**ORIGINAL PAPER**

# Is the readability of abstracts decreasing in management research?

Lorenz Graf-Vlachy[1,2]

**Abstract**
The readability of scientific texts is critical for the successful distribution of research findings. I replicate a recent study which found that the abstracts of scientific articles in the life sciences became less readable over time. Specifically, I sample 28,345 abstracts from 17 of the leading journals in the field of management and organization over 3 decades, and study two established indicators of readability over time, namely the Flesch Reading Ease and the New Dale–Chall Readability Formula. I find a modest trend towards less readable abstracts, which leads to an increase in articles that are extremely hard to read from 12% in the first decade of the sample to 16% in the final decade of the sample. I further find that an increasing number of authors partially explains this trend, as do the use of scientific jargon and corresponding author affiliations with institutions in English-speaking countries. I discuss implications for authors, reviewers, and editors in the field of management.

**Keywords** Readability · Management research · Replication

**JEL Classification** M100

## 1 Introduction

The successful dissemination of research findings is critical to scientific and societal progress. Aside from grand issues like the public's waning trust in science as an institution (Haerlin and Parr 1999) and thus greater difficulty of convincing audiences of the accuracy of scientific findings, lie very pragmatic concerns.

---

---

✉ Lorenz Graf-Vlachy
lorenz.graf-vlachy@tu-dortmund.de

1   TU Dortmund University, Dortmund, Germany

2   ESCP Business School, Paris, France

⌀ Springer

In particular, it is essential that scientific texts, such as scientific articles and their abstracts, be comprehensible for fellow scientists (Loveland et al. 1973), and, ideally, interested laypeople (Scharrer et al. 2013). Prior research has therefore studied the readability of the abstracts of scientific papers. In particular, Plavén-Sigray et al. (2017) recently identified a substantial downward trend in readability across a broad range of scientific disciplines. This finding has already sparked further studies, like similar work in other fields (Yeung et al. 2018), as well as research into how authors can craft abstracts that are more readable (Freeling et al. 2019) and how the reporting of scientific results can be improved in general (Hanel and Mehler 2019).

However, Plavén-Sigray et al.'s (2017) study has several shortcomings from the perspective of a management and organization scholar. First and foremost, it is completely unclear whether their results also apply to management and organization research. For one, their study focused mostly on life sciences and thus does not necessarily generalize readily to other fields. For another, even within their own study, they found differences in the trends of different scientific disciplines, suggesting heterogeneity between fields. This makes it even less clear whether their results hold in our field as well. Given that publications in management and organization research, as the output of an applied science, should be particularly accessible to non-scientists, such a trend in our field would be very disturbing.

Second, while the results of Plavén-Sigray et al.'s (2017) are interesting and their methods sound, there are also opportunities for improvement from theoretical and methodological points of view. The authors, for example, formally explore only one potential antecedent of decreasing readability, and merely discuss but do not formally test a second explanation. Further, they rely on only two related readability measures and do not offer robustness checks regarding alternative measures.

To address all these issues, I replicate and extend the study of Plavén-Sigray et al. (2017) using a corpus of scientific texts from the management and organization literature, aiming not to assess the direct reproducibility or their research (Begley and Ioannidis 2015) but to enhance the generalizability (Block and Kuckertz 2018) of the original study. In addition, I formally test additional explanatory variables and perform a comprehensive array of robustness checks.

With this article, I make two specific contributions. First, I contribute to the general literature on the readability of scientific texts (Freeling et al. 2019; Plavén-Sigray et al. 2017). I demonstrate that the downward trend in readability that was observed in a variety of scientific disciplines also exists in the field of management and organization research, that it is robust, and that key findings from an important study replicate in another context. I confirm that the trend towards less readable abstracts is not only reliably associated with an increasing number of co-authors, but I also introduce the affiliation of authors with an institution in an English-speaking country as a novel predictor of readability.

Second, I contribute to and reinvigorate the specific meta-scientific debate in management and organization research on the accessibility of its body of scientific work (Loveland et al. 1973). In particular, I demonstrate that a part of the downward trend is related to an increase in the use of management-specific scientific jargon.

The remainder of this paper is structured as follows: Sect. 2 introduces the original study and briefly reports its method, results, and key findings. Section 3

then describes the overall methodological approach and derives the hypotheses to be tested. Section 4 introduces the method I employ in the replication, including information on sample and analyses. Section 5 reports the results of the replication, and Sect. 6 reports a host of robustness checks. Finally, Sect. 7 discusses the findings, compares them to those of the original study, highlights limitations, and develops implications for future research.

## 2 Original study

Plavén-Sigray et al. (2017) studied the readability of a large sample of article abstracts from a variety of disciplines in the life sciences. Their corpus was created from a semi-automatic partial download of the *PubMed* index and contained 709,577 abstracts from 123 highly cited journals between 1880 and 1995. None of the journals pertained to the field of management and organization.

They preprocessed the abstracts to clean them from distorting text and then calculated readability according to two widely used and accepted measures, the Flesch Reading Ease (FRE; Flesch 1948) and the New Dale–Chall Readability Formula (NDC; Chall and Dale 1995; Kincaid et al. 1975). The measures assess readability in slightly different ways, but they are both fundamentally determined by the length of sentences and the difficulty of the words used in the focal texts. Lower readability is indicated by lower FRE and higher NDC. Both values were correlated at $r = -0.72$ ($p < 0.001$).

The authors of the original study analyzed the readability of abstracts over time using mixed effects models. For both measures, the fixed effect of the year of publication was significantly related to readability. Overall, they thus found support for a trend towards lower readability of abstracts over time. This trend is present (although in different magnitudes) across all sampled disciplines. Only two journals exhibited an increasing FRE over time.

To explain the observed trend, Plavén-Sigray et al. (2017) put forward two potential explanations. First, they assessed whether the decrease in readability was driven by an increase in the number of co-authors over time and found that more authors were indeed associated with reduced readability, but that the time trend nevertheless remained significant when controlling for the number of authors. They thus rejected this explanation. Second, they studied whether an overall increase in the use of scientific vocabulary drove the reduction in readability. They manually derived word lists that represent common scientific words and general scientific jargon. Both types of words were found to become more frequent over time, suggesting that this explains the downward trend in readability. Notably, however, they did not test this notion in a regression analysis.

As an indication of the generalizability of their findings, Plavén-Sigray et al. (2017) performed a supplementary analysis to assess whether the readability of abstracts correlated with the readability of the corresponding articles. In a

sample of six journals, they found substantial significant correlations of $r = 0.60$ and $r = 0.63$ for FRE and NDC, respectively.

## 3 Replication approach and hypotheses

Replications are an existential element of scientific progress (Jasny et al. 2011; Tsang and Kwan 1999) because they can, among other things, allow to see if the results of individual studies hold in similar settings and whether results generalize, for example to other points in time, empirical contexts, or operationalizations of key variables. Unfortunately, some scientific fields like psychology currently find themselves in the midst of what some call a "replicability crisis" (Pashler and Harris 2012), in which long-held truths suddenly appear much less certain because they fail to replicate. While the current perception in the field of management and organization research is less dire, replication studies are also increasingly explicitly called for by management and organization scholars (Anderson et al. 2019; Bettis et al. 2016; Evanschitzky et al. 2007; Hubbard et al. 1998; Tsang and Kwan 1999) to improve or weed out deficient theories (Tierney et al. 2020).

The central objective of this article is to replicate the key analysis of Plavén-Sigray et al. (2017). I do this first by employing a different sample of abstracts—from management and organization articles—but using the same measures as the original study. Thus, my replication initially constitutes what Lindsay and Ehrenberg (1993) call a "differentiated replication" or what Tsang and Kwan (1999) would refer to as an "empirical generalization." In addition, however, I extend the ideas of Plavén-Sigray et al. (2017) to further identify causes of a potential trend in readability and perform various robustness checks, adding elements of what Tsang and Kwan (1999) term a "generalization and extension."

Specifically, I put forward and test several hypotheses in this article. The first key idea of Plavén-Sigray et al. (2017) is that there exists a trend towards lower readability in scientific abstracts. As the intention of these authors is exploratory, they do not propose any theoretical rationale for this notion. One could, however, speculate that it might be driven by an increased stratification of the field and a subsequent increase in the complexity of specialized vocabulary or by researchers feeling increasingly pressured to sound more sophisticated to "sell" their results (Vinkers et al. 2015). Plavén-Sigray et al. (2017) found empirical support for such a downward trend. The first hypothesis to be tested is thus:

**Hypothesis 1 (H1)** There exists a downward trend in the readability of abstracts over time.

Plavén-Sigray et al. (2017) further speculate that such a trend might also be partially caused by an increase in the average number of authors per paper (Drenth 1998; Epstein 1993). The key idea behind this conjecture is that larger author teams might lead to "too many cooks spoiling the broth" (Kelly 2014). Specifically, it is conceivable that when more people work together, it becomes increasingly difficult

to accommodate everybody's editing suggestions without making text harder to read. Especially, authors' concerns about being right that might lead to adding what one personally deems necessary might combine with concerns about being liked that limit criticism of others' opaque writing (Insko et al. 1985). A constant addition of material over multiple rounds of reading and writing within an author team might thus lead to decreased readability. Alternatively, an increase in the number of authors might lead to a diffusion of responsibility (Beyer et al. 2017), ultimately making the outcome worse than if a single person would be in charge. Plavén-Sigray et al. (2017) found support for a negative influence of the number of authors on abstract readability. I correspondingly propose:

**Hypothesis 2a (H2a)** The number of authors is negatively associated with the readability of abstracts.

It is worth reiterating that Plavén-Sigray et al. (2017) find that while the number of authors matters, the time trend remained significant. Analogously, I thus also hypothesize:

**Hypothesis 2b (H2b)** There exists a downward trend in the readability of abstracts over time even after controlling for the number of authors.

Plavén-Sigray et al. (2017) further propose that an increase in science-specific jargon might explain the time trend. While Plavén-Sigray et al. (2017) claim that there exists "evidence in favor of the hypothesis that there is an increase in general scientific jargon which partially accounts for the decreasing readability" (p. 4), they do not formally test this idea in a regression model. Borrowing from Plavén-Sigray et al.'s (2017) logic, but moving beyond the idea of a pure replication study, I explicitly formulate the corresponding hypothesis:

**Hypothesis 3 (H3)** The share of management-specific scientific jargon is negatively associated with the readability of abstracts.

Finally, Plavén-Sigray et al. (2017) are silent on the role of the origin of authors. Going clearly beyond a pure replication, I propose that there is one additional potential antecedent of readability that has thus far not received attention in the literature. This antecedent is the fact whether the authors of a paper are from an English-speaking environment.[1] Literature on English as a second language indicates that texts produced by authors with different levels of English proficiency may differ along such dimensions as the number of words per clause or the use of complex nominals (Lu and Ai 2015). If a papers' authors are based in a country in which English is a national language, this might allow them to make use of a broader and more complex vocabulary, and they may be able and motivated to express complex

---

[1] I would like to acknowledge an anonymous reviewer who raised an important question that gave rise to this idea.

ideas in fewer sentences. Both tendencies would likely lead to reduced readability. I formally hypothesize:

**Hypothesis 4 (H4)** Authors' affiliation with institutions in English-speaking countries is negatively associated with the readability of abstracts.

## 4 Method

In the following, I explain the sample of abstracts I used in this study, the way I preprocessed the abstract text, and the calculation of the readability measures. Finally, I describe the performed econometric analyses.

### 4.1 Replication sample

For my replication, I selected some of the top journals in the field of management and organization. First, I retained all management and organization journals from the journal list of the University of Texas at Dallas, which is commonly used in tenure decisions in the United States. I then complemented this list of journals with any remaining management and organization journals that are listed on the Financial Times 50 list, which is commonly used to assess research performance in European schools. Finally, I removed any journals for which there were less than ten years of data available, and I removed any non-peer-reviewed journals. In sum, my sample of journals included 17 publications, specifically *Academy of Management Journal*, *Academy of Management Review*, *Administrative Science Quarterly*, *Entrepreneurship Theory and Practice*, *Journal of Applied Psychology*, *Journal of Business Ethics*, *Journal of Business Venturing*, *Journal of International Business Studies*, *Journal of Management*, *Journal of Management Studies*, *Management Science*, *Organization Science*, *Organization Studies*, *Organizational Behavior and Human Decision Processes*, *Research Policy*, *Strategic Entrepreneurship Journal*, and *Strategic Management Journal*.

I obtained article data from *Web of Science* in February 2019. I manually downloaded information about each article, including author names, publication time, and the full text of the abstract. The initial sample consisted of 44,858 records from *Web of Science*. After removing various non-article records (e.g., book reviews, editorial material, etc.), I retained records for 35,391 articles. All articles before 1990 were missing abstracts. Removing them further decreased the sample to 28,904 articles. In addition, I required every year in the sample to be represented with at least 100 articles to prevent outliers from distorting estimations. This led to the exclusion of the year 1990, leaving a final sample of 28,874 abstracts. While this sample is substantially smaller than that of Plavén-Sigray et al. (2017), it is still a considerable sample which should allow for the observation of trends if they are of a meaningful magnitude. The lowest number of
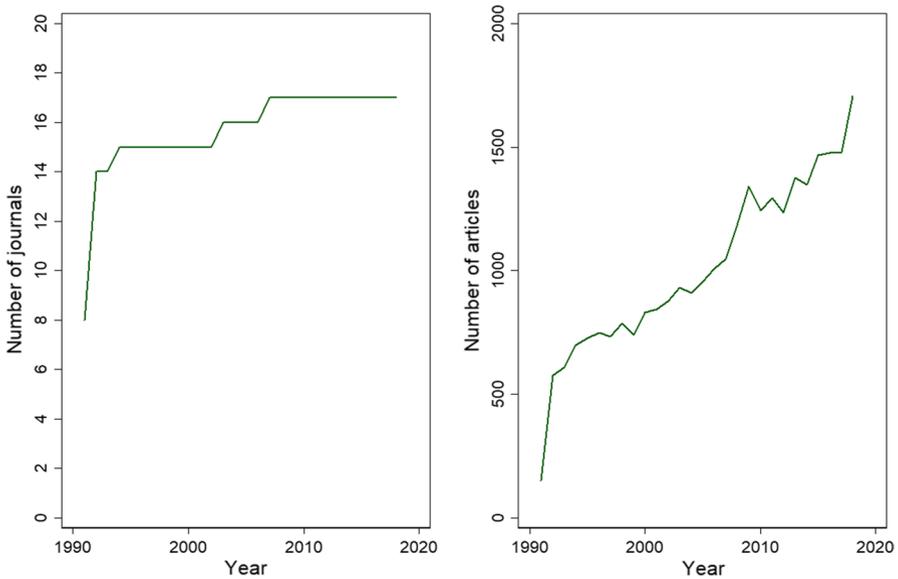
**Fig. 1** Number of journals and articles per year over time

articles per journal was 224 at *Strategic Entrepreneurship Journal*. This exceeds the criterion of 100 articles applied by Plavén-Sigray et al. (2017).

## 4.2 Text preprocessing

Plavén-Sigray et al. (2017) pre-processed their abstracts in a fairly specific fashion. I diverged from their procedure in several ways because most problems they address through preprocessing stem from the fact that they work with abstracts from scientific disciplines largely outside the social sciences. This required, for example, the removal of nucleic acid sequences or periods arising from binomial nomenclature. Such cleaning procedures appeared unnecessary in my sample.

Instead, I followed the spirit of the original study and reviewed 100 randomly selected abstracts from my sample to identify any problems that might be specific to my sample. I identified several publisher copyright statements and included their removal in my preprocessing steps. Further preprocessing measures I conducted included the removal of periods from abbreviations (such as "U.S.") which could confound the identification of sentences. I further removed all enumerators like (1), (2) or (a), (b), which were common in the abstracts. I also expanded some common abbreviations such as "vs." and I removed all percentages and numbers and replaced all hyphens with blank spaces. Finally, all text in square brackets and parentheses was removed as some journals included references to other papers in this format. Preprocessing was performed using a custom-written Perl program.[2] Since

---

[2] The corresponding program can be found in the online appendix. The use of Perl in this article is somewhat ironic as Perl code has the reputation of being particularly hard to read. Some programmers half-jokingly refer to Perl as a "write-only language" (Cozens 2000).

the number of sentences is critical to the calculation of the readability measures described below, I dropped all abstracts that contained fewer than three sentences. Inspection of such cases showed that there was usually clearly missing punctuation in these abstracts, likely due to data entry errors in *Web of Science*. Removing such cases reduced the sample by 529 to 28,345 abstracts. Figure 1 shows the distribution of journals and articles in the final sample over time.

### 4.3 Readability calculations and other variable operationalizations

I assessed readability primarily using the same two indicators that were used in the original study, i.e., the Flesch Reading Ease (FRE) and the New Dale–Chall Readability Formula (NDC). FRE was calculated as follows (Flesch 1948):

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

NDC was calculated according to this formula (Chall and Dale 1995; Kincaid et al. 1975):

$$NDC = \begin{cases} 0.1579\left(\frac{difficult}{words} * 100\right) + 0.0496\left(\frac{words}{sentences}\right) + 3.6365 \ if\left(\frac{difficult}{words}\right) > 5\% \\ \\ 0.1579\left(\frac{difficult}{words} * 100\right) + 0.0496\left(\frac{words}{sentences}\right) if\left(\frac{difficult}{words}\right) \leq 5\% \end{cases}$$

Both formulas contain various constants to calibrate their results and make them easier to interpret, for instance in terms of audience education level. For example, scores between 30 and 0 on the FRE are considered "very difficult" and appropriate only for college-educated audiences (Flesch 1948).

Counting the number of words was performed in a straightforward fashion. As contractions (e.g., "I've") indicate two words, they were counted as such in the calculation of FRE. Since the NDC word list, however, contained expressions like "here's," I had to adjust the word count logic for the NDC calculation. Given that contractions were very rare in the scientific abstracts of my sample, the correlation between the two word counts was almost perfect at $r = 0.99$ ($p < 0.001$). Thus, this should not affect overall results. Sentences were identified by splitting the text at periods, exclamation points, question marks, as well as semicolons. Sentence identification was performed after preprocessing. Counting of syllables was performed using the Perl module *Lingua::EN::Syllable*. The number of *difficult* words was determined as the number of words in an abstract that were not on the NDC list of common words. I employed the exact same 2949-word-list as did Plavén-Sigray et al. (2017), which was obtained from the Python package *textstat*. All counts and computations were performed using custom-written Perl programs.

The number of authors was computed by simply counting the number of individual authors listed in each paper's author string as obtained from *Web of Science*.

I measured management-specific scientific jargon following Plavén-Sigray et al.'s (2017) method. Initially, I developed a list of management-science-specific common

words. To do so, I first obtained the frequencies of all words used in any of the abstracts and retained the most frequent words which were not simultaneously part of the NDC common word list. Plavén-Sigray et al. (2017) only obtained word frequencies of a random subsample of their abstracts, but it is not clear why this would be preferable to an exhaustive analysis, so I diverged here. Plavén-Sigray et al. (2017) retained 2949 words (as many as are on the list of NDC common words). I retained 2953 words as several words at the very end of the ranking had the same frequency and there was no immediately apparent logic to decide which ones to exclude. Plavén-Sigray et al. (2017) further used multiple raters to distill their list of science-specific common words into a list of general scientific jargon. I refrained from developing such a list from my sample as the process in the original study is not very well specified. The rules for decisions on word inclusion/exclusion are not completely clear and initial discussions with a fellow researcher immediately triggered substantial disagreements about several words. In fact, the original study does not report interrater reliability and admits that raters were not fully independent. Finally, I divided the number of management-science-specific common words by the number of total words for each abstract to arrive at the measure for management-specific scientific jargon.

To identify whether an author team was from an English-speaking country, I relied on the institutional information of the corresponding author as listed in *Web of Science*. While it would have been preferable to identify the institutional affiliations of all authors, the corresponding data was unfortunately not available. For each article, I extracted the country and compared it to a list of countries in which English is a *de jure* or *de facto* national language.

### 4.4 Econometric approach and implementation

Precisely following Plavén-Sigray et al. (2017), I employed linear mixed effects models to estimate the effect of time on readability, while accounting for the hierarchical data structure, i.e., the fact that multiple abstracts belong to the same journal while different journals span different year ranges. Specifically, I first estimated three models for each readability measure to test H1. The first model was a null model in which the readability measure was predicted only by the journal as a random effect with varying intercepts. The second model added a fixed effect of publication year. The third model additionally allowed for varying slopes for the random effect of the journal. To replicate the analysis on the role of the number of co-authors and test H2a and H2b, I estimated an additional model for each readability measure. These models were identical to the fully specified models described above but additionally include the number of authors of each paper as a second fixed effect. To test H3, I specified additional models that include the share of scientific jargon words instead of the number of authors. Finally, H4 is tested using fully specified models that include both the number of authors and a dummy variable indicating whether the corresponding author came from an English-speaking country. Note that these models do intentionally not include the share of scientific jargon words as an additional control because this variable, while potentially informative with regard to the

**Table 1** Correlation matrix

|  | FRE | NDC | Publication year | Number of authors | Scientific jargon | English-speaking country affiliation |
|---|---|---|---|---|---|---|
| FRE | 1.000 | | | | | |
| NDC | − 0.687* | 1.000 | | | | |
| Publication year | − 0.097* | 0.122* | 1.000 | | | |
| Number of authors | − 0.067* | 0.096* | 0.187* | 1.000 | | |
| Scientific jargon | − 0.538* | 0.771* | 0.064* | 0.117* | 1.000 | |
| English-speaking country affiliation | − 0.007 | 0.011* | − 0.190* | − 0.013* | − 0.014* | 1.000 |

*$p < 0.05$. FRE = Flesch Reading Ease, NDC = New Dale–Chall Readability Formula

readability of the abstracts, is likely not a causal antecedent. This is because it manifests in the same text the readability of which is the dependent variable. All other antecedents, in contrast, exist independently and temporally prior to the focal text.

All estimations were performed in Stata 16.1 using the *mixed* command with the exact same econometric specifications Plavén-Sigray et al. (2017) used. I ran several tests using R 3.6.1 and the *lme4* package to ensure that the results were identical between the software packages.

# 5 Results

Table 1 shows a correlation matrix of all relevant variables. Most notably, I observe a correlation of $r = − 0.69$ ($p < 0.001$) between the FRE and DCF scores, which is similar to the correlation of $r = − 0.72$ ($p < 0.001$) observed by Plavén-Sigray et al. (2017), giving further credibility to my implementation of readability score calculations.

Figure 2 shows the average FRE and average NDC values per year across all journals in the sample. This visualization—as well as the significant correlations observable in Table 1—suggests a downward trend in FRE and an upward trend in NDC, both implying a decrease in readability over time.

This appears to hold for most individual journals. Figure 3 shows FRE by journal over time, and Fig. 4 shows NDC by journal over time. As is evident from the trend lines included in the figures, most journals exhibit a mild trend towards reduced readability, although some journals certainly show a stronger trend than others. Exceptionally, the *Academy of Management Journal* exhibits an increase in mean readability according to FRE (although not NDC) and the *Strategic Entrepreneurship Journal* shows a mild increase in readability according to NDC (although not FRE).

Table 2 depicts the results of the linear mixed effect models regarding FRE, and Table 3 shows the same results for NDC. The models M0 are the null models. The models M1 add a fixed effect for time (by including the publication year of
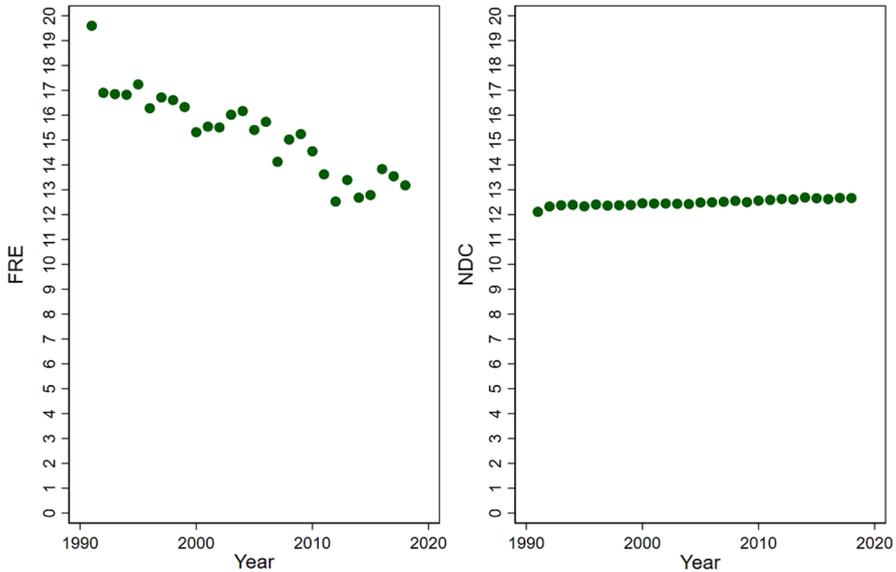
**Fig. 2** Mean Flesch Reading Ease (FRE) and mean New Dale–Chall (NDC) Readability Formula over time

each abstract), and the models M2 add varying slopes for the journal. For both FRE and NDC, it is evident from the differences in AIC and BIC that models M2 provide the best fit for the data (Burnham and Anderson 2004; Kuha 2004). In all models that include the publication year of an article as a predictor, it is highly significantly related to readability ($p < 0.001$). For FRE, the coefficients of year are consistently negative, indicating reduced reading ease, and for NDC, the coefficients of year are consistently positive, indicating increasing difficulty. This means that there is indeed a trend over time towards less readable abstracts, supporting H1.

The two readability measures consist of three components, i.e., the number of syllables per word (FRE), the number of words per sentence (FRE and NDC), and the share of difficult words (NDC). A natural question is thus which of the components changed to drive the overall decrease in readability. As is evident from Fig. 5, all three components increased over time, negatively affecting readability. These findings are consistent with the results of Plavén-Sigray et al. (2017), including the increase in words per sentence, which is observable in the original study's data after the 1960s.

As the average number of co-authors increased over time in their sample, Plavén-Sigray et al. (2017) considered this as a possible explanation for the observed time trend. In my data, the average number of co-authors similarly increased by about a third during the sample timeframe, which comprises close to 3 decades (see Fig. 6).

To formally test the role of the number of authors, I added the number of authors as a fixed effect in the models M3 shown in Tables 2 and 3. As hypothesized, the coefficient of the number of authors was significant and negative for FRE ($p < 0.001$)
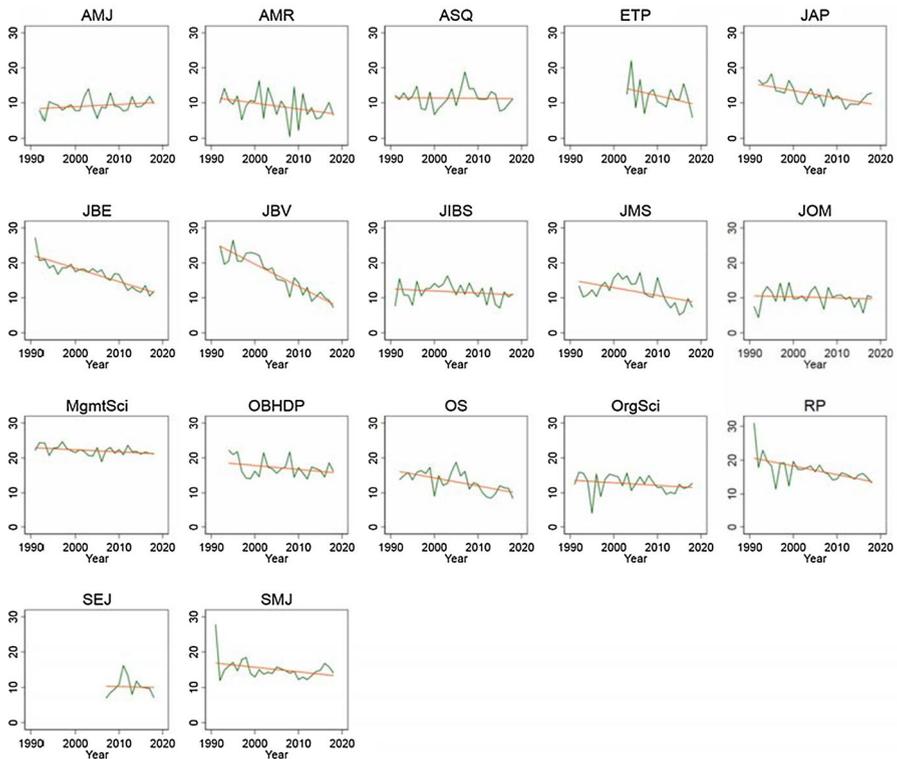
**Fig. 3** Mean Flesch Reading Ease (FRE) over time by journal. *Notes*: Green line represents readability, red line represents trend. AMJ, Academy of Management Journal; AMR, Academy of Management Review; ASQ, Administrative Science Quarterly; ETP, Entrepreneurship Theory & Practice; JAP, Journal of Applied Psychology; JBE, Journal of Business Ethics; JIBS, Journal of International Business Studies; JMS, Journal of Management Studies; JOM, Journal of Management; MgmtSci, Management Science; OBHDP, Organizational Behavior and Human Decision Processes; OS, Organization Studies; OrgSci, Organization Science; RP, Research Policy; SEJ, Strategic Entrepreneurship Journal; SMJ, Strategic Management Journal

and significant ($p < 0.001$) and positive for NDC. This shows that the number of authors indeed has a negative relationship with readability, indicating support for H2a.

At the same time, the results from the initial models remained stable when adding the number of authors. The coefficient of publication year remained significantly negative ($p < 0.001$) for FRE (Table 2) and significantly positive ($p < 0.001$) for NDC (Table 3). This shows that the time trend was not exclusively driven by the increase in mean authors per paper and thus provides support for H2b as well.

Figure 7 reports the changes of the shares of management-specific scientific jargon over time. In addition, to allow comparisons with Plavén-Sigray et al. (2017), it also shows NDC common words over time. Similar to what was found in the original study, the use of science-specific vocabulary increases, whereas the use of NDC common words decreases.
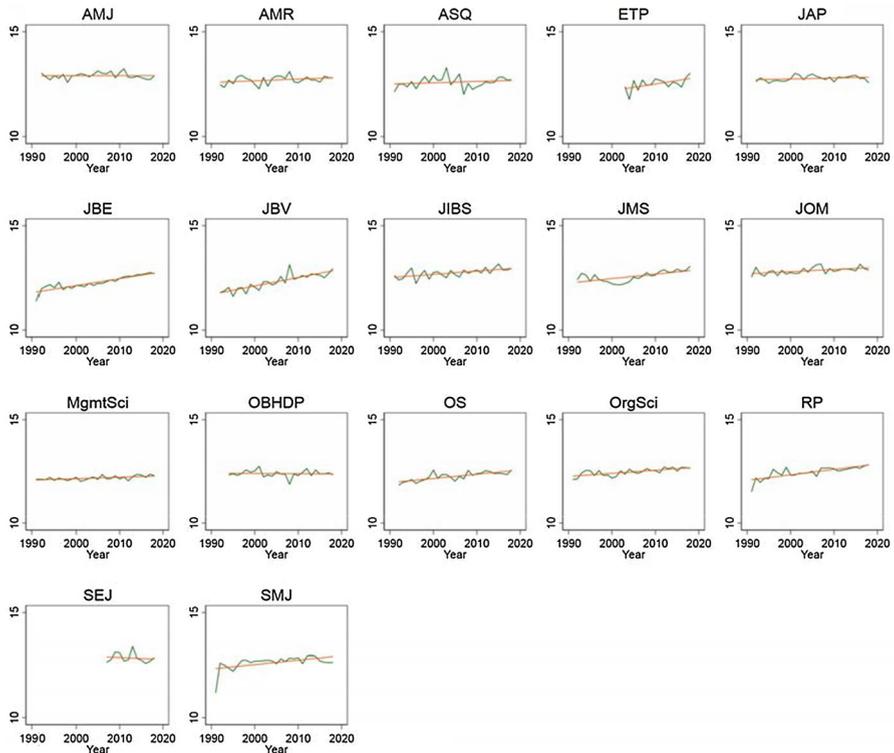
**Fig. 4** Mean New Dale–Chall (NDC) Readability Formula over time by journal. *Notes*: Green line represents mean readability, red line represents trend. Y-axes are truncated and do not include zero. AMJ, Academy of Management Journal; AMR, Academy of Management Review; ASQ, Administrative Science Quarterly; ETP, Entrepreneurship Theory & Practice; JAP, Journal of Applied Psychology; JBE, Journal of Business Ethics; JIBS, Journal of International Business Studies; JMS, Journal of Management Studies; JOM, Journal of Management; MgmtSci, Management Science; OBHDP, Organizational Behavior and Human Decision Processes; OS, Organization Studies; OrgSci, Organization Science; RP, Research Policy; SEJ, Strategic Entrepreneurship Journal; SMJ, Strategic Management Journal

While this already is suggestive evidence for H3, I formally tested it in the models M4 in Tables 2 and 3. Again, the coefficients for management-specific scientific jargon were significantly negative ($p < 0.01$) for FRE (Table 2) and significantly positive ($p < 0.001$) for NDC (Table 3). Both thus suggest that an increase in jargon decreases readability, and thus support H3.

Finally, H4 proposed that corresponding authors with an institutional affiliation in an English-speaking country would also have a detrimental effect on readability. Models M5 in Tables 2 and 3 test this hypothesis. Due to limited data availability on corresponding author addresses in *Web of Science*, the sample was reduced to 27,924 abstracts. As is evident from the tables, the relevant coefficients were significantly negative ($p < 0.002$) for FRE (Table 2) and significantly positive ($p < 0.001$) for NDC (Table 3). These findings provide support for H4.

**Table 2** Linear mixed effect models estimating readability as measured by FRE

|  | M0 | M1 | M2 | M3 | M4 | M5 |
|---|---|---|---|---|---|---|
| Publication year | | −0.173*** (0.01) | −0.164*** (0.04) | −0.151*** (0.04) | −0.137*** (0.02) | −0.161*** (0.04) |
| Number of authors | | | | −0.377*** (0.06) | | −0.370*** (0.06) |
| Scientific jargon | | | | | −105.277*** (1.03) | |
| English-speaking country affiliation | | | | | | −0.596** (0.20) |
| Constant | 13.160*** (0.80) | 360.682*** (21.13) | 342.804*** (79.31) | 317.380*** (78.76) | 333.177*** (47.04) | 337.324*** (79.53) |
| N | 28,345 | 28,345 | 28,345 | 28,345 | 28,345 | 27,924 |
| AIC | 227,733.4 | 227,465.9 | 227,314.0 | 227,280.6 | 218,470.2 | 223,822.9 |
| BIC | 227,758.1 | 227,498.9 | 227,363.6 | 227,338.4 | 218,527.9 | 223,888.8 |

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Standard errors in parentheses

For simplicity, only the fixed effects components of the models are reported

**Table 3** Linear mixed effect models estimating readability as measured by NDC

|  | M0 | M1 | M2 | M3 | M4 | M5 |
|---|---|---|---|---|---|---|
| Publication year | | 0.015*** (0.00) | 0.014*** (0.00) | 0.013*** (0.00) | 0.012*** (0.00) | 0.013*** (0.00) |
| Number of authors | | | | 0.032*** (0.00) | | 0.031*** (0.00) |
| Scientific jargon | | | | | 10.052*** (0.05) | |
| English-speaking country affiliation | | | | | | 0.060*** (0.01) |
| Constant | 12.593*** (0.05) | −17.564*** (1.38) | −14.927** (5.74) | −12.805* (5.70) | −14.802*** (3.63) | −14.477* (5.74) |
| N | 28,345 | 28,345 | 28,345 | 28,345 | 28,345 | 27,924 |
| AIC | 73,184.0 | 72,710.8 | 72,472.0 | 72,415.6 | 48,099.7 | 71,216.3 |
| BIC | 73,208.7 | 72,743.8 | 72,521.5 | 72,473.3 | 48,157.5 | 71,282.1 |

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Standard errors in parentheses

For simplicity, only the fixed effects components of the models are reported
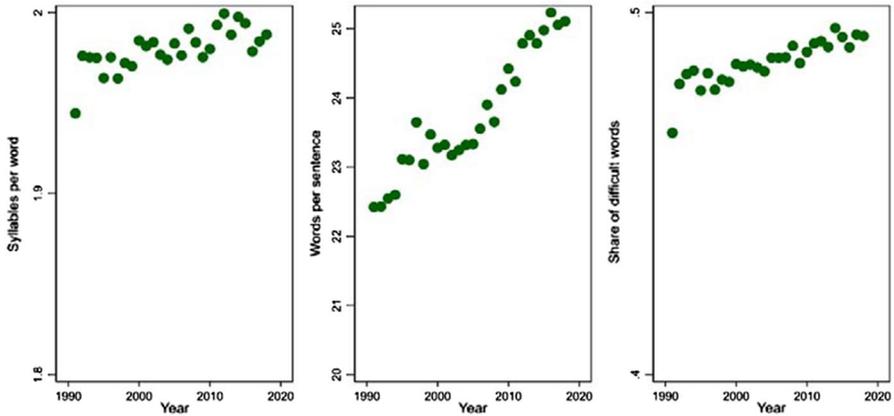
**Fig. 5** Mean values of readability measure components over time. *Notes*: Y-axes are truncated and do not include zero
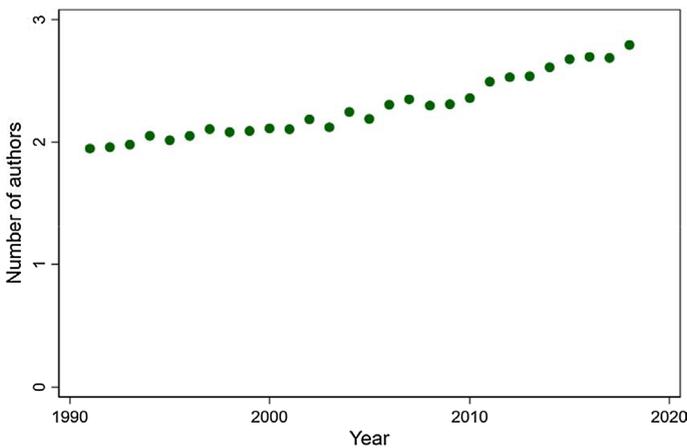


**Fig. 6** Mean number of authors per article over time

## 6 Robustness checks

I performed various checks to ensure the robustness of my findings. First, I obtained FRE and NDC scores for each abstract from readable.com, a commercial provider of readability scores. My FRE scores and those from readable.com correlated at $r = 0.98$ ($p < 0.001$), and NDC scores correlated at $r = 0.93$ ($p < 0.001$). The observed minor differences are likely due to potential additional preprocessing performed by readable.com or due to slightly different treatment of word boundaries, syllable counting, or similar steps in the calculations. I repeated all analyses using the scores from readable.com and obtained fully consistent results
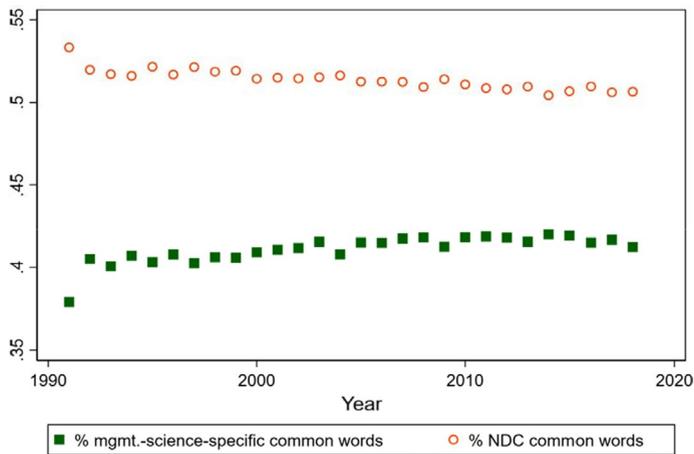
**Fig. 7** Shares of management-specific scientific jargon and NDC common words over time. *Notes*: Y-axis is truncated and does not include zero

in all estimations. Overall, this gives me confidence that the results are not driven by my implementation of the readability measures.

Second, while Plavén-Sigray et al. (2017) only relied on FRE and NDC, I employed four additional readability measures. Specifically, I implemented the Flesch-Kincaid Grade Level (Kincaid et al. 1975) measure, the Automated Readability Index (ARI) (Senter and Smith 1967), the Gunning Fog Index (Gunning 1969), and the Simple Measure of Gobbledygook (SMOG; McLaughlin 1969). When re-running the models with these alternative readability measures, results were almost totally consistent. The sole exception was that the coefficient of publication year failed to reach significance ($p > 0.05$) in M3 for the ARI measure. It remained, however, highly significant ($p < 0.001$) in all other model specifications. Together, these robustness checks provide substantial evidence that the observed findings are not only a function of the specific readability measures chosen.

Third, going beyond the general question of whether authors are based in an English-speaking country, one might wonder if the results hold controlling for country of origin more broadly. I thus ran models that included a set of dummy variables encoding the country of the corresponding author's institution in addition to the number of authors. Country dummies were—as expected—partly significant, and both year of publication and number of authors retained their significant influence on readability. This demonstrates that the results regarding publication year and the number of authors are robust to the inclusion of further controls. I naturally did not run a model that additionally included the dummy indicating the corresponding author's affiliation with an institution in an English-speaking country because the explanatory power of such a dummy is negligible if there are already dummies included for each individual country.

Fourth, mixed effects models are more prevalent in "micro" management and organization research (such as organizational behavior), whereas "macro" researchers (such as strategy scholars) frequently employ other types of models that account for hierarchical data structures. As an additional robustness check, I therefore specified linear regression models with a fixed effect for the journal and with standard errors clustered at the journal, using Stata's *areg* command. The insignificant result of a Hausman test confirmed that it was appropriate to use fixed effects models. The results I obtained were again fully consistent with those from the mixed models for all specifications and all readability measures.

Finally, as is evident in some of the figures shown above, the year 1991 appears to be an outlier year, I also ran all models excluding this year (omitting 151 abstracts). Again, I obtained consistent results.

# 7 Discussion

## 7.1 General discussion and contributions

Management and organization scholars lamented a decreasing readability of their journals already half a century ago (Loveland et al. 1973). Disturbingly, my findings suggest that this trend continues to this day. The absolute magnitude of the trend is admittedly not very large on average. In the fixed effects models, for example, one decade is associated with a decrease in readability by about 1.73 on the FRE measure, while most of Flesch's (1948) categories of readability span a range of 10. Nevertheless, it is quite consistent, and by no means inconsequential. For example, a FRE score below zero indicates a text that one can no longer expect a college graduate to understand (Flesch 1948; Kincaid et al. 1975). In my sample of articles until 2000, such a low FRE score occurred for 12 percent of abstracts. This share increased to 16 percent in the papers published after 2010.

This article makes two key contributions. First, it contributes to the general literature on the readability of scientific texts (Freeling et al. 2019; Plavén-Sigray et al. 2017). It shows that the downward trend in readability observed in a plethora of scientific disciplines also exists in the field of management and organization research. This trend is limited in magnitude but very robust. The article also shows that other key findings of Plavén-Sigray et al. (2017) replicate in a different context. Specifically, it shows that the trend towards less readable abstracts is associated with an increasing number of co-authors. Further, it demonstrates that, possibly somewhat counterintuitively, the affiliation of authors with an institution in an English-speaking country also reduces readability.

Second, this paper revives the old debate specifically in the field of management and organization research on the accessibility of its body of scientific work (Loveland et al. 1973). Specifically, the article demonstrates that a part of the decrease in readability is related to an increased use of management-specific scientific jargon.

## 7.2 Practical implications for management and organization researchers

The observed decrease in readability of scholarly work in the field of management and organization research is problematic because it makes research results less accessible to a host of important constituents. This includes fellow researchers (who also suffer from comprehension problems if articles are hard to read and who might cite articles more frequently if their abstracts are easy to read; Freeling et al. 2019; Hartley 1994), practitioners (whose likely already limited inclinations to consume scientific articles are probably further attenuated if they are hard to read), and journalists (who play a crucial role in relaying scientific findings to a broader audience; Bubela et al. 2009). The field thus has a vested interest in increasing readability.

Consequently, this article holds implications for all actors in management research. First and foremost, it is a call to authors to strive for greater readability. While a certain amount of complexity is inevitable in scientific writing (Knight 2003), authors can certainly pay more attention to readability when crafting abstracts (Hartley 1994). The use of free online tools to assess readability can be helpful in the process.[3] Authors might also find it advantageous to let friendly reviewers not only comment on the quality of an article's theory development or the appropriateness of chosen methods, but they may also explicitly solicit input on readability.

Second, this study has implications for reviewers, who greatly shape the articles that are ultimately published. Consequently, they might gently nudge (or resolutely push if required) authors towards simple expressions where appropriate. In particular, they may flag excessive use of jargon or overly long sentences.

Finally, journal editors as the ultimate arbiters have great influence on readability. On the one hand, they can use their journals' author guidelines to explicitly encourage authors to strive for readability, and the reviewer guidelines to sensitize reviewers for this matter. Of course, editors can also, and even more directly and emphatically than reviewers, enforce readability. On the other hand, editors can support authors in crafting readable abstracts. For instance, they could relax abstract word count maxima. This might, for example, allow authors to compose more easy-to-read sentences instead of fewer ones that are long and cumbersome to read. Editors can also introduce structured abstracts, which have been shown to help readability (Hartley 2003; Hartley and Benjamin 1998). Finally, they can follow the lead of the *Strategic Management Journal* (and other journals outside the field of management; Kuehne and Olden 2015) and offer "managerial summaries" in addition to traditional summaries. A quick comparison of the average readability (FRE) between the two abstract types (preprocessed) for the seven articles published in a recent issue (August 2019) revealed that whereas the traditional research abstracts yielded a score of 19.93, the managerial summaries achieved a substantially higher readability of 31.34. Anecdotal evidence from conversations with various management scholars even suggests that many read the managerial summaries first to decide if reading the regular abstract is even worth their time and effort.

---

[3] Such tools are available, for example, at https://www.readabler.com/ or https://app.readable.com/.

### 7.3 Limitations and further research

As every empirical research project, this article has limitations. Some of them are shared between the original study and this replication. These include, for example, the potential concern that the employed readability measures may not capture every facet of readability (Benjamin 2012). For instance, the length of causal chains influences readability (Otero et al. 2004) but is not assessed in either FRE or NDC. Also, readability measures and subjective assessments of readability may not necessarily concur (Griesinger and Klene 1984).

On a technical note, there are likely slight differences in preprocessing and potentially in the calculation of readability measures between the original study and this replication. Despite use of the exact same formulas, differences may arise, for example through differences in determining sentence, word, and syllable counts, all of which are nontrivial problems in natural language processing that have each attracted research interest in their own right (He and Kayaalp 2006). Nevertheless, such slightly different calculation procedures are very unlikely to produce substantially diverging results. The fact that the correlation between FRE and NDC as observed in my dataset and that of Plavén-Sigray et al. (2017) are highly similar, and that my measures correlate strongly with those obtained from a commercial service suggests that differences in measures are unlikely to distort results.

A further limitation is that at least a part of the observed decrease in readability over time may be driven by an actual increase in the complexity of the content of the journals. It might be argued that some journals in the sample became more "scientific" and less "practitioner-focused" over time. It would be interesting (and would likely enable provocative research) to design a measure for the scholarliness of a journal without relying on measures of mere text complexity or readability.

An additional limitation is that I could not re-create a meaningful analysis that would assess how strongly the readability of abstracts correlates with the readability of full articles. In a supplementary analysis, Plavén-Sigray et al. (2017) selected six open access journals to address this question for their fields of interest. The selected journals make their articles available in HTML format and could thus be comparably easily downloaded and analyzed. Unfortunately, there are not a sufficient number of open access journals in the field of management and organization research that (1) are of adequate quality to be comparable to the journals I analyzed and (2) make the full text of their articles available in a format that lends itself to easy processing. It is worth noting that Plavén-Sigray et al. (2017) do not consider this analysis to be the centerpiece of their study and caution the reader that their selection of journals is limited to relatively young ones, and thus not representative of their journal selection in their main analysis. Thus, the absence of a comparable analysis in my replication is likely not critical, but future researchers with access to a sufficiently large corpus of full-text articles might wish to explore the linkage between readability of abstracts and articles for the field of management and organization research.

Final potential limitations stem from limited data availability. While *Web of Science* contains data going back to 1980, abstracts appear to be available only from 1990 on. The observed time frame of about 3 decades is thus shorter than that of Plavén-Sigray et al. (2017). Nevertheless, it is a substantial time frame, and given

the changes the field of management and organization research underwent since the 1990s, I contend that my analysis is still meaningful. Still, it would of course be interesting to study whether the observed trends hold true even for articles published before the beginning of my sampling timeframe.

Future opportunities that go beyond the immediate limitations of this study abound as well. For instance, it could be interesting to see if an author's gender or an author team's gender composition has an influence on readability. Similarly, it would be fascinating if authors' academic seniority or the status of their institution has an influence on readability. Relatedly, various diversity aspects within an author team might help explain readability.

Finally, it is noteworthy that not all sampled journals exhibited a consistently negative trend in all readability measures. The *Academy of Management Journal*, for example, even showed a slight upward trajectory regarding the FRE measure. Future researchers might thus conduct in-depth analyses of the editorial teams or editorial policies of such exceptional cases.

In sum, this article suggests that there is an opportunity and a need for us, as management and organization scholars, to strive for improved readability in our field. Making abstracts easier to read is in the interest of authors who may enjoy greater reception of their work (Freeling et al. 2019), and it is also not against the interests of journals, as there is no adverse effect of readability on journal prestige (Hartley et al. 1988). Finally, it most certainly is in the interest of our readers. Let's keep it simple.

# References

Anderson BS, Wennberg K, McMullen JS (2019) Enhancing quantitative theory-testing entrepreneurship research. J Bus Ventur 34:105928

Begley CG, Ioannidis JPA (2015) Reproducibility in science: improving the standard for basic and pre-clinical research. Circ Res 116:116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Benjamin RG (2012) Reconstructing readability: recent developments and recommendations in the analysis of text difficulty. Educ Psychol Rev 24:63–88. https://doi.org/10.1007/s10648-011-9181-8

Bettis RA, Helfat CE, Shaver JM (2016) The necessity, logic, and forms of replication. Strat Manag J 37:2193–2203

Beyer F, Sidarus N, Bonicalzi S, Haggard P (2017) Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring. Soc Cogn Affect Neurosci 12:138–145

Block J, Kuckertz A (2018) Seven principles of effective replication studies: strengthening the evidence base of management research. Manag Rev Q 68:355–359. https://doi.org/10.1007/s11301-018-0149-3

Bubela T, Nisbet MC, Borchelt R, Brunger F, Critchley C, Einsiedel E, Geller G, Gupta A, Hampel J, Hyde-Lay R, Jandciu EW, Jones SA, Kolopack P, Lane S, Lougheed T, Nerlich B, Ogbogu U, O'Riordan K, Ouellette C, Spear M, Strauss S, Thavaratnam T, Willemse L, Caulfield T (2009) Science communication reconsidered. Nat Biotechnol 27:514–518. https://doi.org/10.1038/nbt0609-514

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304. https://doi.org/10.1177/0049124104268644

Chall JS, Dale E (1995) Readability revisited: the New Dale–Chall readability formula. Brookline Books, Cambridge

Cozens S (2000) Ten Perl Myths. https://www.perl.com/pub/2000/01/10PerlMyths.html. Accessed 29 July 2019

Drenth JP (1998) Multiple authorship: the contribution of senior authors. J Am Med Assoc 280:219–221. https://doi.org/10.1001/jama.280.3.219

Epstein RJ (1993) Six authors in search of a citation: villains or victims of the Vancouver convention? BMJ 306:765–767. https://doi.org/10.1136/bmj.306.6880.765

Evanschitzky H, Baumgarth C, Hubbard R, Armstrong S (2007) Replication research's disturbing trend. J Bus Res 60:411–415

Flesch R (1948) A new readability yardstick. J Appl Psychol 32:221–233. https://doi.org/10.1037/h0057532

Freeling B, Doubleday ZA, Connell SD (2019) Opinion: How can we boost the impact of publications? Try better writing. Proc Natl Acad Sci USA 116:341–343. https://doi.org/10.1073/pnas.1819937116

Griesinger WS, Klene RR (1984) Readability of introductory psychology textbooks: flesch versus student ratings. Teach Psychol 11:90–91. https://doi.org/10.1207/s15328023top1102_8

Gunning R (1969) The fog index after twenty years. J Bus Commun 6:3–13

Haerlin B, Parr D (1999) How to restore public trust in science. Nature 400:499. https://doi.org/10.1038/22867

Hanel PH, Mehler DM (2019) Beyond reporting statistical significance: identifying informative effect sizes to improve scientific communication. Public Underst Sci 28:468–485. https://doi.org/10.1177/0963662519834193

Hartley J, Trueman M, Meadows AJ (1988) Readability and prestige in scientific journals. J Inf Sci 14:69–75. https://doi.org/10.1177/016555158801400202

Hartley J (1994) Three ways to improve the clarity of journal abstracts. Br J Educ Psychol 64:331–343. https://doi.org/10.1111/j.2044-8279.1994.tb01106.x

Hartley J (2003) Improving the clarity of journal abstracts in psychology. Sci Commun 24:366–379. https://doi.org/10.1177/1075547002250301

Hartley J, Benjamin M (1998) An evaluation of structured abstracts in journals published by the British Psychological Society. Br J Educ Psychol 68:443–456. https://doi.org/10.1111/j.2044-8279.1998.tb01303.x

He Y, Kayaalp M (2006) A Comparison of 13 Tokenizers on MEDLINE. The Lister Hill National Center for Biomedical Communications Technical Report LHNCBC-TR-2006-003

Hubbard R, Vetter DE, Little EL (1998) Replication in strategic management: scientific testing for validity, generalizability, and usefulness. Strateg Manag J 19:243–254

Insko CA, Smith RH, Alicke MD, Wade J, Taylor S (1985) Conformity and group size: the concern with being right and the concern with being liked. Pers Soc Psychol Bull 11:41–50

Jasny BR, Chin G, Chong L, Vignieri S (2011) Again, and again, and again. Science 334:1225

Kelly C (2014) Too many cooks. https://www.youtube.com/watch?v=QrGrOK8oZG8. Accessed 23 Mar 2018

Kincaid JP, Fishburne RP, Rogers RL, Chissom BS (1975) Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Naval Technical Training Command Millington TN Research Branch RBR-8-75

Knight J (2003) Scientific literacy: clear as mud. Nature 423:376–378. https://doi.org/10.1038/423376a

Kuehne LM, Olden JD (2015) Opinion: Lay summaries needed to enhance science communication. Proc Natl Acad Sci USA 112:3585–3586. https://doi.org/10.1073/pnas.1500882112

Kuha J (2004) AIC and BIC: comparisons of assumptions and performance. Sociol Methods Res 33:188–229. https://doi.org/10.1177/0049124103262065

Lindsay RM, Ehrenberg ASC (1993) The design of replicated studies. Am Stat 47:217–228. https://doi.org/10.1080/00031305.1993.10475983

Loveland J, Whatley A, Ray B, Reidy R (1973) An Analysis of the readability of selected management journals. Acad Manag J 16:522–524. https://doi.org/10.5465/255014

Lu X, Ai H (2015) Syntactic complexity in college-level English writing: differences among writers with diverse L1 backgrounds. J Second Lang Writ 29:16–27

McLaughlin GH (1969) SMOG grading—a new readability formula. J Read 12:639–646

Otero J, Caldeira H, Gomes CJ (2004) The influence of the length of causal chains on question asking and on the comprehensibility of scientific texts. Contemp Educ Psychol 29:50–62. https://doi.org/10.1016/S0361-476X(03)00018-3

Pashler H, Harris CR (2012) Is the replicability crisis overblown? Three arguments examined. Perspectives in Psychological Science 7:531–536

Plavén-Sigray P, Matheson GJ, Schiffler BC, Thompson WH (2017) The readability of scientific texts is decreasing over time. Elife 6:e27725. https://doi.org/10.7554/eLife.27725

Scharrer L, Britt MA, Stadtler M, Bromme R (2013) Easy to understand but difficult to decide: information comprehensibility and controversiality affect Laypeople's science-based decisions. Discourse Process 50:361–387. https://doi.org/10.1080/0163853X.2013.813835

Senter RJ, Smith EA (1967) Automated readability index. Aerospace Medical Research Laboratories AMRL-TR-66-220

Tierney W, Hardy JH III, Ebersole CR, Leavitt K, Viganola D, Giulia C, Gordon M, Dreber A, Johannesson M, Pfeiffer T, Collaboration HDF, Uhlmann EL (2020) Creative destruction in science. Organ Behav Hum Decis Process 161:291–309

Tsang EWK, Kwan K-m (1999) Replication and theory development in organizational science: a critical realist perspective. Acad Manag Rev 24:759–780. https://doi.org/10.5465/amr.1999.2553252

Vinkers CH, Tijdink JK, Otte WM (2015) Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. BMJ. https://doi.org/10.1136/bmj.h6467

Yeung AWK, Goto TK, Leung WK (2018) Readability of the 100 most-cited neuroimaging papers assessed by common readability formulae. Front Hum Neurosci 12:308. https://doi.org/10.3389/fnhum.2018.00308