

Distinguishing articles in questionable and non-questionable journals using quantitative indicators associated with quality

Dimity Stephen*

*stephen@dzhw.eu

0000-0002-7787-6081

German Centre for Higher Education Research and Science Studies, Berlin, Germany

This ongoing study investigates the viability of distinguishing articles in questionable journals (QJs) from those in non-QJs on the basis of quantitative indicators typically associated with quality, and what can be deduced about the quality of articles in QJs based on the differences observed. I contrast the length of abstracts and full-texts, prevalence of spelling errors, text readability, number of references and citations, and other characteristics of 1,714 articles from 31 QJs, 1,691 articles from 16 journals indexed in Web of Science (WoS), and 1,900 articles from 45 non-WoS/non-QJs, all in the field of psychology. Initial results indicate that there are differences between QJs and non-QJ samples, however these are relatively small, perhaps indicating that QJs may not substantially differ from non-QJs on these quantitative indicators of quality. However, I intend to use additional analyses to further explore any potential differences.

1. Introduction

Questionable journals (QJs) demonstrate dishonest or dubious publishing practices, such as bypassing peer review and soliciting researchers for submissions. Consequently, the academic community is currently concerned that QJs may contaminate the discourse with poor quality research and about the flow-on effects this may have for the integrity of the academic system. However, only a few studies to date have examined the quality of articles in QJs to determine whether they actually pose a threat to academia and the broader community, and these have had mixed findings. For instance, a comparison of 410 randomised control trials (RCTs) in physical therapy found that articles in QJs complied with reporting guidelines significantly more poorly than articles in non-QJs (Bianchini et al., 2020). Yan et al (2018) similarly rated the reporting quality of 3 orthopaedics RCTs in QJs lower than that of 6 RCTs in non-QJs. In a blind comparison of the theoretical grounding, methodology, and presentation of 25 articles from questionable and non-questionable psychology journals, articles in QJs were scored significantly more poorly on all measures (McCutcheon et al., 2016). Using similar criteria, 48.4% of 353 articles in predatory nursing journals were rated as poor quality and 5% of articles contained information that could be potentially harmful to patients. However, 47.9% of articles were average and 3.7% were excellent quality (Oermann et al., 2017). As such, results regarding differences in quality between articles in questionable and non-questionable journals are mixed and often based on small sample sizes.

Further research is thus required to investigate the quality of articles in QJs to understand their potential impact on academia and the broader community. However, manual review of QJs is resource-intensive. As such, I seek in this study to determine whether articles in QJs are distinguishable from those in non-QJs based on quantitative indicators typically associated with quality, and what can be deduced about the quality of research in QJs from any differences. This ongoing study thus examines quantitative indicators such as the length of abstracts and articles, spelling errors, the quantity and characteristics of references, text readability and complexity, statistical reporting, citations, and other features of articles in samples sourced from i) QJs, ii) non-QJs indexed in the Web of Science (WoS), and iii) non-QJs not indexed in WoS. These categories represent a potential spectrum of quality; WoS-indexed journals are those selected by reviewers as high quality and central to their fields, non-WoS and non-QJs represent a middle tier of journals that are neither questionable nor top-tier, and the demonstrably

deceptive practices of QJs has cast doubt on the quality of their articles. The two non-QJ samples thus constitute a baseline for the characteristics of articles published in peer-reviewed journals that we can use to assess whether articles in questionable journals deviate from this standard. This research both informs current questions about the quality of articles in QJs, and also explores the viability of using quantitative indicators to distinguish QJs from non-QJs.

2. Methods

2.1. Sample selection and collection

I defined QJs as those listed in Cabell's Predatory Reports (CPR). CPR is a subscription-based service that classifies journals as "predatory" via manual review of the journal against more than 60 criteria, such as publishing practices and accurate use of metrics¹. I selected CPR as the basis for identifying QJs as it is current and transparent in its justification for classifying journals as questionable. To identify in-scope journals, I extracted the title and ISSNs of all journals with content published between 2010 and 2020 from Dimensions. I used Dimensions as the basis for bibliometric data because it contains a broader sample of research than the more curated content of Scopus and WoS bibliometric databases (Visser et al., 2021). Dimensions data were obtained from Digital Science by the Kompetenznetzwerk Bibliometrie (KB)² as a snapshot of data up to April 2021. As the process to identify journals in CPR was largely manual, I first reduced the sample of journals by excluding those in the Directory of Open Access Journals (DOAJ) or WoS, or published by reputable publishers such as Wiley and Springer as these journals are unlikely to be questionable and searching these publishers' names returned no results in CPR. Due to space restrictions, I refer readers to Stephen (2022a) for a complete description of this process, extended here to all journals in Dimensions, not just those with articles published by authors at German institutions. I then searched for the journals' titles/ISSNs in CPR to identify those classified as questionable as of 16 August 2022.

It was important that the journals in each sample were from the same discipline to account for any differences in article characteristics between disciplines. As a large proportion of articles from the QJs were missing discipline information in Dimensions, I restricted the samples to journals with "psycholo*" in the title, capturing those in the field of psychology. Further, I required that the article's full-text was available in order to be included in the study. Unpaywall³ is a publicly available database of Open Access (OA) content from more than 50,000 publishers that facilitates access to scholarly content. I matched in-scope articles from Dimensions and WoS with KB's snapshot of Unpaywall as of July 2022 via DOI and then extracted the URL to full-texts of articles from Unpaywall. As such, the inclusion criteria applied to all samples were i) an article published between 2010 and 2020 ii) in a journal with "psycholo*" in the title and iii) a URL to a full-text in Unpaywall. In addition, the QJ sample originated from journals classified as "predatory" by CPR, the WoS sample was drawn from journals indexed in WoS using the KB's in-house snapshot as of April 2022 and listed as OA in Unpaywall, and the non-WoS/non-QJ sample originated from journals indexed in the DOAJ. The OA inclusion criterion was included for the two non-QJ samples to account for any citation (dis)advantage as all QJs were OA.

I then extracted the journal title, article title, and the URL to the full-text from Unpaywall for all articles within scope. As I could not filter to only the "article" document type in either Unpaywall or Dimensions, I excluded as much non-research material as possible by removing

¹ <https://www2.cabells.com/about-predatory>

² <https://bibliometrie.info/>

³ <https://unpaywall.org/>

documents with keywords in the titles such as “editorial”, “book review”, “letter to/from the editor”, “correction”, “opinion”, etc. I also removed articles published in languages other than English based on titles to ensure consistency in calculating language-based indicators. I used the tidyverse package (Wickham et al., 2019) and base functions in R (R Core Team, 2023) for this processing.

2.2. Data extraction

Having identified the in-scope articles in each sample, I then downloaded a PDF version of the full-text of each article using the `download.file` function of the `utils` package in R (R Core Team, 2023). Articles that timed out or were unavailable at this URL were excluded. I then converted the PDFs to XML format via the open source software Content ExtRactor and MINER (CERMINE; Tkaczyk et al., 2015), as this format is able to be read by the JATsdecoder R package (Böschén, 2022) to extract features of academic articles. Using JATsdecoder I extracted the articles’ titles, abstracts, references, and full-texts. I manually examined the articles’ extracted titles and abstracts and corrected any missing entries or misspellings introduced by issues reading PDFs. During this cleaning, I also excluded any additional articles identified as non-research material or written in languages other than English. For articles with abstracts in more than one language, I retained the English version. I did not undertake any additional cleaning for the full-texts.

2.3. Analysis

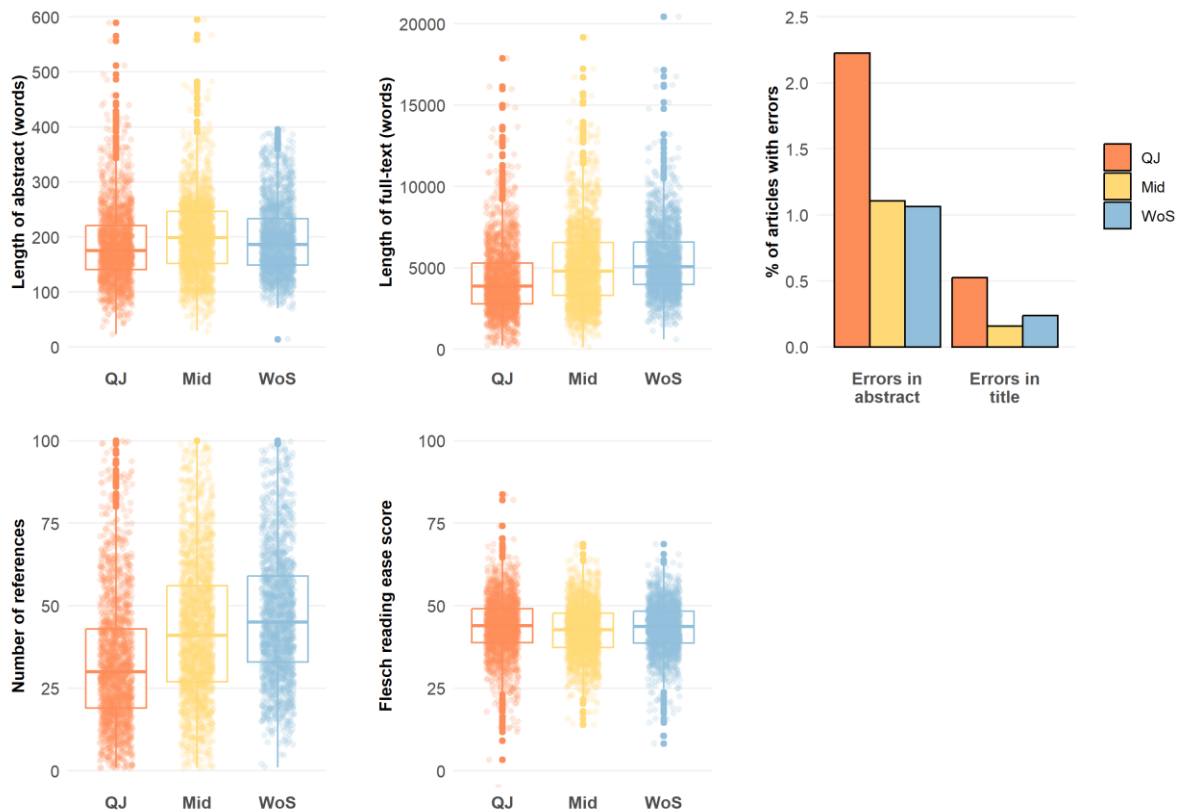
I calculated the length of the titles and abstracts of each article as the number of words contained in each using `stringr` in the tidyverse package (Wickham et al., 2019) in R (R Core Team, 2023). I also identified misspelt words in the titles and abstracts using the `hunspell` package (Ooms, 2023). I considered misspellings as those words where the intended word was discernible but misspelt (e.g., “absence”, “accomodation”, “fourty”), including proper nouns (e.g., “Chronbach”, “Bronfennbrenner”), or the wrong word was used (e.g., “exisntening”, “illuded”, “meaningly”). I accepted either British or American English spellings and ignored scientific words and jargon terms. I also calculated the number of references extracted for each article. I calculated readability scores for the abstracts and full-texts of the articles using the Flesch Reading Ease score (FRE) in the `sylcount` package (Schmidt, 2022). The FRE compares the number of syllables to the number of words in the text, and the number of words to the number of sentences (Kincaid, et al., 1975), as shorter words and sentences are more readable. Most texts score between 10 and 100, with 60-70 being an average score indicating an 8th grade reading level. However, academic articles could be expected to be more complex and aimed at college-level readers, associated with scores of 50 or lower.

3. Preliminary results and discussion

The study includes 3 samples: 1,714 articles from 31 QJs, 1,691 articles from 16 WoS-indexed journals, and 1,900 articles from 45 non-WoS/non-questionable journals. The distributions of the lengths of abstracts and full-texts, the number of references, and the FRE scores, and the percentage of articles with spelling errors in QJ, “mid”-tier (non-WoS, non-QJ) journals, and WoS-indexed journals are shown in Figure 1. Abstracts and article full-texts in QJs were typically shorter (mean = 186 and 4,229 words) than in WoS-indexed (196 and 5,434) or mid-tier (201 and 5,143) journals. Titles and abstracts of articles in QJs were also more likely than articles in the other two groups to contain spelling errors, although these percentages were low in all groups. QJs also typically had fewer references (mean = 35) than mid-tier (46) or WoS-indexed articles (50). No notable differences between groups were

observable for FRE scores, with mean scores of 42.5-43.6 in all groups. Although articles in QJs had a greater range of scores than the non-QJ groups.

Figure 1: Distributions of the lengths of abstracts and full-texts, the number of references, and the FRE scores, and the percentage of articles with spelling errors in QJ, “mid” journals, and WoS-indexed journals.



These preliminary results demonstrate that QJs do diverge from the standard set by non-QJs on quantitative indicators, such as text length, spelling errors, and the number of references. This might reflect the absence of editorial processes or peer review, as the length of manuscripts and the number of references have previously been observed to increase after peer review (Stephen, 2022b; Akbaritabar, et al., 2022). However, text characteristics, such as the FRE, did not differ.

As this study is currently ongoing, I intend to examine additional features of the articles, such as the complexity and other text-based characteristics of the articles, their structure, citation counts, accuracy in reporting statistics, and the distinct and overlapping references cited by the articles in each sample. I will also contextualise the samples by examining the most prominent countries of authors and the countries studied in the articles. The effect of journals' policies on these measures, such as the sharp cut-off of abstracts at 400 words in WoS journals, will be considered and accounted for where possible. A limitation of the study thus far is that reading PDFs inherently introduces some error into the data due to mistakes in optical character recognition or unexpected formatting. However, applying the same process to all articles means that no particular set is (dis)advantaged by any errors introduced. This and other considerations will be more fully examined in the completed study.

Open science practices

I made use of several open access data sources for this study, including the Unpaywall and Dimensions databases. These sources were instrumental to the identification of articles within the scope of my study or facilitating access to them. Further, I also used open access software packages to analyse the data, including CERMINE, R, R packages such as tidyverse, JATSdecoder, sylcount, and hunspell, and RStudio. Unfortunately, because WoS and Cabell's Predatory Reports are proprietary databases, I cannot make the raw data from my study openly available. Once the study is finalised, I will submit a preprint version of the manuscript to arXiv and also seek OA publication in a journal, including supplementary, aggregated datasets.

Competing interests

I have no competing interests to declare.

References

- Akbaritabar, A., Stephen, D., & Squazzoni, F. (2022). A study of referencing changes in preprint-publication pairs across multiple fields. *Journal of Informetrics*, 16(2), 101258. <https://doi.org/10.1016/j.joi.2022.101258>.
- Bianchini, C., Cosentino, C., Paci, M., & Baccini, M. (2020). Open Access Physical Therapy Journals: Do Predatory Journals Publish Lower-Quality Randomized Controlled Trials? *Archives of Physical Medicine and Rehabilitation*, 101(6), 969-977. <https://doi.org/10.1016/j.apmr.2019.12.012>.
- Böschén, I. (2022). JATSdecoder: A Metadata and Text Extraction and Manipulation Tool Set. R package version 1.1. <https://github.com/ingmarboeschen/JATSdecoder>.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. (1975). *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for Navy enlisted personnel*. Report No. RBR-8-75. Naval Technical Training Command Millington TN Research Branch.
- McCutcheon, L., Aruguete, M., Jenkins, W., Williams, J., Rivardo, M., & Shaughnessy, M. (2016). How Questionable Are Predatory Social Science Journals? *North American Journal of Psychology*, 18(3), 427-440.
- Oermann, M., Nicoll, L., Chinn, P., Ashton, K., Conklin, J., Edie, A., Amarasekara, S., & Williams, B. (2017). Quality of articles published in predatory nursing journals. *Nursing Outlook*, 66(1), 4-10. <https://doi.org/10.1016/j.outlook.2017.05.005>.
- Ooms, J. (2023). hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker. R package version 3.0.2. <https://hunspell.github.io>.
- R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Schmidt, D. (2022). sylcount: Syllable Counting and Readability Measurements. R package version 0.2-5. <https://cran.r-project.org/web/packages/sylcount/index.html>.

Stephen, D. (2022a). Medical articles in questionable journals are less impactful than those in non-questionable journals but still extensively cited. arXiv. <https://doi.org/10.48550/arXiv.2210.15350>.

Stephen, D. (2022b). Peer reviewers equally critique theory, method, and writing, with limited effect on the final content of accepted manuscripts. *Scientometrics* 127, 3413–3435. <https://doi.org/10.1007/s11192-022-04357-y>.

Tkaczyk, D., Szostek, P., Fedoryszak, M., Jan Dendek, P., & Bolikowski, L. (2015). CERMINE: Automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition*, 18(4), 317-335. <https://doi.org/10.1007/s10032-015-0249-8>.

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20-41. https://doi.org/10.1162/qss_a_00112.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>.

Yan, J. R., Baldawi, H., Lex, J. R., Simchovich, G., Baisi, L-P, Bozzo, A., & Ghert, M. (2018). Predatory publishing in orthopaedic research. *The Journal of Bone and Joint Surgery*, e138(1), 1-48. <http://doi.org/10.2106/JBJS.17.01569>.