RESEARCH ARTICLE

Relative prolixity in journals with different citation impact values: an evidence-based scientific writing assessment

[v1; peer review: 2 approved with reservations, 1 not approved]

Gyzelle P.V. Nascimento1, Daniel C. Moreira2, Alexis F. Welker1

1Programa de Pós-Graduação em Ciências e Tecnologias em Saúde, Faculdade de Ceilândia, Universidade de Brasília, Brasília, DF, 72220-275, Brazil
2Área de Morfologia, Faculdade de Medicina, Universidade de Brasília, Brasília, DF, 70910-900, Brazil

Abstract

Background: Scientific writing guidelines recommend that a scientific text should be straightforward, without prolixity, and informative, without obscurity. However, the extent to which researchers follow these recommendations is unknown. Considering that the most cited journals provide more detailed instructions for authors, we aimed to investigate the degree of relative prolixity (i.e., length versus amount of information) among journals with different citation impact scores.

Methods: We analyzed journals whose articles follow the classic Introduction, Method, Results, and Discussion structure, written in English and with a CiteScore value ≥ 0.01 classified in the ‘Pharmaceutical Science’ area. Relative prolixity was calculated as the ratio between the number of characters and the number of citations contained in the introductory section of original articles. Additionally, we collected the number of paragraphs and words.

Results: The number of characters, words and citations in the Introduction section were significantly higher in the journals with higher CiteScore values. The median number of paragraphs in the Introduction was not affected by the citation impact of the journals. The degree of relative prolixity in the Introduction section of the articles was negatively correlated with the CiteScore values.

Conclusions: Articles published in journals with higher CiteScore values have lower degrees of relative prolixity (i.e., shorter texts to transmit a certain amount of information) and obscurity.

Keywords

Academic writing, Bibliometrics, Health, Impact factor, Scientometrics
Introduction

A scientific text should be straightforward, without prolixity (Matthews & Matthews, 2014), and informative, without obscurity (Annesley, 2010a; Hirst et al., 2015). However, the extent to which researchers follow these recommendations is unknown. Considering that the most cited journals provide more detailed instructions for authors (Gasparyan et al., 2017), we hypothesized that articles published in such journals would be less prolix and more informative than those published in journals with fewer citations. Therefore, the objective of this study was to investigate the degree of relative prolixity (i.e., length versus amount of information) among journals with different CiteScore™.

Methods

Experimental design

To verify the degree of relative prolixity of original articles among journals with different citation impact scores, we chose to analyze journals of a specific area whose publications had their quality recently assessed – ‘Pharmaceutical Science’ area (indexed in the Scopus database) (Bohannon, 2013; Xia et al., 2015). We evaluated 101 journals (from 305) whose articles follow the classic Introduction, Method, Results, and Discussion (IMRAD) structure, written in English and with a CiteScore value ≥ 0.01.

The degree of relative prolixity was calculated as the ratio between the number of characters, which increase prolixity (Hirst et al., 2015), and the number of citations to any reference, which make the text more informative (Annesley, 2010b; Katz, 2009). Additionally, we collected the number of paragraphs and words. These analyses were done in the Introduction because this is the only section for which scientific writing literature recommends a length limit. To count the number of characters (including spaces), words, paragraphs and citations, the entire body of text of the Introduction section of each article was copied and pasted into Microsoft Office Word 2010 (Microsoft Corporation, Redmond, WA, USA). Each journal was evaluated using the median of the last three published articles in 2018.

Data analysis

Journals were grouped into quartiles of distribution based on their CiteScore values (Fernandez-Llimos, 2018). Differences between groups were assessed using the Kruskal-Wallis test and the correlation between the journals’ CiteScore and the degree of relative prolixity was assessed by the Spearman’s rank correlation test using SPSS software version 21 (IBM, Armonk, NY, USA). For all analyses, p values below 0.05 were considered statistically significant. Data are presented as mean ± standard error of the mean.

Results

The mean CiteScore values of the ‘Pharmaceutical Science’ journals divided into the four quartiles were 0.34, 1.37, 2.40 and 4.14, respectively. The median number of paragraphs in the Introduction of articles did not differ significantly between quartiles and ranged from 3.92 to 4.84 (p = 0.102; Figure 1a). Both the number of characters (including spaces) and citations (21, on average) in the Introduction were significantly higher in the journals with the highest CiteScore (p < 0.001; Figure 1b and c). The number of words in the Introduction gradually increased from the first to the fourth quartile (p < 0.001), whose average number was 442.69, 512.24, 591.00, and 721.60 words. No differences were detected between the first and second, and second and third quartiles (p > 0.05).

The degree of relative prolixity in the Introduction of the articles presented a negative correlation with the CiteScore value of the journals (p = 0.017; Figure 2).

Discussion

This study showed that the articles published in pharmaceutical science journals with higher values of CiteScore have an Introduction with a lower degree of relative prolixity and more characters, words, and citations. A low degree of relative prolixity matches the consensual recommendation that the
Correlation between the degree of relative prolixity (the ratio between the number of characters and the number of citations) in the Introduction section and the CiteScore values of journals indexed in the Pharmaceutical Science category (indexed in the Scopus database: https://www.scopus.com/sources), Spearman’s p = 0.017.

Figure 2. Correlation between the degree of relative prolixity (the ratio between the number of characters and the number of citations) in the Introduction section and the CiteScore values of journals indexed in the Pharmaceutical Science category (indexed in the Scopus database: https://www.scopus.com/sources), Spearman’s p = 0.017.

Introduction should be as short as possible (Annesley, 2010b; Armağan, 2013; Katz, 2009; Liumburo et al., 2013; Matthews & Matthews, 2014). One hypothesis to explain our results is that authors who publish in journals with higher citation impact would have better writing skills. In fact, people with less cognitive ability use more words to express information (Saling et al., 2012; Saling et al., 2016). Another hypothesis is that articles that cite more references would simply be more cited, which in fact occurs (consequently increasing CiteScore values) (Fox et al., 2016; Gargouri et al., 2010).

The higher number of references in articles from journals with higher CiteScore indicates that such articles refer to more experimental findings to support their statements. This agrees with the recommendation that a scientific text should cite the most relevant findings that directly relate to the particular scope of the study (Cals & Kotz, 2013; Seals & Tanaka, 2000; Thrower, 2008), i.e., the “closest information available in the scientific literature” (Katz, 2009). However, the citations may refer to sources that should be avoided, such as books, review articles, and papers whose degree of proximity to the original findings is low (Bavdekar, 2015; Katz, 2009). The average number of words in the Introduction found here (442.69–721.60) was higher than the recommendation of 250–300 words (Bahadoran et al., 2018; Kallestinova, 2011) and disagrees with the consensus that the Introduction should be short (Kallestinova, 2011; Liumburo et al., 2013; Matthews & Matthews, 2014; Seals & Tanaka, 2000). Furthermore, the number of paragraphs in the Introduction found in the present study (3.92–4.84) was also higher than the recommended number of up to three paragraphs (Annesley, 2010b; Bahadoran et al., 2018; Katz, 2009; Liumburo et al., 2013; Matthews & Matthews, 2014; Thrower, 2008; Vitse & Poland, 2017), which indicates that many authors are not aware of or ignore this recommendation.

In conclusion, articles in the pharmaceutical sciences journals with higher CiteScore values show lower degrees of relative prolixity (i.e., shorter texts to transmit a certain amount of information) and obscurity.

Data availability
Underlying data
Figshare: Nascimento et al. Relative prolixity Dataset 1 CC0. https://doi.org/10.6084/m9.figshare.8872253 (Nascimento et al., 2019).

This project contains the following underlying data:
- Nascimento et al. Relative prolixity Dataset 1 CC0.xlsx (Dataset containing the number of characters (including spaces), paragraphs, and citations in the Introduction of the analyzed papers)

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Grant information
The author(s) declared that no grants were involved in supporting this work.

References

Annesley TM: The title says it all. Clin Chem. 2010a; 56(3): 357–360. Published Abstract | Publisher Full Text

Annesley TM: “It was a cold and rainy night”: set the scene with a good introduction. Clin Chem. 2010b; 56(5): 708–713. Published Abstract | Publisher Full Text

Armağan A: How to write an introduction section of a scientific article? Turk J Urol. 2013; 39(Suppl 1): 8–9. Published Abstract | Publisher Full Text

Bahadoran Z, Jeddi S, Mirmian P, et al.: The Principles of Biomedical Scientific Writing: Introduction. Int J Endocrinol Metab. 2018; 16(4): e84795. Published Abstract | Publisher Full Text | Free Full Text

Bavdekar SB: Writing Introduction: Laying the Foundations of a Research Paper. J Assoc Physicians India. 2015; 63(7): 44–46. Published Abstract

Bohannon J: Who’s afraid of peer review? Science. 2013; 342(6154): 60–65. Published Abstract | Publisher Full Text

Cals JW, Katz D: Effective writing and publishing scientific papers, part III: introduction. J Clin Epistemol. 2013; 6(7): 702. Published Abstract | Publisher Full Text

Fernandez-Limos F: Differences and similarities between journal impact factor and CiteScore. Pharm Pract (Granada). 2018; 16(2): 1282. Published Abstract | Publisher Full Text | Free Full Text

Fox CW, Païne CET, Sauterey B: Citations increase with manuscript length, author number, and references cited in ecology journals. Ecol Evol. 2016; 6(21): 7717–7726. Published Abstract | Publisher Full Text | Free Full Text

Gargouri Y, Hajem C, Larivière V, et al.: Self-selected or mandated, open access increases citation impact for higher quality research. PLoS One. 2010; 5(10): e13636. Published Abstract | Publisher Full Text | Free Full Text

Gasparyan AY, Nurmashev B, Yessirkepov M, et al.: The journal impact factor: Moving toward an alternative and combined scientometric approach. J Korean Med Sci. 2017; 32(2): 173–179. Published Abstract | Publisher Full Text | Free Full Text
Kallestinova ED: How to write your first research paper. Yale J Biol Med. 2011; 84(3): 181–190.
PubMed Abstract | Publisher Full Text

Katz MJ: From research to manuscript: A guide to scientific writing. Springer, Netherlands. 2009.
Publisher Full Text

Liumbruno GM, Velati C, Pasqualetti P, et al.: How to write a scientific manuscript for publication. Blood Transfus. 2013; 11(2): 217–226.
PubMed Abstract | Publisher Full Text | Free Full Text

Matthews JR, Matthews RW: Successful scientific writing: a step-by-step guide for the biological and medical sciences. Cambridge University Press. 2014.
Publisher Full Text

Nascimento GPV, Moreira D, Welker AF: Nascimento et al. Relative proximity Dataset 1 CC0. figshare. Dataset. 2019.
http://www.doi.org/10.6084/m9.figshare.8872253.v1

Saling LL, Willis A, Saling MM: Do the Elderly Get the Message? A Comparative Study of Stories Produced Verbally and as a Text Message. J Psycholinguist Res. 2016; 45(6): 1419–1425.
PubMed Abstract | Publisher Full Text

Saling LL, Laroo N, Saling MM: When more is less: failure to compress discourse with re-telling in normal ageing. Acta Psychiatr (Amst). 2012; 139(1): 220–224.
PubMed Abstract | Publisher Full Text

Seals DR, Tanaka H: Manuscript peer review: a helpful checklist for students and novice referees. Adv Physiol Educ. 2000; 23(1): 552–558.
PubMed Abstract | Publisher Full Text

Thrower PA: Writing a scientific paper: II. Introduction and references. Carbon. 2008; 46(2): 183–184.
Publisher Full Text

Vitse CL, Poland GA: Writing a scientific paper-A brief guide for new investigators. Vaccine. 2017; 35(5): 722–728.
PubMed Abstract | Publisher Full Text

Xia J, Harmon JL, Connolly KG, et al.: Who publishes in “predatory” journals? J Assoc Inf Sci Technol. 2015; 66(7): 1406–1417.
Publisher Full Text
Kristin Lynn Sainani
Department of Epidemiology and Population Health, Stanford University, Stanford, CA, USA

I commend the authors on addressing a critical problem in the scientific literature: the poor quality of scientific writing. We need more empirical studies that address this important issue, and I hope that the authors will continue to pursue this line of research. I also appreciate that the authors have written the paper in a clear and concise manner. The data were also easy to access and understand. The main finding of interest is that there is a weak inverse correlation between CiteScore (a measure of a journal's impact) and prolixity (a measure of wordiness). This may be evidence that journals with better writing get more citations, although the evidence for this conclusion is fairly weak. I believe that the paper would be strengthened by considering additional metrics of readability. Specific comments:

1. I probably would have chosen the word “wordiness” over “prolixity” in keeping with the theme of encouraging authors to write in a straightforward manner. Prolixity is a more academic term, whereas wordiness is simpler and easier to understand.

2. I'm not sure why the authors have looked at the number of paragraphs and number of characters as outcomes in themselves. The length of the introduction doesn't tell us about writing quality or wordiness. A long piece can be well-written and a short piece can be poorly written. Also, different journals have different constraints (such as constraints on word counts) that may affect the length of the introduction. Thus, I don't think these outcomes are very informative, and need not be highlighted in graphics.

3. Why were characters, numbers, and citations analyzed by CiteScore quartile but relative prolixity analyzed treating CiteScore as a continuous variable? In Figure 1, I expected to see a Figure 1d that showed the average prolixity per quartile.

4. This paper would be strengthened by considering other metrics of readability beyond just the ratio of characters to citations. There are numerous online tools that allow one to measure readability with validated measures such as the Flesh Reading Ease Scale. See for example, this tool: http://www.checktext.org/. Examining the correlation between CiteScore and readability scores would add to the impact of this paper.
5. Journals sometimes have limits on the number of references, which would influence the number of citations appearing in the introduction section. The authors should state whether any of the journals examined had such limitations.

6. The aim of this study was to gauge the association between a journal’s CiteScore and a measure of the journal’s prolixity. But only 3 samples per journal were taken. Given that prolixity may vary widely from paper to paper within the same journal, I believe that a larger sample size per journal would have strengthened this study. It’s not clear that the last three papers published are going to be representative of all studies in a journal. The authors should comment on how they arrived at the choice of 3 studies per journal.

7. The authors should give the magnitude of the correlation coefficient between CiteScore and prolixity. I believe it’s about -0.20, which would be considered an extremely weak correlation even though it is statistically significant. The authors should comment on the fact that the correlation is weak. Also, there is one study with high prolixity (>600) that is making it hard to see the pattern of correlation in Figure 2; consider presenting the graph both with and without this point.

8. Bar graphs are not ideal. I would recommend that the authors replace bar graphs with box plots with individual points overlaid, as these are more informative.

9. The authors should read and reference this study on text readability in the scientific literature (Plavén-Sigray et al., 2017).

References
1. Plavén-Sigray P, Matheson G, Schiffler B, Thompson W: The readability of scientific texts is decreasing over time. eLife. 2017; 6. Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly
**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** athletic injuries, epidemiology, statistics, scientific communication

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 25 November 2019

https://doi.org/10.5256/f1000research.21971.r56409

© 2019 Habibzadeh F. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Farrokh Habibzadeh

1 The International Journal of Occupational and Environmental Medicine (The IJOEM), Shiraz, Iran
2 Past President, World Association of Medical Editors (WAME), Shiraz, Iran
3 R&D Headquarters, Petroleum Industry Health Organization, Shiraz, Iran
4 Editorial Consultant, The Lancet, Shiraz, Iran

In this manuscript, the authors were supposed to examine the relationship between the "relative prolixity" and the CiteScore, an index reflecting the journal quality. They found that articles published in journals with higher CiteScore have lower relative prolixity. There are, however, many major problems with the hypothesis and the methodology used.

**Major Problems:**

1. The relative prolixity was defined “as the ratio between the number of characters and the number of citations contained in the introductory section of original articles.” The definition was based on the assumption that increasing the number of characters would increase prolixity, which might be correct, and that citing any reference would make the text more informative, which is correct but NOT relevant to the hypothesis of this study. In fact, although citing any reference would add some information, the number of citations does not necessarily reflect the amount of information conveyed in an article. There is no evidence indicating that the number of citations in the Introduction is associated with the information provided by an article, to the best of my knowledge. Most of the information conveyed in an article is presented in the Results section. Therefore, reasonable surrogates for the amount of information conveyed in an article would be the number of figures in Tables, the number of curves, number of columns in a bar chart, number of points in a scatter plot, number of p values presented (number of hypotheses tested), to name just a few.

2. As all the conclusions were made based on an invalid index, the study has neither internal nor external validity.
3. In Statistical Methods, the authors should have considered correction of p values reported for multiple comparisons made.

4. Having extreme values would increase the absolute value of the correlation coefficient. This should be considered in interpreting the results presented in Fig. 2.

**Minor Problems:**
1. The Spearman’s rho is not reported.

2. There is no definition for “obscurity.” The authors should thus not comment on the obscurity of articles.

3. In journalism, we usually refer to journals with the highest CiteScore as Q1, not Q4. In this manuscript, the authors did categorize in an opposite way.

4. As Richard Smith, the former Editor of BMJ, has pointed out, many authors do not read the instructions for authors of journals at all, no matter how detailed they are.

**Is the work clearly and accurately presented and does it cite the current literature?**
Partly

**Is the study design appropriate and is the work technically sound?**
No

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Partly

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
No

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Medical Journalism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.
Asghar Ghasemi  
Endocrine Physiology Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

The paper by Nascimento et al. analyzed whether “relative prolixity” of published papers in the field of pharmaceutical science is related to quality of Journals (assessed by CiteScore values). Although, efforts on improving science of writing are acknowledged, in my opinion, the presented paper is subjected to some criticisms as below.

**Major issues:**
1. The index that the authors used to assess prolixity of the published papers seems not to be valid to lead to a legitimate conclusion. The authors did not provide a good background for the index. I could not find any information about “relative prolixity index” in the references that the authors refer to (Hirst et al., 2015; Annesley, 2010; Katz, 2009). There are other scores that authors can use; e.g. “Fog index” (Gunning, 1969[ref-1]) which was developed by Robert Gunning to test readability of a paragraph or passage.

2. The conclusion of the study is mostly based on the results of a correlation between CiteScore and relative prolixity; the number of journals with high cite score (~10) seems to be low (~10) that affect the correlation coefficient. In addition, correlation per se cannot provide a basis for conclusion. In fact, in this study the authors conclude that the more the references in the introduction, the less the relative prolixity. This is not straightforward as other factors such as topic presented, number of hypotheses being tested, etc. can affect the number of references in the introduction section of a paper. I suggest to consider relevancy of the citations in papers studied.

3. Some reasoning provided by the authors for the observed association between the index and Journals’ CiteScore values needs revision; the authors stated that “people with less cognitive ability use more words” in their writing. There are more important factors like being a non-native English speaker that influence quality of writing.

**Minor issues:**
1. Correlation coefficient and “n” need to be reported in Figure 2.

2. Multiple comparisons need to be mentioned for ANOVA.

3. Guides for authors may also affect the results.

**References**
1. Gunning R: The Fog Index After Twenty Years. *Journal of Business Communication*. 1969; 6 (2): 3-13 Publisher Full Text

*Is the work clearly and accurately presented and does it cite the current literature?*
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am working on Nitric oxide and diabetes but also have some experience in scientific writing. I teach Scientific writing and have some publications in this field.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com