Scientometrics
https://doi.org/10.1007/s11192-019-03332-4

Write better, publish better

Diego Marino Fages[1]

Received: 20 September 2019

Abstract
There is evidence that having more readable abstracts and introductions help authors get cited. I show that, in economics, there is also an effect of readability on the probability of publishing in a Top 5 journal (and in a higher-ranked journal in general). I compute readability measures for a set of working papers and examine the journals in which they get published. My results suggest that previous estimates of the effect on citations are downward biased, as higher-ranked journals are more widely read and cited.

Keywords Readability · Journals · Ranking · Top 5 · NBER

JEL Classification A1

Introduction

In *The Tyranny of the Top Five Journals*,[1] James Heckman and Sidharth Moktan state that "Without doubt, publication in the T5 [Top 5] is a powerful determinant of tenure and promotion in academic economics" . They discuss the consequences of this state of the profession and propose solutions. Unfortunately, until the tyranny is brought down, it will still be of great importance for researchers, especially at the beginning of their careers, to publish in the top five journals. In this article, I ask whether writing clearly increases the likelihood of publishing in one of them.

Relying on formal readability measures, previous studies have shown that readability has a positive effect on citations (Dowling et al. 2018; McCannon 2019), on winning more awards (Sawyer et al. 2008) and on having more downloads (Guerini et al. 2012). Journals also benefit from this, as shown by Richardson (1977) and Swanson (1948), the more readable academic journals enjoy larger readerships.

Complementing Dowling et al. (2018) and McCannon (2019) who work with the abstracts from the *Economics Letters* and the introductions of the *American Economic*

[1] Heckman and Moktan (2018).

✉ Diego Marino Fages
diego.marinofages@nottingham.ac.uk
https://sites.google.com/view/diegomarinofages

[1] School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK

Review, respectively, I focus on the abstracts of working papers published in the *National Bureau of Economic Research* (NBER) website. It must be noted that only affiliated members can publish working papers in the NBER. On the website one can read that "the Research Associates at NBER hold tenured positions at their home institutions" and that they "are the leading scholars in their fields" . There are two reasons to focus on these working papers. First, it allows me to study a broader spectrum of journals. Second, since the affiliated members are highly biased towards top institutions, the sample should be more homogeneous in terms of researcher abilities, resources, and so on, which helps mitigate the potential selection problem that would arise if the quality of the writing is correlated with the quality of the research.

Methodology

As Dowling et al. (2018) argue, abstracts are the portion of an article that is the most widely read. Furthermore, there is some evidence suggesting the possibility of a correlation between the style of the abstract and other sections of a paper (Hartley et al. 2003).[2] As long as this correlation applies in the economics context, studying the abstracts would be justified. In the rest of the paper I focus on the abstracts.

I scrape abstracts of working papers from the website of the NBER for the period 1998–2019. However, in the analysis, I restrict the sample to working papers issued before 2015 since they might still be in publication process.[3] NBER working papers can only be uploaded by its members, which allows me to control for researcher quality, as they belong to the top institutions in the US. After cleaning the data, the final sample consists of 9757 working papers written by 7621 authors (Marino Fages 2019).

In linguistics, there is a myriad of measures of writing quality, ranging from basic readability indices (length of syllables, words, sentences, number of complex words and so on) to the use of adjectives or adverbs (Okulicz-Kozaryn 2013) and linguistic complexity (Lu et al. 2019). In this paper, I focus on traditional readability measures and throughout the paper I use the terms "higher readability" and "better-written" interchangeably. However, the reader should notice that these measures cover only some aspects of the quality of writing.

Since there are hundreds of readability measures, I follow Hengel (2017), who claims that the most widely used, tested and reliable measures are the following: Flesch Reading Ease (FRES), Flesch–Kincaid (FKS), Gunning Fog (FOG), Simple Measure of Gobbledegook (SMOG) and Dale–Chall (DCS). The exact formulae are the following:[4]

[2] The study relies on papers from the Journal of Education Psychology.

[3] Earlier working papers may also be in publication process, however, I consider this to be a reasonable cut-off.

[4] I compute the scores using textatistic 0.0.1 for python (See https://pypi.org/project/textatistic/).

$$\text{DCS} = \begin{cases} 1 & \text{if } \left(\frac{\text{difficult}}{\text{words}}\right) > 0.05 + 15.79 * \left(\frac{\text{difficult}}{\text{words}}\right) + 0.0496\left(\frac{\text{words}}{\text{sentences}}\right) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{FRES} = 206.835 - 1.015 * \left(\frac{\text{words}}{\text{sentences}}\right) - 84.6 * \left(\frac{\text{syllables}}{\text{words}}\right)$$

$$\text{FKS} = -15.59 + 0.39 * \left(\frac{\text{words}}{\text{sentences}}\right) + 1.18 * \left(\frac{\text{syllables}}{\text{words}}\right)$$

$$\text{FOG} = 0.4\left[\left(\frac{\text{words}}{\text{sentences}}\right) + 100\left(\frac{\text{complex}}{\text{words}}\right)\right]$$

$$\text{SMOG} = 3.1291 + 1.043 * \sqrt{\text{complex} * \frac{30}{\text{sentences}}}$$

where a complex word is defined as one with three or more syllables excluding common endings (e.g. '-ing') and difficult words are those not found on a list of 3000 words understood by 80 percent of fourth-grade readers (aged 9–10).

Since readability has a direct relation to the FRES but an inverse relation to the other measures, I multiply all the others by − 1 such that a larger score indicates better readability. After the transformation, all measures correlate strongly and positively.

Descriptive statistics

Table 1 presents the means and standard deviations of the main variables for the full sample and the Top 5 journals (See "Appendix"). The Top 5 journals represent 17% of the sample, have higher readability scores in all measures and less single-authored papers. The last two variables are the Scimago Journal Rank Indicator (1999) and the Impact Factor (1997). In both cases, a higher value represents a better journal. I use these years to prevent

Table 1 Descriptive statistics

	Full sample		Top 5	
	Mean	SD	Mean	SD
Words	151.91	(52.53)	152.66	(51.84)
Sentences	6.21	(2.29)	6.36	(2.26)
Pages	45.23	(16.31)	46.99	(17.68)
DCS	− 10.87	(0.99)	− 10.86	(0.99)
FRES	41.21	(11.55)	41.95	(11.53)
FKS	− 13.82	(2.82)	− 13.58	(2.8)
FOG	− 17.37	(3.15)	− 17.14	(3.15)
SMOG	− 15.14	(2.09)	− 15.00	(2.1)
Single authored	0.21	(0.41)	0.17	(0.38)
Scimago 1999	5.87	(4.02)	9.50	(3.95)
Impact Factor 1997	1.52	(0.97)	2.08	(0.66)
Observations	9757		1621	

Table 2 Top 5 journals statistics

Journal	*N*	DCS	FRES	FKS	FOG	SMOG	Single authored	Scimago	Impact Factor
AER	787	− 10.85	42.14	− 13.64	− 17.15	− 14.97	0.20	6.15	1.63
ECTA	134	− 11.22	38.48	− 13.89	− 17.59	− 15.47	0.14	11.87	1.71
JPE	248	− 10.89	41.85	− 13.59	− 17.18	− 15.03	0.15	11.20	2.85
QJE	317	− 10.67	43.43	− 13.28	− 16.78	− 14.76	0.14	16.28	3.07
ReStud	135	− 10.98	41.00	− 13.57	− 17.34	− 15.24	0.18	7.94	1.42
EcLetters	30.00	− 11.08	39.50	− 14.47	− 18.16	− 15.60	0.30	0.80	0.26

the indexes from being affected by the papers in the sample.[5] The correlation between the Scimago Index and the Impact Factor is 0.82 .[6]

Table 2 summarizes the means of the readability scores, the proportion of single authors and the Scimago Index and Impact Factor for the Top 5 journals separately. *Econometrica* scores comparatively low in all measures, bringing the averages presented in Table 1 closer.

I also include *Economics Letters* as it allows me to compare abstracts from working papers with the published version. Dowling et al. (2018), using the published version (from 2003 to 2012) obtains a FRES = 42.46, FOG = −16.08 and SMOG = −14.20. All of them are better than in the working paper version, which is consistent with the submission process improving the readability.

Analysis

In Table 3 (columns 1–5), I estimate by OLS a linear probability model[7] of *Top 5* on each of the readability measures: i.e. $\text{Prob}[\text{Top5} = 1] = \alpha + \beta * \text{ReadabilityScore} + \delta * \text{Controls} + \epsilon$. Since the results virtually do not change, I only report the full specification regressions, where I control for JEL codes, number of pages, number of coauthors, its square, and dummy variables for the working paper's year and the year of publication.

Almost all regressions show a significant positive effect of the readability on the probability of becoming a Top 5 that ranges between 0.1 and 0.6% points. This is not a small effect considering the standard deviations in Table 1 and the fact that it is not related to the actual scientific contribution of the papers. For instance, increasing one standard deviation of the FRES is associated with a 1.16% points increase in the probability of being in a Top 5 journal.

The number of coauthors has a positive and decreasing effect on Top 5 (this is consistent with Hollis 2001). In non-reported results, I also find that the number of coauthors is also strongly associated with better readability measures (and also decreasing).

[5] For the analysis with the Scimago Journal Rank Indicator, I restrict the sample to working papers issued since 2000.

[6] For a comparison of the pros and cons of these indexes see Falagas et al. (2008).

[7] Results of probit models are similar.

Table 3 OLS regressions of *Top 5* on readability scores

VARIABLES	(1) Top 5	(2) Top 5	(3) Top 5	(4) Top 5	(5) Top 5	(6) Top 5	(7) Top 5	(8) Top 5	(9) Top 5	(10) Top 5
DCS	0.00567 (0.00387)					0.00770 (0.00587)				
FRES		0.00108*** (0.000328)					0.00156*** (0.000509)			
FKS			0.00491*** (0.00134)					0.00566*** (0.00218)		
FOG				0.00413*** (0.00121)					0.00539*** (0.00193)	
SMOG					0.00577*** (0.00182)					0.00854*** (0.00280)
No. Authors	0.0422*** (0.0125)	0.0409*** (0.0125)	0.0411*** (0.0125)	0.0413*** (0.0125)	0.0414*** (0.0125)	− 0.0151 (0.0255)	− 0.0171 (0.0256)	− 0.0163 (0.0256)	− 0.0162 (0.0255)	− 0.0164 (0.0255)
No. Authors squared	− 0.00608*** (0.00222)	− 0.00589*** (0.00223)	− 0.00591*** (0.00223)	− 0.00597*** (0.00223)	− 0.00598*** (0.00223)	0.00266 (0.00479)	0.00300 (0.00480)	0.00286 (0.00480)	0.00282 (0.00479)	0.00284 (0.00479)
Constant	0.0623 (0.0940)	− 0.0422 (0.0838)	0.0716 (0.0857)	0.0761 (0.0865)	0.0916 (0.0887)	0.258 (0.176)	0.116 (0.159)	0.256 (0.164)	0.275* (0.164)	0.314* (0.166)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9685	9685	9685	9685	9685	6034	6034	6034	6034	6034
R-squared	0.037	0.038	0.039	0.038	0.038	0.059	0.061	0.060	0.061	0.061

Robust standard errors in parentheses

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table 4 OLS regressions of Scimago index on readability scores

VARIABLES	(1) Scimago	(2) Scimago	(3) Scimago	(4) Scimago	(5) Scimago
DCS	0.117** (0.0546)				
FRES		0.0176*** (0.00456)			
FKS			0.0787*** (0.0181)		
FOG				0.0627*** (0.0163)	
SMOG					0.0854*** (0.0249)
No. Authors	0.198 (0.246)	0.174 (0.246)	0.183 (0.246)	0.185 (0.246)	0.184 (0.245)
No. Authors squared	− 0.0346 (0.0475)	− 0.0304 (0.0476)	− 0.0318 (0.0476)	− 0.0326 (0.0474)	− 0.0325 (0.0474)
Constant	4.683*** (0.880)	2.725*** (0.667)	4.570*** (0.701)	4.560*** (0.714)	4.731*** (0.758)
Other controls	Yes	Yes	Yes	Yes	Yes
Observations	5072	5072	5072	5072	5072
R-squared	0.134	0.136	0.136	0.135	0.135

Robust standard errors in parentheses

$p < 0.01$; $^{**}p < 0.05$, $^{*}p < 0.1$

As a robustness check, I standardize the five measures and take the average of them. The full regression with the JEL codes gives a coefficient of 0.019 (p value = 0.143) which means that increasing the readability in 1 standard deviation in this global measure would increase the probability of being a Top 5 of 1.9% points.

As a second robustness check, in Table 3 (columns 6–10), I restrict the sample to working papers that were published in highly ranked journals (see "Appendix"). The effects do not change significantly and, if anything, get slightly stronger. However, the significance of the number of authors disappears.

Next, I extend the analysis to check if readability is associated with the ranking of the journal in general (i.e. not only Top 5). For this, I use two measures. First, I use the Scimago Journal Rank Index (Table 4). To prevent the papers in the sample to affect the index, I use the earliest available (1999) and restrict the sample to the papers issued since 2000 onwards. Again, I only present the full specification controlling for JEL codes, number of pages, number of coauthors and its square, and dummy variables for the working paper's year and the year of publication. For all measures, the readability has a significant effect on the Scimago index. On the other hand, the number of coauthors presents no effect. Second, I use the Impact Factor from 1997. In this case I do not need to drop the years before 2000, but the sample is smaller because there are some journals with no Impact Factor. The coefficients are all positive and significant at the 1% level in all cases (the results are available in the do file). The number of coauthors in this case has a negative effect.

Finally, instead of focusing on papers, I aggregate at the journal level, keeping the score of the median paper. This allows me to check whether journals that are better written are

Table 5 OLS regressions of Scimago index on median journal score

VARIABLES	(1) Scimago	(2) Scimago	(3) Scimago	(4) Scimago	(5) Scimago
Median DCS	− 0.0276 (1.251)				
Median FRES		0.189* (0.111)			
Median FKS			1.464*** (0.521)		
Median FOG				0.755* (0.378)	
Median SMOG					0.601 (0.609)
Constant	3.455 (13.74)	− 3.955 (4.373)	23.72*** (7.352)	16.76** (6.735)	12.83 (9.359)
Observations	54	54	54	54	54
R-squared	0.000	0.032	0.080	0.030	0.011

Robust standard errors in parentheses

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

higher in the ranking. In spite of having only 54 observations (see the list in the "Appendix") three out of five measures are positive and significant (Table 5). As a robustness check, I run the same regression using the Impact factor of 1997 as a dependent variable. All the results (available in the do file) have positive coefficients but, probably because of a lack of power (only 43 observations) none of them can be rejected.[8] The smallest p value is 0.12 (results available in the dofile).

Discussion

In the conclusion of his paper, McCannon (2019) argues that his estimates might be downward biased because he is not taking into account the effect of the selection of papers by the journals. In this article, I confirm his hypothesis.

Using a fairly homogeneous set of papers in terms of quality, I provide suggestive evidence that the quality of the writing, measured by formal readability scores, is associated with better publications. However, because of the lack of a credible identification strategy, I cannot claim any causality in the relations. In particular, reverse causality might be an issue if researchers adapt the quality of the writing depending on their expectations of the quality of the papers.

I find a positive and significant effect of having better-written abstracts on the probability of being published in a Top 5 journal. The effect seems to be of great magnitude considering that the measures do not include anything related to the actual scientific contribution of the papers.

[8] I get the same results if I use the Impact Factor of 1999 instead.

More generally, I find that higher readability is associated with a higher-ranked journal (even when restricting the sample to the very top). These results still apply when I look at the journal level, journal readability correlates with its Scimago index.

Finally, the number of collaborators in a paper has a positive and significant effect on the readability measures and on the probability of becoming Top 5, however, I find no effect on the Scimago index and even a negative effect on the Impact Factor.

Although the data does not allow me to explore the mechanisms, there may be two possible reasons for my results. First, a psychological explanation would say that easier papers to read are received more favorably by editors and reviewers because they require less effort. Second,[9] if the papers are circulated as working papers, then they can already start being cited. In a model in which editors are only interested in getting the journal cited, they may only accept the already highly cited working papers. In this case, the effect of readability on the ranking of the journal in which the article gets published comes indirectly from the effect of readability on citations as a working paper. Contrasting these two mechanisms would be an interesting next step.

Acknowledgements I thank Facundo Albornoz, Federico Bernini, Kris Gulati, Gabriele Luchetti and two anonymous referees for useful comments; and Federico Bennett, Katie Harrison and Richard Mills for their help in improving the readability.

Appendix

See Table 6

[9] I thank one of the referees for this insight.

Table 6 List of highly ranked journals with Scimago Index (1999) and Impact Factor (1997) *Sources*: https://www.scimagojr.com/journalrank.php?category=2002&year=1999. https://jcr.clarivate.com/JCRJournalHomeAction.action?pg=JRNLHOME&categoryName=ECONOMICS&year=2018&edition=SSCI&categories=GY

Journal	Scimago Index 1999	Impact Factor 1997	Journal	Scimago Index 1999	Impact Factor 1997
American Economic Review	6.146	1.630	Journal of Financial Economics	9.621	2.506
Canadian Journal of Economics	0.759	0.153	Journal of Financial Intermediation	2.843	–
Econometrica	11.873	1.712	Journal of Human Resources	3.948	1.233
Economic Development and Cultural Change	1.781	0.493	Journal of Industrial Economics	1.882	0.636
Economic Inquiry	1.484	0.516	Journal of International Economics	3.812	0.840
Economic Journal	2.698	1.158	Journal of International Money and Finance	0.907	–
Economica	1.232	0.500	Journal of Labor Economics	4.500	0.800
Economics Letters	0.798	0.258	Journal of Law and Economics	1.991	1.146
Economics of Education Review	0.604	0.284	Journal of Monetary Economics	4.002	0.905
European Economic Review	2.772	0.692	Journal of Money, Credit and Banking	2.533	–
Games and Economic Behavior	3.728	0.699	**Journal of Political Economy**	11.201	2.854
International Economic Review	3.291	0.500	Journal of Public Economics	2.566	0.713
Journal of Accounting and Economics	4.031	0.509	Journal of Regulatory Economics	0.933	0.270
Journal of Applied Econometrics	1.337	0.971	Journal of Risk and Uncertainty	0.975	0.583
Journal of Banking and Finance	0.795	0.351	Journal of Urban Economics	1.387	0.548
Journal of Business and Economic Statistics	1.919	1.295	Labour Economics	1.595	–
Journal of Development Economics	2.257	0.590	Macroeconomic Dynamics	1.137	–
Journal of Econometrics	3.314	1.011	Oxford Economic Papers	1.264	0.641
Journal of Economic Behavior and Organization	0.973	0.352	Pacific Economic Review	0.631	–
Journal of Economic Dynamics and Control	1.928	0.312	**Quarterly Journal of Economics**	16.278	3.074
Journal of Economic Growth	11.491	–	RAND Journal of Economics	4.275	1.012
Journal of Economic Literature	10.402	5.818	Review of Economic Dynamics	0.590	–

Table 6 (continued)

Journal	Scimago Index 1999	Impact Factor 1997	Journal	Scimago Index 1999	Impact Factor 1997
Journal of Economic Perspectives	7.823	2.771	**Review of Economic Studies**	7.939	1.421
Journal of Economic Surveys	1.221	–	Review of Economics and Statistics	3.436	1.163
Journal of Economic Theory	4.790	0.606	The Review of Financial Studies	6.273	–
Journal of Environmental Economics and Management	1.037	2.064	World Bank Economic Review	3.747	1.643
Journal of Finance	9.553	–	World Development	1.529	0.818

Top 5 journals in bold.

References

Dowling, M., Hammami, H., & Zreik, O. (2018). Easy to read, easy to cite? *Economics Letters*, *173*, 100–103.

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB Journal*, *22*(8), 2623–2628.

Guerini, M., Pepe, A., & Lepri, B. (2012). Do linguistic style and readability of scientific abstracts affect their virality? In *Sixth international AAAI conference on weblogs and social media* (pp. 475–478).

Hartley, J., Pennebaker, J. W., & Fox, C. (2003). Abstracts, introductions and discussions: How far do they differ in style? *Scientometrics*, *57*(3), 389–398.

Heckman, J. J, & Moktan, S. (2018). *Publishing and promotion in economics: The tyranny of the top five*. Technical Report, National Bureau of Economic Research.

Hengel, E. (2017). *Publishing while Female. Are women held to higher standards? Evidence from peer review*.

Hollis, A. (2001). Co-authorship and the output of academic economists. *Labour Economics*, *8*(4), 503–530.

Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., et al. (2019). Analyzing linguistic complexity and scientific impact. *Journal of Informetrics*, *13*(3), 817–829.

Marino Fages, D. (2019). Data for: Write Better, Publish Better, Mendeley Data, V1. https://doi.org/10.17632/mmfttcywgy.1.

McCannon, B. C. (2019). Readability and research impact. *Economics Letters*, *180*, 76–79.

Okulicz-Kozaryn, A. (2013). Cluttered writing: Adjectives and adverbs in academia. *Scientometrics*, *96*(3), 679–681.

Richardson, J. V, Jr. (1977). Readability and readership of journals in library science. *Journal of Academic Librarianship*, *3*(1), 20–22.

Sawyer, A. G., Laran, J., & Xu, J. (2008). The readability of marketing journals: Are award-winning articles better written? *Journal of Marketing*, *72*(1), 108–117.

Swanson, C. E. (1948). Readability and readership: A controlled experiment. *Journalism Bulletin*, *25*(4), 339–343.