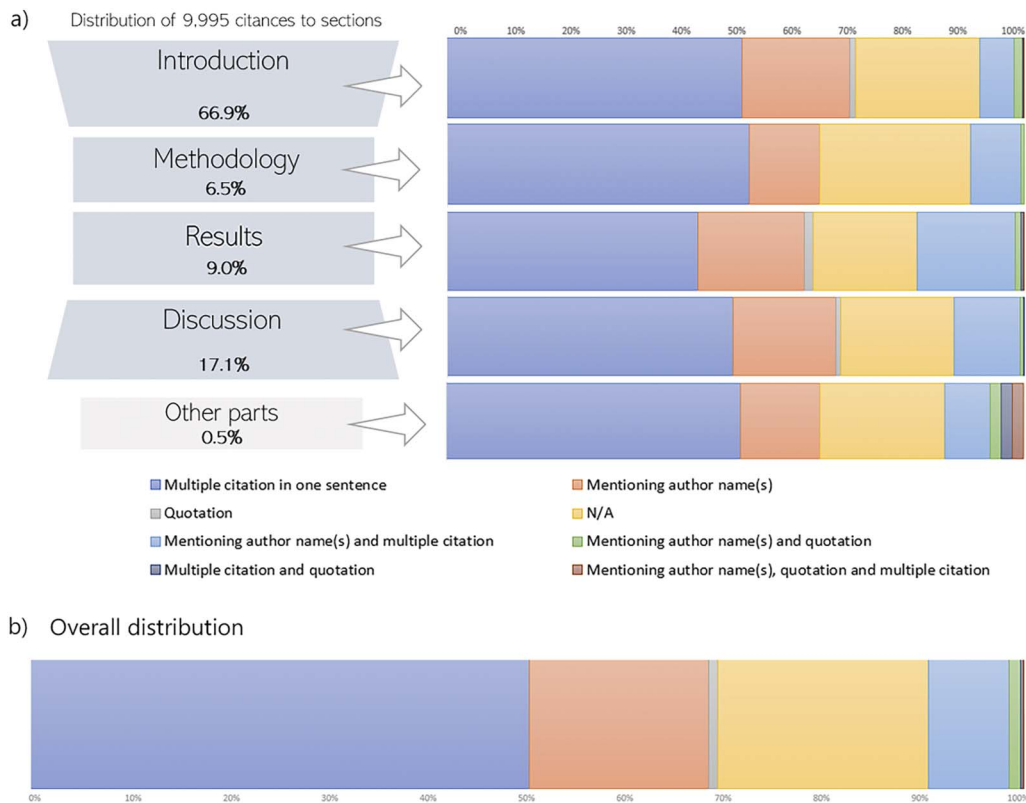


Figure 5. Distribution of citances in the shape and array classes.

(i.e., with similar studies in the literature). This finding can open up a new discussion for future studies regarding the citing behaviors of authors. For instance, it could be investigated whether the authors cite articles that support/validate their hypotheses without considering the publication venue.

3.1.3. Shapes of citations

There are many ways to cite others' publications. Some researchers have indicated that the most valuable citation types are those mentioning authors' names and those with quotations (Bonzi, 1982; Zhu, Turney et al., 2015). However, with the massive increase in the number of publications in all scientific fields, researchers have started to cite papers without reading them (Simkin & Roychowdhury, 2015), with multiple citations in a citance potentially signaling this phenomenon. From this point of view, multiple citations have less importance in scientific writing. This study produced interesting results about the shapes of the analyzed citances (see Figure 5). Although the chi-square test results pointed to a difference between the IMRaD sections and citation shapes ($\chi^2(12) = 36.644, p < 0.001, V = 0.037$)¹, this difference was not as significant as that found for the citation purposes.

Unlike the literature studies reporting that citances mentioning author names comprise the most common citation shape (Bonzi, 1982; Taşkın & Al, 2018), in the present study, almost half of the citances in the data set were multiple citations. Moreover, quotations were

¹ Mentions of the author name(s), multiple citations, quotations, and N/A classes were included for the statistical tests. When all classes were included, the test results were $\chi^2(28) = 230.336, p < 0.001, V = 0.076$.

extremely rare. As previously mentioned, the high rate of multiple citations (e.g., citations such as “there are many studies in the literature on this subject” and many cited articles) could indicate citations made without reading the corresponding article. They could also be coercive citations requested by editors or reviewers. As indicated by Yu, Yu, and Wang (2014), abnormal citing behaviors are common for coercive citation practices. Therefore, future investigations on the citing behaviors of authors who cite multiple sources could be useful.

3.2. Content Analysis of Citances

Table 2 presents the top 20 countries mentioned in the analyzed citances. The “N of occurrences” column shows the total number of citances including country names. However, although some citances include examples from various countries, some are unique for a specific country. Therefore, we added a column to Table 2 to show single mentions (SMs) of countries. For example, while 156 citances mentioned India, 56 of these included other

Table 2. The most frequently mentioned countries in the analyzed citances

Country	N of occurrences	N of occurrences in articles for which the corresponding author was from a given country	N of single mentions (SM)	N of occurrences in articles for which the corresponding author was from a given country (SM)
India	156	113 (72.4%)	100	89 (89.0%)
Malaysia	129	85 (65.9%)	81	71 (87.7%)
Iran	120	69 (57.5%)	92	61 (66.3%)
Saudi Arabia	119	84 (70.6%)	88	77 (87.5%)
Pakistan	117	65 (55.6%)	80	57 (71.3%)
United States	111	44 (39.6%)	62	33 (53.2%)
Nigeria	106	83 (78.3%)	88	74 (84.1%)
China	88	25 (28.4%)	42	23 (54.8%)
Turkey	88	57 (64.8%)	49	44 (89.8%)
Ghana	83	51 (61.4%)	58	48 (82.8%)
Thailand	54	30 (55.6%)	29	27 (93.1%)
Bangladesh	54	32 (59.3%)	31	30 (96.8%)
Indonesia	48	22 (45.8%)	19	15 (78.9%)
Ethiopia	47	33 (70.2%)	28	24 (85.7%)
Taiwan	45	16 (35.6%)	20	12 (60.0%)
South Africa	41	14 (34.1%)	15	8 (53.3%)
Jordan	40	14 (35.0%)	31	13 (41.9%)
Egypt	36	10 (27.8%)	15	8 (53.3%)
Vietnam	36	14 (38.9%)	25	13 (52.0%)
Australia	34	8 (23.5%)	15	8 (53.3%)

country names as well; thus, only 100 of the citations were solely regarding India. Table 2 also shows the country self-citation rates for each occurrence (all and SMs).

The economic positions of the countries (Table 2) vary, and the only countries on the list that can be classified as Western are the United States and Australia. The other Western countries with the most occurrences are United Kingdom (31) and Italy (30). Central countries, such as Germany, Canada, and Spain, have fewer than 15 occurrences each. This highlights a geo-political bias, because, even with the shifting advantage of research in Western countries in terms of funding or the number of publications, there is still significant Western cultural hegemony in science (Marginson, 2021).

Moreover, for some countries, such as India and Thailand, the vast majority of their referent citations came from authors affiliated with these countries. This was not the case for China or Egypt, which were more frequently mentioned by authors from outside these countries. This observation highlights the heterogeneity of the positions of non-Western countries in Western-centered academic publishing.

Another important finding was related to cross-country comparisons or examples illustrated in the citations. As shown in Table 2, some citations included information about more than two countries. We created a co-occurrence network using the country names mentioned in the citations, the results of which are shown in Figure 6. As demonstrated in the figure, the regional

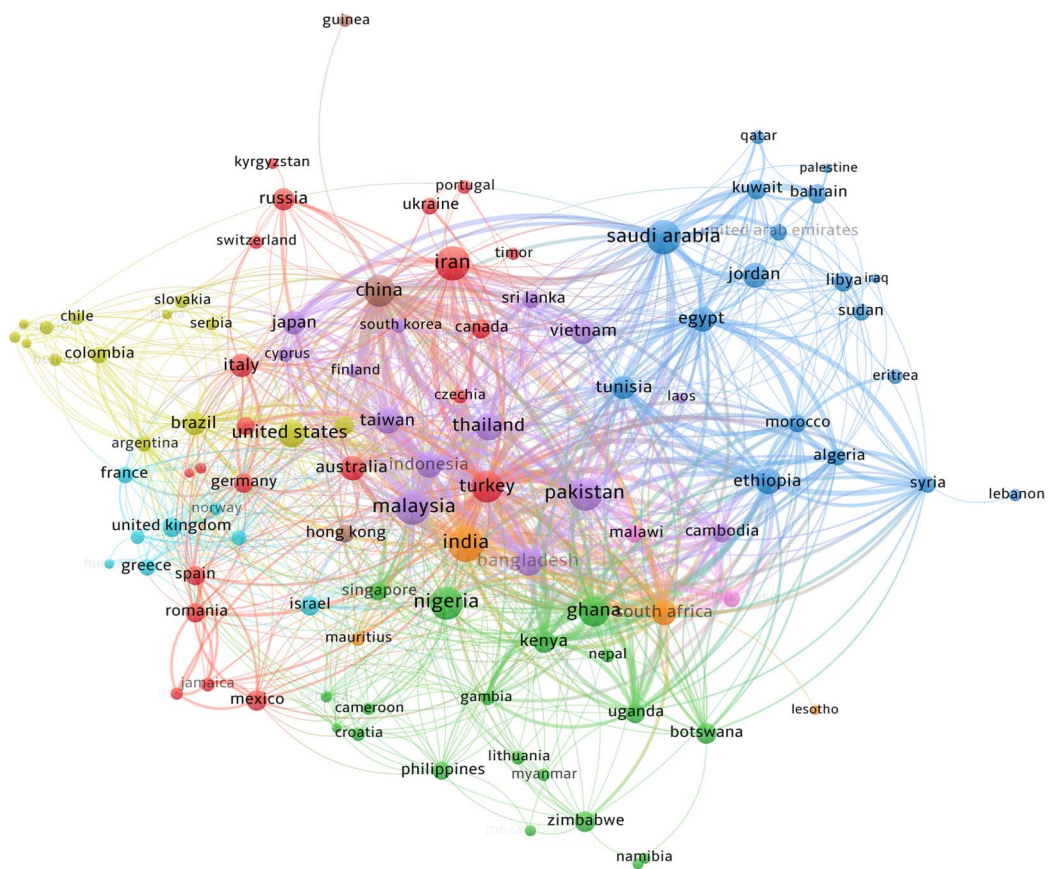


Figure 6. Co-occurrence of countries in the citations.

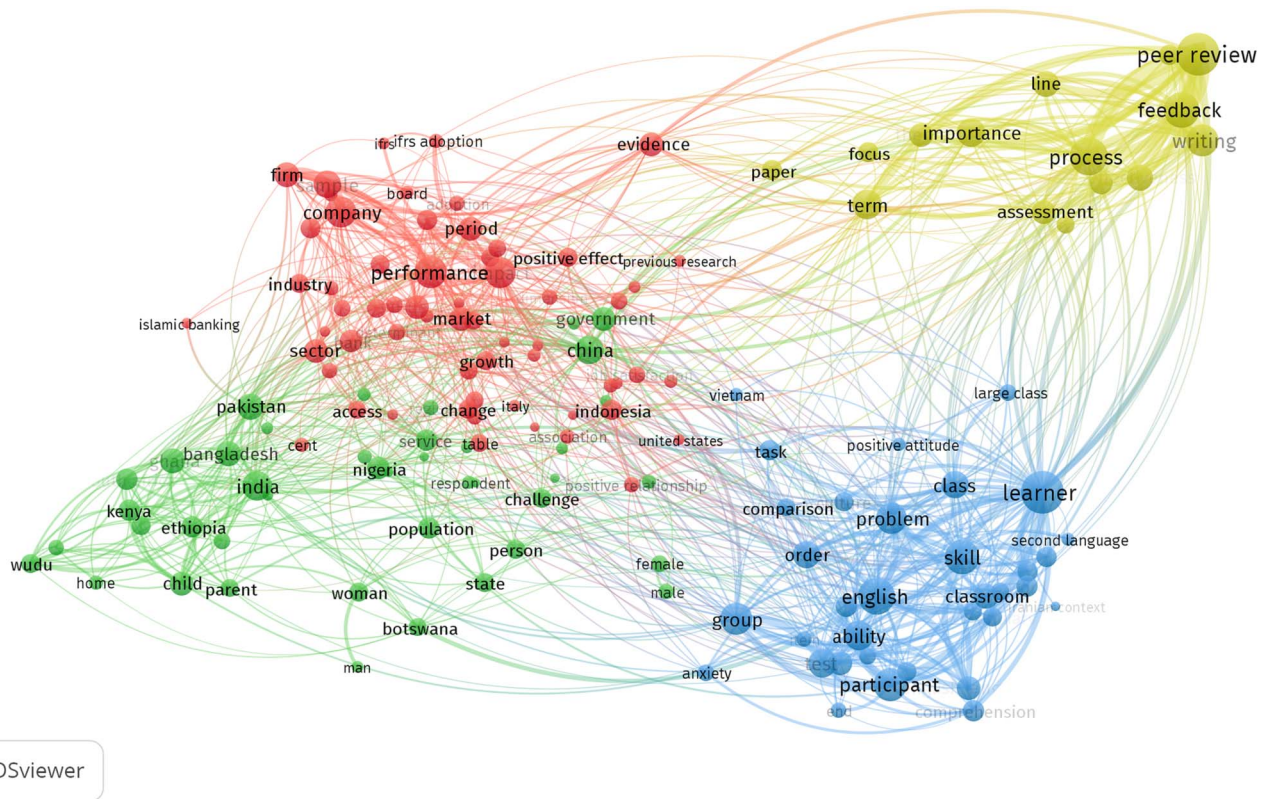


Figure 7. Co-occurrence of keywords in the citances covering country names.

distribution is obvious. The authors cited publications in questionable journals to make comparisons or describe the current state of a specific subject in a particular region. These citations could be explained by the fact that questionable journals are the only easily available journals for many non-Western scholars, and, in the end, questionable journals are the potential publication venues for many publications covering data about these countries.

As an addition to Figure 6, Figure 7 shows a co-occurrence map of the keywords. Four main subject categories were determined: economics (red), second language (blue), education (yellow), and geography-based challenges (green). The clusters show the connections between the countries and the subjects, such as economic development or academic publishing. The green cluster deserves substantial attention, as the citances in this cluster cited papers about populations, gender issues, and governments in Africa and Asia. As presented by Canagarajah (2002), there are many barriers for scholars from the periphery to successfully publish in mainstream journals from the center (e.g., different writing style requirements, biases of the reviewers, or language issues). Consequently, it is likely that many of those papers on societies in Africa or Asia have been published in questionable journals, mostly because of geopolitical inequalities in academic publishing (i.e., the authors did not meet the central journals' expectations). We have assumed that these papers are being cited because these subjects are important for understanding the current situation in the world outside Europe and North America.

To verify this assumption, we again analyzed citances that mentioned India, China, or the United States. We chose these countries because India was the most frequently mentioned country, China was the most frequently mentioned non-Western country with the relatively

Downloaded from http://direct.mit.edu/qss/article-pdf/4/1/44/2078463/qss_a_00242.pdf by guest on 07 September 2023

lowest percentage of mentions from authors of the same country, and the United States was the most frequently mentioned Western country. Unsurprisingly, most of the citations did not mention why information about a given country was taken from that and not the other journals. Nonetheless, we found a few examples of authors acknowledging that the literature on a given country was scarce. For instance, in three papers that mentioned India, the authors stated that the available literature for citing was “very limited” (Ismail & Ahmed, 2019, p. 228) or that they used it because “no exact data exist on the Indian traditional medicine industry” (Kloos, 2017, p. 1). Additionally, in one of the papers mentioning China in the citations, the lack of data on China is mentioned. Articles mentioning the United States show that mentions of a lack of literature are not specific to non-Western countries. However, only one paper mentioned a lack of studies on a certain topic in the United States and then cited a paper from a questionable journal on the topic centered around the United States. Two other citations mentioning a lack of studies on certain topics in the United States cite papers from questionable journals as proof that such studies are present in the context of other countries.

Figure 8 shows the distribution of citations in the purpose class in terms of their mentioning country names. The most interesting result was the rate of citation. Although the number of cases for the data class was lower compared to the other citation classes, the data implies that authors cite the statistics of developing countries (e.g., population, demographics, and economics) to show the current situation in these countries. This validates our previous statement. The chi-square test also confirmed the difference between the class of citations and mentioning country names ($\chi^2(10) = 152.357, p < 0.001, V = 0.087$).

The results in this section highlight researchers’ need for information about noncentral countries, which is supported by free access to knowledge provided by the transition to the open-access model in scholarly communication. However, this need leads to the practice of citing papers from journals accused of being predatory. One possible explanation for this is that researchers are often unable to find information about themes such as “Islamic banking”

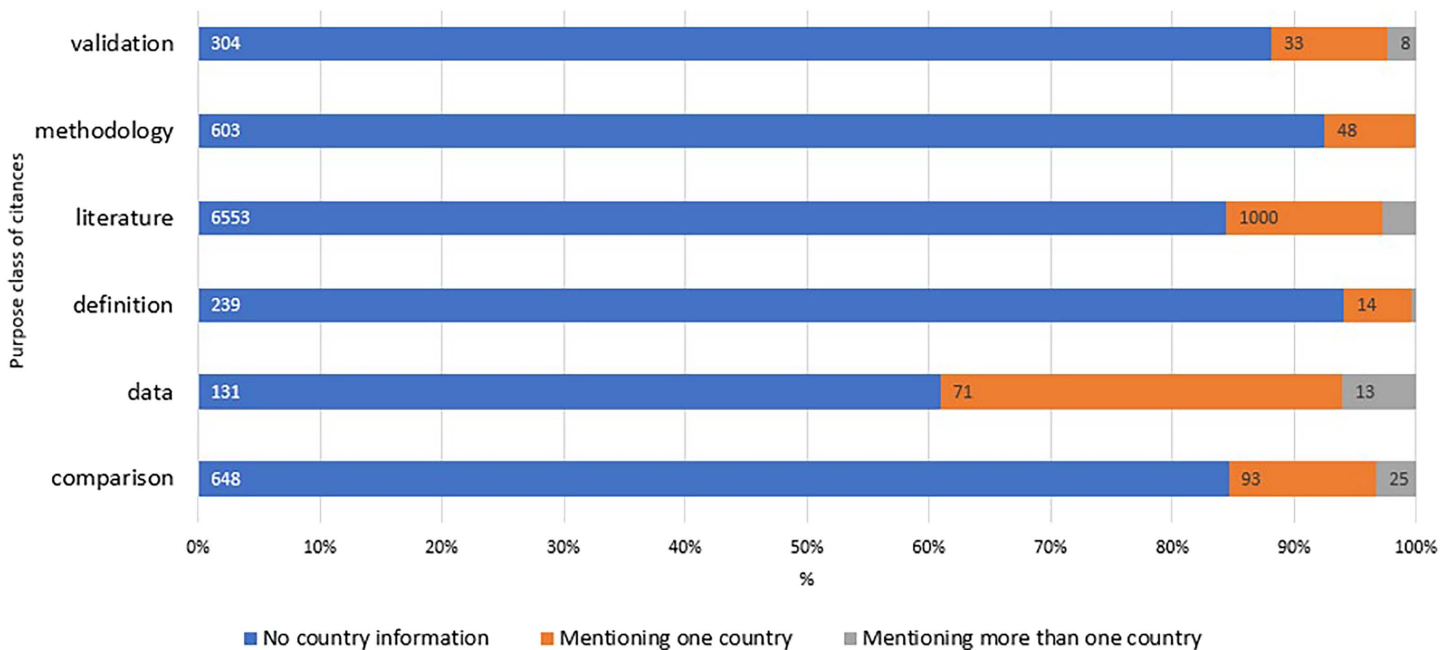


Figure 8. Distribution of citations in the purpose class in terms of including country names.

or about the language skills of students in non-Western countries in the mainstream literature. To understand this issue more clearly one can use the concept of the mislocated center of scholarly communication, which does not refer to the quality of the journals but to their position in geopolitical relations of power. From this perspective, the study results enable us to point out an important contradiction in the system: the strong delegitimization of certain journals that are mislocated centers of scholarly communication and that US scholars place on predatory journal lists and other scholars' need to cite papers that provide information about noncentral countries that is less frequently found in more Eurocentric central journals.

4. DISCUSSION AND CONCLUSION

This study aims to understand the content of the citations in WoS-indexed journals referring to questionable journals and to reveal whether these journals are as “worthless” as they are often perceived. Overall, in the present study, the citations referring to questionable journals did not show substantial differences from the legitimate content-based analysis literature. Moreover, the distribution of the citations in the citation classes closely followed that reported in the literature. In the current content-based citation analysis literature, positive and negative citations are extremely rare. For example, Taşkın and Al (2018) found that 2.0% of citations in the library and information science literature are positive and 0.2% are negative. These rates are similar to those found in the present study. However, this result was expected.

Questionable publishing is characterized by publication and editorial practices such as illegitimate peer review or misleading advertisements. A journal can be considered questionable by soliciting articles without considering their quality or contributions to science. Moreover, a conference organizer can be predatory by organizing more than 100,000 conferences in 3 years without considering the scientific contributions of the proceedings. Researchers can use questionable publication channels by publishing papers with the sole aim of obtaining tenure or other incentives. However, a scientific article cannot be questionable in the sense implied by the discussion on “predatory publishing,” especially in the unequal world of the publishing sector. The current scholarly publishing sector does not commonly consider the quality levels of articles (although there are various article-level indicators) or the contributions to science. The main evaluation mechanism of scholarly articles is still a publication venue and its metrics. However, scientific articles are a contribution to current scientific heritage, spreading knowledge across disciplines, sharing research findings, and creating new paths for new studies. The present study supports arguments regarding the importance of including a geopolitical dimension to the analysis of questionable journals. Such analysis would not be limited to assessing the difference between legitimate and questionable journals or developing new metrics but would also consider the enhancement of the accessibility (in terms of authorship or readership) of academic publishing to scholars from all regions of the world.

The practice of counting citations as well as the term *predatory journals* leads to a simplified conception of the issue of citing questionable journals. Studies previously analyzing citations referring to predatory journals have not questioned the *predatory* label—regardless of their findings. When Moussa (2021) observed a high number of citations referring to predatory journals in marketing, he stated that the risk of “infecting” (p. 503) scholarly literature is high. Additionally, when Frandsen (2017) found a low number of citations, she stated that the risk of danger from these journals is lower. However, considering the content of citations enables us to look beyond the issue of assumed predators.

An important finding of this study is that the majority of noncentral countries were mentioned by authors affiliated with these same countries. This leaves room for further studies on

regional networks of citations that are not influenced as much by the international prestige of the journal in which the cited article is published. However, this finding was not the same for every country in this study. Although 89.1% of mentions of India were written by Indian authors, only half of the mentions of China were written by Chinese authors. This does not undermine our finding that the need for information about non-Western countries is at least partially fulfilled by citing articles from journals accused of being predatory. Such findings reveal the complexity of the geopolitical issues surrounding academic publishing. One such issue is the often-biased processes of legitimization and delegitimization of journals and articles which can be influenced by the arbitrary academic writing norms (Canagarajah, 2002) or prejudices against open access journals (Krawczyk & Kulczycki, 2021a). Another issue is a general perception of which countries deserve to be mentioned in scholarly articles that also could influence observed citation patterns. If we reduce these complex issues to a blanket warning against citing predatory journals, we will only deepen geopolitical inequalities in academia instead of counteracting them.

However, paradoxically, without addressing the contradiction between the practice of accusing journals of being predatory and the practice of citing papers from these same journals, an unequal division between the centers and peripheries of science will again be supported. Our findings show that understanding some journals as mislocated centers of scholarly communication is relevant for analyzing questionable journals. When knowledge from the center is an important source of legitimization outside it (Rodriguez Medina, 2014), from the perspective of some scholars in the peripheries, mislocated centers of scholarly communication seem to be part of the center while they are mostly invisible or considered illegitimate by scholars in the center (Krawczyk & Kulczycki, 2021b). A few citations in good articles in central journals can lead scholars from the periphery to believe they are published in the “right” journal. At the same time, however, from the perspective of many central scholars or institutions using lists of predatory journals, these same scholars will be suspected of fraudulent behavior because they published in a journal on the lists.

Such findings prove the usefulness of content-based citation analysis, and this study contributes to further development in this area of study. As an important difference from the legitimate content-based citation analysis literature, the present study found that validation or comparison citations in the discussion and conclusion sections were more common. This may indicate that the authors’ main purpose behind their citations was to support or counter their views—regardless of the publication venue. This approach may be open to criticism, but, as suggested by the San Francisco Declaration on Research Assessment (San Francisco Declaration on Research Assessment, n.d.), output-based assessments instead of journal-based metrics should be used to assess the quality of articles. However, all studies on predatory journals evaluate journals, not articles. To be able to consider the articles themselves worthless, we must focus at that level.

Another important difference is the high rate of multiple citations. Authors often tend to cite collectively when they have not read the cited sources (Simkin & Roychowdhury, 2015). This may be explained by changes in authors’ motivations to cite in the “publish-or-perish” world, but it also provides important insights into the problems of citation-based performance evaluation models. Our findings validated that not all citations used as quality indicators in academia are of equal value. Citation counts are just numbers, and they do not describe the quality of the articles.

One of the important findings of this study was the existence of citations in the reference list that were not cited in the text. Although citing behaviors of some fields (e.g., various subfields

of history) is to show that the author is aware of the cited literature, the fact that these missing citations were frequently in the journals indexed in the ESCI may indicate that the editorial processes of the journals in this index are more superficial than those of JCR journals. For this reason, to minimize editorial errors, editors or editorial boards must work with a checklist and ensure the accuracy of citations.

This study showed that obtaining knowledge about non-Western countries is an important part of the phenomenon of citing questionable journals. This finding can help us argue that the main question we are dealing with is not how to eliminate all questionable or predatory journals most efficiently but, rather, how to provide better ways to communicate knowledge from many regions in Asia or Africa. To minimize the problems created by this situation, research performance evaluation models that take into account local publication practices should be developed, the diamond open-access action plan based on community publishing should be supported (Ancion, Borrell-Damián et al., 2022), and researchers should be prevented from losing their valuable work to questionable publishers. In this way, effective publishing practices will become widespread, and the problem of questionable journals will be minimized.

5. LIMITATIONS

5.1. Conceptual Limitations to Overcome in Future Studies

In this study, we analyzed the connections between WoS-indexed and questionable journals referred to by citations. However, we ignored factors such as the status, reputation, or level of journals in each category. For this reason, future investigations and multidimensional analyzes are needed to consider all angles of the subject, including author groups, publication languages, and center and periphery collaborative papers.

We evaluated the contents of citations referring to questionable journals and revealed some geographical findings for peripheral countries. However, to make accurate comparisons, some follow-up analysis for the articles in the legitimate literature is needed.

5.2. Methodological Limitations of Content-Based Citation Analysis

5.2.1. Understanding the positive and negative meanings of citations

The citation meaning class includes positive, negative, neutral, and positive and negative citations. The main aim of this classification is to understand the perceived sentiments of citers when engaging in citing. The tagging results showed that only 1.7% of the citances were positive and 0.8% of the citances were negative. This distribution was expected. Studies in the literature have revealed that positive and, especially, negative citations are extremely rare (Lacetera & Oettl, 2015; Spiegel-Rosing, 1977; Taşkın & Al, 2018). However, we would like to highlight a more important issue: the challenges of understanding the meanings of the citations.

All citances in the data set were tagged by the authors of this study between January and November 2021. Each citance was tagged by one person. Then, a list of positive and negative citances and 273 randomly selected neutral citances was sent to all three authors to confirm the citation classes. Other classes of the classification scheme were not validated in this procedure because of the nonproblematic nature of deciding, for example, in which part of the paper a citance is included. Figure 9 shows the agreement rates for the meaning class. The figure proves that understanding the meanings of the citations regarding the papers is not

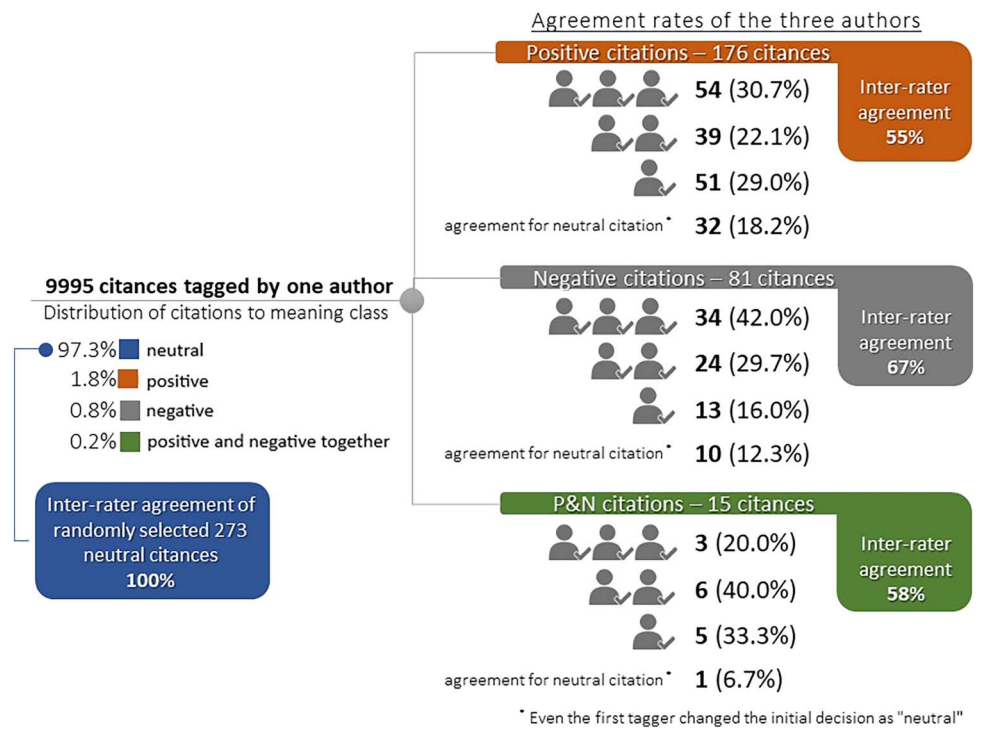


Figure 9. Distribution of citations in the meaning class and agreement rates.

an easy task. All authors agreed on the neutral citations, but the interrater agreement rates were low for the other classes. There are several reasons for this:

1. Even the first tagger disagreed with their initial decision in the next tagging session and changed some tags to neutral. This was seen in all citation classes, but predominantly for the positive citations. This means that the meaning of the citation can change even for the same person based on the tagger's mood on the day of the tagging, the noise in the environment in which the tag is made, or other reasons. This is important in terms of showing the difficulty of accurate classification in content-based citation analysis, especially for positive and negative citations.
2. Some words, such as *useful*, *comprehensively*, *significantly*, or *influential*, seemed to be used frequently around citations. These words were not always used to describe the cited article. For this reason, it is often difficult to understand to what the related words refer. The main motivation for the citation can be understood only by asking the author; however, it is not possible for researchers who cite an average of 30 sources in each article and read more to remember their views on the sources they cite. This highlights the difficulty of the meaning-based classification of citations.
3. Although some contradictory findings were presented, and comparisons were made between the cited and citing papers, some taggers considered these citations negative, while others indicated that these citations could not be considered negative. In such cases, it was difficult to distinguish between contradictory results and negative citations.
4. As evidenced by previous studies in the literature (Taşkın & Al, 2018), positive and especially negative citations are made in a very polite and implicit way. It is always difficult to understand the positive or negative intentions of citers in one sentence. For this reason, it is not only challenging for machine learning algorithms but also for humans to distinguish the true meaning of citations.

5.2.2. Finding citances in the texts

As content-based citation analyses become automated, automatic citance extraction from full texts is essential for further processing. Correctly extracting citances is a prerequisite for successful content-based citation analysis. Although it seems a simple task to extract citances according to the surnames of authors in citation styles such as the American Psychological Association (APA) or to match them with numbers in number systems, our tagging experience revealed that this process is problematic for the following reasons:

1. A citance can consist of one or more sentences. Although linking words, such as *although*, *however*, and *those*, are helpful for citances containing more than one sentence, this method does not always work. With the scope of the present study, whole paragraphs were considered in the tagging process, but it is difficult to perform the same process in automated systems because there is a need to propose rule lists for automated systems. To ensure their correct classification, it is important to be able to understand where the citance begins and ends.
2. Special characters in surnames or mistakes made by citers (e.g., citations of a first name, not a surname) create problems for finding citances in the full texts. This was one of the main limitations of this study. Mistakes by authors and the lack of control of editors can make content-based citation analysis complicated. To remedy this issue, referencing styles must be applied correctly by authors and editors alike.

5.2.3. Classifying multiple citations

As presented in the previous sections, multiple citations were a common practice of citances in the data set. However, it was very difficult to distinguish to which source the authors were referring, especially in citances containing two different interpretations (Abu-Jbara & Radev, 2012). Overcoming this challenge in content-based citation analysis is difficult, because, if the author did not comment by pointing specifically, it is impossible to identify to which work the author referred.

It is obvious that the future of citation analysis lies in content-based citation analysis. This study showed that such analysis helps to go beyond simplified divisions in highly cited and predatory journals. However, this study also confirmed that content-based citation analyses have many challenges, from data quality issues to difficulties in understanding the content. Therefore, it is important to solve these issues before applying content-based analysis to current performance evaluation systems. We expect machines to classify citations in terms of their meanings, but even experts cannot do this accurately. Considering that machine learning systems are trained by humans, there is a need for more developments in machine classification systems before using these schemes in research evaluation systems.

ACKNOWLEDGMENTS

We would like to thank Marek Hołowiecki and Abdulkadir Taşkın for their support in creating data sets, collecting data, and designing databases and interfaces.

AUTHOR CONTRIBUTIONS

Zehra Taşkın: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing. Franciszek Krawczyk: Data curation, Investigation, Writing—original draft, Writing—review & editing.

Emanuel Kulczycki: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Writing—original draft, writing—review & editing.

COMPETING INTERESTS

The authors have no conflicts of interest.

FUNDING INFORMATION

This work was financially supported by the National Science Centre in Poland (Grant Number UMO-2017/26/E/HS2/00019).

DATA AVAILABILITY

Full data (coded citations, list of articles) for this project are available at <https://osf.io/chsgp>.

REFERENCES

- Abu-Jbara, A., & Radev, D. (2012). Reference scope identification in citing sentences. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 80–90). <https://aclanthology.org/N12-1009>
- Ancion, Z., Borrell-Damián, L., Mounier, P., Rooryck, J., & Saenen, B. (2022). *Action plan for diamond open access*. <https://doi.org/10.5281/zenodo.6282403>
- Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 Student Session* (pp. 81–87). <https://aclanthology.org/P11-3015>
- Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179. <https://doi.org/10.1038/489179a>, PubMed: 22972258
- Beall, J. (2018). Scientific soundness and the problem of predatory journals. In A. B. Kaufman & J. C. Kaufman (Eds.), *Pseudoscience: The conspiracy against science*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/10747.003.0018>
- Bonzi, S. (1982). Characteristics of a literature as predictors of relatedness between cited and citing works. *Journal of the American Society for Information Science*, 33(4), 208–216. <https://doi.org/10.1002/asi.4630330404>
- Bonzi, S., & Snyder, H. W. (1991). Motivations for citation: A comparison of self citation and citation to others. *Scientometrics*, 21(2), 245–254. <https://doi.org/10.1007/BF02017571>
- Brooks, T. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37(1), 34–36. <https://doi.org/10.1002/asi.4630370106>
- Canagarajah, A. S. (2002). *A geopolitics of academic writing*. Pittsburgh: University of Pittsburgh Press. <https://muse.jhu.edu/book/27073>. <https://doi.org/10.2307/j.ctt5hjn6c>
- Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, 40(4), 284–290. [https://doi.org/10.1002/\(SICI\)1097-4571\(198907\)40:4<284::AID-ASI10>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(198907)40:4<284::AID-ASI10>3.0.CO;2-Z)
- Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423–441. <https://doi.org/10.1177/030631277500500403>
- Clarivate. (2021). New WoS April 29 Release Notes. Web of Science Group. <https://clarivate.com/webofsciencegroup/release-notes/wos/new-wos-april-29-release-notes/>
- Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation*, 37(1), 16–24. <https://doi.org/10.1108/eb026703>
- Demir, S. B. (2018). Predatory journals: Who publishes in them and why? *Journal of Informetrics*, 12(4), 1296–1311. <https://doi.org/10.1016/j.joi.2018.10.008>
- Eve, M. P., & Priego, E. (2017). Who is actually harmed by predatory publishers? *TripleC*, 15(2), 755–770. <https://doi.org/10.31269/triplec.v15i2.867>
- Eykens, J., Guns, R., Rahman, A. I. M. J., & Engels, T. C. E. (2019). Identifying publications in questionable journals in the context of performance-based research funding. *PLOS ONE*, 14(11), e0224541. <https://doi.org/10.1371/journal.pone.0224541>, PubMed: 31703069
- Frandsen, T. F. (2017). Are predatory journals undermining the credibility of science? A bibliometric analysis of citers. *Scientometrics*, 113(3), 1513–1528. <https://doi.org/10.1007/s11192-017-2520-x>
- Garfield, E. (1970). Can citation indexing be automated? *Essays of an Information Scientist*, 1, 84–90.
- Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G. L., Cukier, S., ... Lalu, M. M. (2019). Predatory journals: No definition, no defence. *Nature*, 576(7786), 210–212. <https://doi.org/10.1038/d41586-019-03759-y>, PubMed: 31827288
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>, PubMed: 25903611
- Iqbal, S., Hassan, S.-U., Aljohani, N. R., Alelyani, S., Nawaz, R., & Bornmann, L. (2021). A decade of in-text citation analysis based on natural language processing and machine learning techniques: An overview of empirical studies. *Scientometrics*, 126(8), 6551–6599. <https://doi.org/10.1007/s11192-021-04055-1>
- Ismail, S., & Ahmed, S. (2019). Economic effects of tariff liberalization of prospective India-GCC FTA: A computable general equilibrium analysis. *Foreign Trade Review*, 54(3), 224–252. <https://doi.org/10.1177/0015732519854934>
- Kloos, S. (2017). The pharmaceutical assemblage: Rethinking Sowa Rigpa and the herbal pharmaceutical industry in Asia. *Current Anthropology*, 58(6), 693–717. <https://doi.org/10.1086/693896>
- Krawczyk, F., & Kulczycki, E. (2021a). How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing. *Journal of Academic Librarianship*, 47(2), 102271. <https://doi.org/10.1016/j.acalib.2020.102271>

- Krawczyk, F., & Kulczycki, E. (2021b). On the geopolitics of academic publishing: The mislocated centers of scholarly communication. *Tapuya: Latin American Science, Technology and Society*, 4(1), 1984641. <https://doi.org/10.1080/25729861.2021.1984641>
- Kulczycki, E., Hołowiecki, M., Taşkın, Z., & Doğan, G. (2022). Questionable conferences and presenters from top-ranked universities. *Journal of Information Science*. <https://doi.org/10.1177/01655515221087674>
- Kulczycki, E., Hołowiecki, M., Taşkın, Z., & Krawczyk, F. (2021). Citation patterns between impact-factor and questionable journals. *Scientometrics*, 126(10), 8541–8560. <https://doi.org/10.1007/s11192-021-04121-8>
- Lacetera, N., & Oettl, A. (2015). The incidence and role of negative citations in science. *Proceedings of the National Academy of Sciences*, 112(45), 13823–13826. <https://doi.org/10.1073/pnas.1502280112>, PubMed: 26504239
- Marginson, S. (2021). What drives global science? The four competing narratives. *Studies in Higher Education*, 47(8), 1566–1584. <https://doi.org/10.1080/03075079.2021.1942822>
- Maričić, S., Spaventi, J., Pavičić, L., & Pifat-Mrzlijak, G. (1998). Citation context versus the frequency counts of citation histories. *Journal of the American Society for Information Science*, 49(6), 530–540. [https://doi.org/10.1002/\(SICI\)1097-4571\(19980501\)49:6<530::AID-ASI5>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4571(19980501)49:6<530::AID-ASI5>3.0.CO;2-8)
- Moussa, S. (2021). Citation contagion: A citation analysis of selected predatory marketing journals. *Scientometrics*, 126, 485–506. <https://doi.org/10.1007/s11192-020-03729-6>
- Nakov, P. I., Schwartz, A. S., & Hearst, M. A. (2004). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the SIGIR'04 Workshop on Search and Discovery in Bioinformatics* (pp. 81–88). <https://biotext.berkeley.edu/papers/citances-nlpbio04.pdf>
- Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., ... Rife, S. C. (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*, 2(3), 882–898. https://doi.org/10.1162/qss_a_00146
- Oermann, M. H., Nicoll, L. H., Ashton, K. S., Edie, A. H., Amarasekara, S., ... Ledbetter, L. S. (2020). Analysis of citation patterns and impact of predatory sources in the nursing literature. *Journal of Nursing Scholarship*, 52(3), 311–319. <https://doi.org/10.1111/jnu.12557>, PubMed: 32346979
- Read the Declaration. (n.d.). DORA. Retrieved 11 March 2022, from <https://sfedora.org/read/>
- Rochmyaningsih, D. (2019). How to shine in Indonesian science? Game the system. *Science*, 363(6423), 111–112. <https://doi.org/10.1126/science.363.6423.111>, PubMed: 30630909
- Rodriguez Medina, L. (2014). *Centers and peripheries in knowledge production*. London: Routledge Taylor & Francis Group. <https://doi.org/10.4324/9780203767016>
- San Francisco Declaration on Research Assessment. (n.d.). Sfdora. Org. <https://sfedora.org/read/>
- Siler, K. (2020). Demarcating spectrums of predatory publishing: Economic and institutional sources of academic legitimacy. *Journal of the Association for Information Science and Technology*, 71(11), 1386–1401. <https://doi.org/10.1002/asi.24339>
- Simkin, M., & Roychowdhury, V. (2015). Do you sincerely want to be cited? Or: Read before you cite. In B. Cronin & C. R. Sugimoto (Eds.), *Scholarly metrics under the microscope: From citation analysis to academic auditing* (pp. 203–210). Information Today.
- Spiegel-Rosing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science*, 7(1), 97–113. <https://doi.org/10.1177/030631277700700111>
- Taşkın, Z., & Al, U. (2018). A content-based citation analysis study based on text categorization. *Scientometrics*, 114(1), 335–357. <https://doi.org/10.1007/s11192-017-2560-2>
- Teixeira da Silva, J. A., Moradzadeh, M., Adjei, K. O. K., Owusu-Ansah, C. M., Balehegn, M., ... Al-Khatib, A. (2022). An integrated paradigm shift to deal with 'predatory publishing'. *Journal of Academic Librarianship*, 48(1), 102481. <https://doi.org/10.1016/j.acalib.2021.102481>
- Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text* [PhD, University of Edinburgh]. <https://www.cl.cam.ac.uk/~sht25/thesis/t1.pdf>
- Tinsley, H. E. A., & Weiss, D. J. (2000). Interrater reliability and agreement. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 95–124). London: Academic Press. <https://doi.org/10.1016/B978-012691360-6/50005-7>
- Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or, Did we op. cit. your idem? *Journal of Academic Librarianship*, 1, 19–21.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., ... Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Yu, T., Yu, G., & Wang, M.-Y. (2014). Classification method for detecting coercive self-citation in journals. *Journal of Informetrics*, 8(1), 123–135. <https://doi.org/10.1016/j.joi.2013.11.001>
- Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, 66(2), 408–427. <https://doi.org/10.1002/asi.23179>

APPENDIX A

Examples of citation sentences for each citation class:

Meaning	
Positive	A well-documented case illustrating such an impostor behaviour is that of BP (Beyond Petroleum; ex British Petroleum: De Wolf and Mejr, 2013).
Negative	High quality intervention studies in other age groups are largely missing. Two studies on young adults (students, age 20–22, novice, and intermediate skiing skills) after a 7-day skiing intervention (Wojtyczek et al., 2014) and on adolescents (age 14, novice skiers) after a 5-day intervention (Camliguney, 2013) reported improved balance skills, however, due to missing control groups these results have to be interpreted with caution.
Positive & Negative	Yunus, Salehi, and Chenzi (2012) witnessed the benefits of online collaboration by integrating SNS tools in an English as a second language (ESL) writing class. Their findings revealed that SNS helped broaden students' knowledge, increased their motivation, and built confidence and clarity as they developed L2 writing skills. However, their study was limited due to their small sample size and lack of either a comparison or control group. Regardless of the limitations, their qualitative findings in the form of semistructured interviews and class observation identified the important utility of SNS as an OCW learning tool as well as the need for future research.
Neutral	In Zimbabwe, one of the largest sources of regional cross-border traders, deindustrialisation associated with the economic collapse under Mugabe caused local shortages of goods and created the opportunity for both the employed and unemployed to engage in trading of multiple goods in short supply, such as food, raw materials, spares and agricultural inputs (Chiliya et al., 2012; Kachere, 2011).

Purpose	
Comparison	The result is in contrary to the submissions of some previous empirical findings that perception and acceptance of Islamic financial products are significantly influenced by the aspects of religiosity (Tara, 2014; Wilson and Liu, 2011; Akhtar et al., 2016; Kapriani et al., 2014). However, the conclusion of Amin et al. (2011) agreed with the findings of this research by stating that religious factors are apparently not always a significant predictor toward the intentions of using Islamic financial instruments.
Data	However, some studies have observed that most nurses reach professional ideals in their interactions with patients or clients during service provision. For instance a study in Mexico by Fusilier et al. [8] found that 81 per cent of health care providers interviewed were willing to provide AIDS care. It indicates that some nurses treat clients with attentive kindness and respect. Similarly, another study in Kentucky, United States of America, by Jaoko [9] noted that the majority of social workers (81%) showed positive attitudes towards persons living with HIV/AIDS.
Definition	Video activism is a means of communicating via video to influence public opinion. It is a method of protest used to counteract an abuse of power or injustice that reflects political beliefs and has the potential to transform politics and generate social change (Mateos & Gaona, 2015; Peña et al., 2015).
Literature	The culture of politics developed and inculcated therein has profoundly impacted Eritrea's internal dynamics and external relations since its independence (ICG, 2010; Nur, 2013). Eritrea's protracted struggle was enthused by a bitter resentment towards the UN, OAU and the international community because Ethiopia was tolerated when it annexed Eritrea by abrogating the federal arrangement decided by the UN.
Methodology	When $v > 0.5$, the value of Q_k will tend toward majority agreement. When $v < 0.5$, the value of Q_k will indicate majority negative attitude. In general, $v = < 0.5$ in empirical research (Mohaghar et al. 2012).
Validation	While general information about the debate topic was provided, students were required to find their own scholarly support for their arguments. This active learning strategy puts students in charge of their own learning and allows them to learn and explore the topic on their own, as opposed to reading an assigned article or listening to a lecture. This is also consistent with Weeks' (2013) hypothesis that because online debates move more slowly than face-to-face, students have more time to reflect on their learning and compose more thoughtful arguments.

Array

IMRaD structure (Introduction, Methodology, Results and Discussion) is followed for the classification. The sections that the citations appeared in the text are considered.

Shape

Mentioning author names	Quite a lot of research has been done into the (positive) results of the use of songs, especially for English. For the literature, see Engh (2013).
Multiple citations	Despite such efforts, questions and doubts on what factors actually are key to promoting research interest of academics in this teaching-oriented, low-research-support academic environment have remained intact. Some previous studies (Kwok et al., 2010; Sam, Zain, & Jamil, 2012; Chen, Sok, & Sok, 2007; Tan & Kuar, 2013) have raised a number of factors believed to explain the lack of engagement and interest in academic research activities of the country's university lecturers.
Quotation	However, the ultimate goal of the cashless policy is to eventually achieve a cashfree economy – i.e. “when all means of payments are carried out without the use of physical cash” (Ayoola, 2013). So, even when some external realities prevent a country's government from forcing their economy to go completely cash free, the rationale behind the policy remains – to constantly push their economy further and further toward making less cash available to the public.
