

# Text and Team: What Article Metadata Characteristics Drive Citations in Software Engineering?

Lorenz Graf-Vlachy\*

Daniel Graziotin

Stefan Wagner

{lorenz.graf-vlachy|daniel.graziotin|stefan.wagner}@iste.uni-stuttgart.de

University of Stuttgart, Institute of Software Engineering

Stuttgart, Germany

## ABSTRACT

*Context:* Citations are a key measure of scientific performance in most fields, including software engineering. However, there is limited research that studies which characteristics of articles' metadata (title, abstract, keywords, and author list) are driving citations in this field. *Objective:* In this study, we propose a simple theoretical model for how citations come to be with respect to article metadata, we hypothesize theoretical linkages between metadata characteristics and citations of articles, and we empirically test these hypotheses. *Method:* We use multiple regression analyses to examine a data set comprising the titles, abstracts, keywords, and authors of 16,131 software engineering articles published between 1990 and 2020 in 20 highly influential software engineering venues. *Results:* We find that number of authors, number of keywords, number of question marks and dividers in the title, number of acronyms, abstract length, abstract propositional idea density, and corresponding authors in the core Anglosphere are significantly related to citations. *Conclusion:* Various characteristics of articles' metadata are linked to the frequency with which the corresponding articles are cited. These results partially confirm and partially go counter to prior findings in software engineering and other disciplines.

## CCS CONCEPTS

• **Software and its engineering**; • **General and reference** → **Empirical studies**;

## KEYWORDS

citations, metadata, abstract, title, keyword, author

### ACM Reference Format:

Lorenz Graf-Vlachy, Daniel Graziotin, and Stefan Wagner. 2022. Text and Team: What Article Metadata Characteristics Drive Citations in Software Engineering?. In *The International Conference on Evaluation and Assessment in Software Engineering 2022 (EASE 2022)*, June 13–15, 2022, Gothenburg, Sweden.

\*Also with TU Dortmund University.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

EASE 2022, June 13–15, 2022, Gothenburg, Sweden

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9613-4/22/06...\$15.00

<https://doi.org/10.1145/3530019.3530022>

Sweden. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3530019.3530022>

## 1 INTRODUCTION

Citations are probably the most important metric used in science to assess researchers' performance [1, 5, 16]. Despite mounting and substantial criticism of citations as an evaluation criterion for both scientific venues and individual researchers [1, 46, 62], they (or related concepts like the h-index [31]) are widely used to make important assessments and subsequently decisions about researchers and even entire research groups [72]. Specifically, citations are frequently considered in hiring and promotion decisions [50]. More generally, citations are widely viewed as a metric that reflects a scientist's impact on their field [4, 49], as is evident, for instance, in the prominent display of citation counts at the top of every Google Scholar profile.

Consequently, prior researchers have sought to determine what drives citations. This debate takes place in various disciplinary fields, including most notably in the meta-scientific fields of bibliometrics and scientometrics [54, 69]. Importantly, several authors identified characteristics of an article's metadata as drivers of citations. For instance, researchers studied the impact of the length of an article's abstract and title on citations across various disciplines [73]. Similarly, researchers have identified the number of authors as being linked to citations [4, 73].

However, there is a dearth of empirical research in software engineering that systematically develops hypotheses on and tests which of such characteristics drive citations in this field. While some groundbreaking and promising work has been done, it is limited in several ways. Some work does not relate to articles' metadata but studies the effect of article content on citations [51]. Other work is mostly descriptive in nature and studies only a small set of drivers [23]. The arguably most extensive exemplar of extant work on the drivers of citations specifically in software engineering is a non-peer-reviewed working paper that studies journal but not conference papers [48].

This limited scholarship is regrettable because just as prior findings on the drivers of citations are plentiful, they are often inconsistent between samples and disciplines. For example, the aforementioned link between the length of the abstract and citations appears to exist in medicine but not in sociology [73]. The association between the length of the title and citations appears to be positive in the field of medicine [73], absent in biology and related disciplines [36] as well as addiction research [60], and even negative in applied physics [73] and marketing [65].

In this article, we contribute by proposing a parsimonious theoretical two-step filtering process model of how citations come to be with respect to article metadata, and by systematically assessing corresponding metadata drivers of citations in articles from some key venues in the field of software engineering. We identify relevant antecedents regarding articles’ text (titles, abstracts, and keywords) and author teams.

In the following, we first provide background on related research, the we develop explicit hypotheses, describe our study design, report our empirical results, discuss them, and conclude with an outlook on future research.

## 2 BACKGROUND AND RELATED WORK

Our central research question is: *Which readily observable characteristics of a software engineering article’s metadata (title, abstract, keywords, and author team) drive citations to the focal article?*

There are at least three streams of research that relate to this question. First, there is extensive research on the drivers of citations in fields outside of software engineering. In fact, researchers in different disciplines have identified at least 28 different antecedents of citation count [53, 54, 69]. Specifically, there appear to be three broad categories of factors that influence citations: paper-related, journal-related, and author-related [69].

On paper-related factors, authors have, for instance, studied characteristics of titles, abstracts, keywords, and the manuscript text of articles. Ayres and Vars [7] found that manuscript length has a positive effect on citations in legal scholarship, whereas title length has a negative effect. Stremersch et al. [65] found a similar negative effect of title length in marketing, as did Subotic and Mukherjee [68] in psychology. Van Wesel et al. [73] showed that abstract length positively affects citations in medicine and applied physics, but not in sociology. Buter and van Raan [11] demonstrated that punctuation marks such as hyphens and colons in titles have a small positive effect on citations in a wide range of disciplines, including clinical medicine and food science, but a negative effect in others like biology. Jamali and Nikzad [36] presented evidence that articles with question marks in the title are downloaded more frequently but ultimately less cited in the field of astronomy. Chakraborty et al. [13] and Rostami et al. [60] showed that the diversity of keywords positively predicts citations in a large multidisciplinary sample of articles. Relatedly, So et al. [64] found, among other things, that the number of keywords is positively associated with citations. Regarding the manuscript text of an article, van Wesel et al. [73] identified the number of pages, the number of sentences, and the number of cited references as factors influencing citations, but they observed substantial variance across academic disciplines.

Regarding journal-related factors, van Dalen and Henkens [71] found that the reputation of a journal (e.g., the journal impact factor or citations of editorial board members) is positively linked to citations of articles appearing in the focal journal in demography. Many others found similar results in other fields (e.g., [44, 74]).

Several author-related factors have also received attention. For instance, the number of authors has frequently been studied. Van Wesel et al. [73] found a positive relationship with citations in sociology, medicine, and applied physics. Ibáñez et al. [33] studied Spanish computer science professors and found that papers by

two authors receive the most citations, and that international collaboration is also positively associated with citations. Stremersch et al. [67] found at least partly contrasting evidence in marketing, suggesting that an increasing number of authors is linked to fewer citations. Stremersch et al. [67] also showed evidence that U.S. affiliations of authors do not drive citations, whereas their own prior work [66] and that of others suggested [71] such a link. Another oft-studied characteristic is the reputation of authors. Van Dalen and Henkens [71], for instance, found that author reputation in the form of accumulated past citations is positively related to citations.

It is worth noting that there exists extensive literature across disciplines and that—as the selected examples above already indicate—there is great variance in the results. For almost every significant finding, there are studies where the same characteristic is not significant or even significant in the opposite direction [69].

Second, there is a rich bibliometric literature in the field of software engineering. For one, researchers have shown an interest in identifying most-cited papers. For example, Wohlin [76–78] conducted a well-known series of studies that identified the most-cited papers in software engineering. Kitchenham [41] also analyzed the key topics, objectives, and content of the papers. Garousi and Fernandes [21] identified the top 100 most-cited papers in software engineering. For another, several bibliometric studies attempted to identify not top papers but top scholars and institutions in the field of software engineering in various geographies. Garousi and Varma [24], for instance, performed such an analysis for Canada, whereas Hummel et al. [32] took a global approach and ranked highly influential software engineering scholars who were program committee or organizing committee members of reputable software engineering venues or ACM SIGSOFT research award winners. Other authors fused these two substreams of research, reporting on both top papers and scholars [14]. Yet another substream takes a broader yet still primarily descriptive perspective on citations. Garousi and Mäntylä [23], for instance, performed a host of descriptive analyses of the citation distributions in software engineering, the studied topics in articles, as well as the field’s publication output over time, while also presenting evidence that review articles are more frequently cited than other article types and that paper age is related to citations. Ahmad and Raulamo-Jurvanen [2, 3] studied *XP Conference* contributions and identify most-cited papers and recently “hot” topics, mapped out citation distributions, and showed linkages between the number of authors and article age and the number of citations.

Finally, the last stream of related research is the intersection of the aforementioned streams, and thus directly concerns itself with the antecedents of citations in software engineering.<sup>1</sup> Specifically, this nascent stream contains several papers of note, which, however, all have relevant limitations.

Garousi and Fernandes [22] studied how different drivers impact citation frequency in software engineering. Specifically, they analyzed the citation impact of different types of venues (journals vs. conferences), different venues, authors’ countries of origin, and “top-10 authors” (as measured by authors’ total number of papers).

<sup>1</sup>Note that some of the articles listed in the category above also occasionally link article characteristics to citations (e.g., [3, 23, 41]). We nevertheless classified them as bibliometric studies because their primary interest was not the identification of drivers of citations.

While this study is highly instructive, it is exploratory as it treated each driver in isolation and did not combine all drivers into one model to account for potentially confounding effects between them.

Molléri et al. [51] analyzed the relationship between articles' reporting quality regarding relevance and rigor, and citations. They found no effect of relevance, but could demonstrate an association between reported rigor and citations. However, they did so without controlling for the potential influence of other drivers like article length, which they found to be positively correlated with rigor.

The paper most closely related to our work is a recent working paper by Mäntylä and Garousi [48]. They found that venues, past citations of authors, article length, and the number and recency of references in an article particularly affect the citations an article attracts. While this study doubtlessly substantially advances the field, it has not been formally reviewed yet, and it considered only journal publications, neglecting conference papers.

In addition, none of these prior studies engaged in formal theory testing [58]. While they all provide well-argued reasons for why certain factors might affect citations, none of them truly specified models *ex ante* and tested them. In fact, the majority of these articles did not jointly test a comprehensive model of drivers, but instead tested each variable in isolation, potentially missing confounding effects between multiple variables. Mäntylä and Garousi [48] overcame this limitation but introduced new variables into their regression models in a step-wise fashion. While this is helpful and appropriate for creating a robust set of predictors of citations, it is epistemologically at odds with our objective of building and testing a theoretical model.

Our work differs from prior work in the following ways. First, we attempt to develop a simple theoretical model of citations, explicitly theorize on drivers of citations, and then subject our hypotheses to empirical testing. Second, in this quest, we purposefully focus on the “early stage” of the citation process. As we detail below, we are particularly interested in metadata that is apparent from search results in scientific search engines, before a potentially citing researcher has read the entire paper. In other words, we are interested in a subset of what Onodera and Yoshikane label “extrinsic factors” [53, 54]. Third, we employ multiple regression analyses to account for multiple article characteristics at the same time, attempting to control for a wide range of potentially confounding factors. Finally, we introduce a hitherto unstudied characteristic of abstracts to the study of citations, namely abstracts' propositional idea density.

### 3 HYPOTHESES

Out of the many possible metadata characteristics of articles, we selected several for inclusion in this study. Our guiding logic here was (a) that these characteristics should be within the realm of influence of the article authors, (b) that we suspected a potential causal link to citation counts, rather than mere correlations, and (c) that the necessary data be available for a large-scale analysis. Specifically, we opted for characteristics related to the author team and the article itself, which are two of the three categories identified by Tahamtan et al. [69]. We decided to exclude the third category, i.e., journal characteristics such as a journal's impact factor. The reason for this is not that we believe these characteristics have no influence. To

the contrary, we believe that there are likely so many strong effects that we should capture them comprehensively through fixed effects, as we will explain in the method section. This does not allow us to hypothesize on and meaningfully test such effects individually, but it enables us to adequately control any confounding effects that might originate from them, thus strengthening our hypothesis tests regarding the other categories.

We propose that there is a two-step filtering process that needs to be passed for an article to be cited (see Figure 1). In a first step, readers must become aware of and attentive to an article to at least consider it citeable. Particularly easily accessible metadata that may, for instance, be shown as results by scientific search engines or that may influence the inclusion of an article in search results—like the article's author(s), keywords, and title—are likely to play a role here. In the second step, readers must be compelled to cite the focal article in their own work. In this step, readers are likely to consult at least the abstract of the article. The greater the likelihood that any given article passes each filter, the greater the likelihood that the article gets cited, and thus the greater its total citations. Note that we purposefully focus on these early steps of the citation process, which leads us to not study other potentially relevant variables that might affect citations but that are not influential in the early stages, e.g., the number of pages in an article [48] or the quality of an article's content [51]. In the following, we develop explicit hypotheses about several characteristics that may influence these two steps.

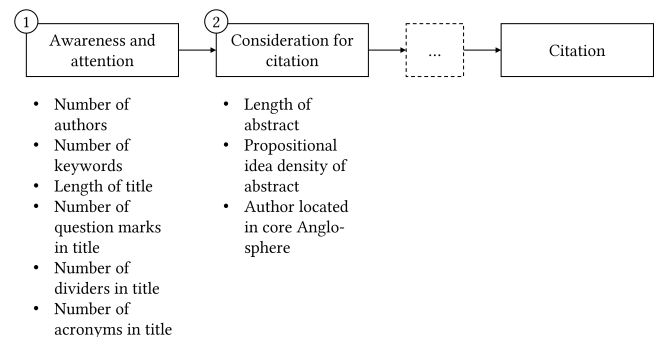


Figure 1: Filtering process

In the first step of the filtering process, awareness and attention, we propose that the *number of authors* has a positive influence. If the author team is larger, this might not only lead to higher-quality work due to the inclusion of diverse perspectives [55], or to more self-citations [29], but a larger author team will also have more opportunities through which to promote their article [45, 67]. Specifically, more authors can increase “knowledge diffusion” [9, p. 832] through presenting their work at more different conferences or workshops, and promoting it more through informal, personal channels [9]. In line with much of prior research, e.g., [4, 64, 73], we thus propose:

**Hypothesis 1a (H1a):** *There is a positive association between the number of authors and citations.*

Another way in which authors can make their work visible to potential readers is by including a *number of keywords*. When an

article includes many keywords, the article has a greater chance of showing up in searches [60], and it might thus have a greater chance of passing the first filtering step. In line with prior research [64], we therefore propose:

**Hypothesis 1b (H1b):** *There is a positive association between the number of keywords and citations.*

We next hypothesize a link between the *length of an article’s title* and citations. Specifically, we argue that longer titles have the potential to convey more, and more specific, information, making the article potentially more attractive to readers. Consequently, we hypothesize in line with some prior results [35, 60]:

**Hypothesis 1c (H1c):** *There is a positive association between the length of the title and citations.*

Another way in which authors might make their papers more appealing is by using certain rhetorical devices. Specifically, authors frequently pose questions (that they may or may not answer in the title), and thus use *question marks*, in an article’s title. Research on citations in computer science overall [18] and news media [43] has shown that the use of questions increases engagement with an article (“the clickbait phenomenon” [42, p. 1301]). We propose that the presence of such questions make an article more interesting and thus more likely to be read and considered by readers. Put formally:

**Hypothesis 1d (H1d):** *There is a positive association between the number of question marks in the title and citations.*

Similarly, authors often use *dividers* such as a colon or a dash in article titles. Composing such “combinational titles” [60, p. 2007] allows them to tease an interesting story (e.g., “Understanding IV&V in a safety critical and complex evolutionary environment: The NASA Space Shuttle program”), communicate the results in a vivid way (e.g., “Same App, Different Countries: A Preliminary User Reviews Study on Most Downloaded iOS Apps” or “Exception Analysis and Points-to Analysis: Better Together”), introduce a potentially useful named artifact (e.g., “MMLT: A mutual multilevel trust framework based on trusted third parties in multicloud environments”), or include methodological details like the type of study (e.g., “Software-testing education: A systematic literature mapping” or “How does combinatorial testing perform in the real world: an empirical study”). In addition, a colon or a dash allows the authors to provide more information, given that they can effectively include two sentences in the title instead of one. All of these cases are likely to make a paper more appealing and thus pass the first filtering process. Compatible with prior research [35, 60], we thus propose:

**Hypothesis 1e (H1e):** *There is a positive association between the number of dividers in the title and citations.*

In contrast to the prior characteristics, we propose that a greater *number of acronyms* in a title will make a paper, in general, less relevant and appealing to readers. Notably, this goes against prior findings in the medical field. However, these findings are likely due to particularities of that field, where acronyms may characterize articles from highly influential multi-center trials or organizations, or about specific genes [35]. We instead argue that, for one, acronyms might be hard (or impossible) to comprehend. For another, acronyms might signal a more specialized or narrower topic that likely only appeals to a smaller subset of readers. We thus formalize:

**Hypothesis 1f (H1f):** *There is a negative association between the number of acronyms in the title and citations.*

We now turn to the second filtering step of Figure 1, consideration for citation. In this step, readers have chosen to engage with an article at least far enough to read the abstract, and might be able to make a decision whether to cite it in their own work. One key determinant is likely the information content of the abstract. Specifically, the greater the *length of an abstract*, the more information it can contain and the more it might give the reader a good sense for an article, potentially even allowing the reader to cite the article without reading the entire text [37, 73]. Consequently, we argue:

**Hypothesis 2a (H2a):** *There is a positive association between the length of the abstract and citations.*

Another factor we consider is the *propositional idea density* [15] of the abstract. Propositional idea density captures the number of propositions or ideas, normalized by the number of words in a text. From the perspective of a stream of psycholinguistic research going back to seminal research by Kintsch and colleagues [38, 40], an idea is whatever can be true or false, and thus the basic unit involved in the understanding and retention of text. Consequently, higher idea density makes text harder to understand, but also richer in content (when keeping text length constant) [38–40]. For example, the two sentences “the previously obsolete system is relevant again” and “there is a system which was previously obsolete, and it is relevant again” have the same informational content but differ in their idea density. Prior literature has linked propositional idea density to cognitive functioning and impairment, even demonstrating, for example, its suitability as a screening tool for detecting Alzheimer’s disease based on a person’s writing [30, 59]. Conversely, we suggest that greater idea density in an abstract indicates a greater density of content and thus a larger number of reasons for which an article could be cited (independent of abstract length), thus leading to a higher number of citations [19, 61]. We posit:

**Hypothesis 2b (H2b):** *There is a positive association between the propositional idea density of the abstract and citations.*

Finally, there might be an additional effect of the authors on citations, which was found in prior work outside software engineering with regard to U.S.-based authors [66] and the Anglo-Saxon world [56]. Specifically, the entire field of computer science is substantially influenced by institutions and scholars that are based in the Anglo-Saxon world. Citing papers with *corresponding authors based in the “core Anglosphere”* [75], i.e., the developed Anglo-Saxon countries, might provide legitimacy to the citing work, as it will signal that the articles build on ideas from a critical country in the field. Authors at such institutions are also more likely to be native English speakers, potentially leading to more elegantly written abstracts and even articles, which again might increase citation rates. Put formally:

**Hypothesis 2c (H2c):** *There is a positive association between the corresponding author being located in the core Anglosphere and citations.*

## 4 STUDY DESIGN

In the following section, we detail the study design. Specifically, we discuss our sample and data, explain how we operationalized our variables, and describe the data analysis.

## 4.1 Sample and data

We obtained a sample of article abstracts and titles from Clarivate’s Web of Science (WoS, formerly known as Web of Knowledge), a database that has been used in recent studies analyzing textual characteristics and/or citations of articles in computer science, information systems, and other disciplines [11, 18, 25, 26, 53, 54, 73]. We opted for WoS since it is superior to Google Scholar in various ways [23, 69] and we did not have access to Scopus. We manually downloaded all WoS records of articles published between 1990 and 2020 and indexed in the category “Computer Science - Software Engineering” as of May 2021. Since this category also includes a host of venues that are not primarily or even tangentially related to software engineering, such as, for example, the *USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet*, we sought to appropriately restrict our sample. Specifically, we identified the top 20 venues according to Google Scholar from the category “Engineering & Computer Science - Software Systems” as of July 2021 (see Table 1), and only retained articles from these venues.

We restricted our sample to articles published after 1990 because the coverage of WoS is very limited before this time [26]. We further restricted our sample to articles published until 2020 for two reasons. First, at the time of data collection, the following year was not finished, and we would thus have necessarily obtained an incomplete sample for that year. Second, since our variable of interest is the number of citations, we can only analyze articles that had a realistic chance of getting cited in the first place. The time window for citations would have been extremely narrow for the most recent articles and would very likely only have introduced more noise into our sample than contribute explanatory value.

We dropped duplicate entries from WoS, removed all records that did not pertain to an “Article,” “Article; Proceedings Paper,” or “Proceedings Paper,” thereby removing other types of publications like book reviews or editorials.

As the same venue might be included in WoS under a host of different names (e.g., *OOPSLA 11: Proceedings of the 2011 ACM International Conference on Object Oriented Programming Systems Languages and Applications* and *OOPSLA 2008 Nashville, Conference Proceedings*) we performed data cleaning and manually identified all records that belong to the 20 venues.

To further clean the data, we performed minor preprocessing on the abstracts (but not the titles) using Perl scripts.<sup>2</sup> Specifically, we removed copyright and classification strings that were frequently included, we expanded some very common non-technical abbreviations like “e.g.” removed enumerations, and made adjustments to facilitate later analysis (such as converting colons to full stops and removing full stops in “et al.” to aid sentence recognition). We manually inspected 500 randomly sampled abstracts to ensure that our preprocessing did not have unintended side effects.

We dropped all abstracts with less than three full stops in them because spot checks indicated that these cases were errors in WoS, e.g., abstracts which contained no punctuation at all. We also dropped all articles which had titles with a length of three or less characters as these were clear errors in WoS.

Our sample in the 20 selected venues comprised 22,984 articles. Of these, we had sufficient data for 16,131 to use in our analyses.

<sup>2</sup>All used scripts are available at <https://doi.org/10.6084/m9.figshare.17209178>.

## 4.2 Operationalization of variables

We measure our dependent variable, the *number of citations*, simply as the total number of citations an article had accrued at the time of data collection, as recorded by WoS. Unlike some prior researchers, we do not log-transform the number of citations [73]. This is because this variable is a count, which has distributional properties that are best addressed by using an appropriate estimator rather than performing transformations [20]. Specifically, transformations cannot produce normally distributed errors which would be necessary for, for instance, ordinary least squares regression.

We measured our independent variables as follows. The *number of authors* is the count of authors as listed in WoS. The *number of keywords* is measured the same way. The *length of the title* is the number of characters in the article’s title, divided by 100 to make regression coefficients more interpretable. The *number of question marks in the title* is again a simple count. The *number of dividers in the title* is the count of colons followed by a blank space plus the count of dashes or double dashes preceded and succeeded by a blank space (to avoid incorrectly counting hyphenated words as in “3D-Color-Structure-Code - A new non-plainness island hierarchy” or “A Computation- and Communication- Infrastructure for Modular Special Instructions in a Dynamically Reconfigurable Processor”). The *number of acronyms in the title* was determined as the total number of all-caps words consisting of at least two characters (with or without separating full stops) in a title, while setting this value to zero for articles that had titles that were entirely in all-caps. The *length of the abstract* was the number of characters in the abstract before any preprocessing, again divided by 100 for interpretability.

To measure the *propositional idea density of the abstract*, we use CPIDR (Computerized Propositional Idea Density Rater) 5, a software specifically built and validated for this purpose [10], as well as used in prior research [17, 30]. The software exploits the fact that propositional idea density—as defined in the present article [38, 70]—can be approximated by relying on syntactic features of text. Specifically, performing part-of-speech tagging of text can provide approximations of the frequencies of various categories of words. Summing up the number of verbs, adjectives, adverbs, prepositions, and conjunctions, and dividing the sum by the total number of words then yields the propositional idea density of a text [63]. Note that Kintsch’s notion of propositional idea density [38] differs from formal semantics in that common nouns are not considered propositions. Similarly, information about verb tense, modality, or co-reference does not enter the count of propositions. The developers of the CPIDR software validated it in two ways [10]. First, they established that the software essentially replicates some examples provided in the literature on propositional idea density [70]. Second, they tested the software against human raters and found that the results of the software correlated very highly ( $r = .97$ ) with consensus ratings of human raters [10]. We standardized the value of propositional idea density for easier interpretation.

To determine whether the *corresponding author was located in the core Anglosphere*, we identified the institutional affiliation of the author listed as the “reprint author” (which indicates the corresponding author) in WoS, and created a binary variable set to one

**Table 1: Software engineering venues included in analysis**

#	Publication	h5-index
1	ACM/IEEE International Conference on Software Engineering	74
2	Journal of Systems and Software	61
3	Information and Software Technology	59
4	ACM SIGSOFT International Symposium on Foundations of Software Engineering	53
5	Empirical Software Engineering	53
6	IEEE Transactions on Software Engineering	52
7	ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)	48
8	ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)	46
9	IEEE/ACM International Conference on Automated Software Engineering (ASE)	45
10	IEEE Software	44
11	Symposium on Operating Systems Principles	42
12	Software & Systems Modeling	41
13	Mining Software Repositories	40
14	International Conference on Software Analysis, Evolution, and Reengineering (SANER)	40
15	International Symposium on Software Testing and Analysis	36
16	International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)	33
17	IEEE International Conference on Software Maintenance and Evolution	33
18	Proceedings of the ACM on Programming Languages	31
19	Software: Practice and Experience	30
20	ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA)	29

if this institution was located in the U.S., Canada, the U.K., Ireland, Australia, or New Zealand [75], and to zero otherwise.<sup>3</sup>

### 4.3 Data analysis

To account for the non-independence of observations at the venue-level, we select a fixed-effects Poisson estimator and calculate standard errors clustered at the venue. This estimator is appropriate for analyses with a dependent variable that is a count [20] such as the number of a paper’s citations. In addition, this estimator is robust to a host of potential problems and specifically more robust than a negative binomial estimator [79, 80]. Our choice to use cluster-robust standard errors accounts for the non-independence of observations, i.e., the fact that that error terms in the regression model might be correlated for articles because they were published in the same venue.

We also include an extensive array of so-called *dummy variables* in our models. Dummy variables are binary variables that are included in regression models to represent categorical values [28]. Typically, dummy variables are used to account for so-called “fixed effects” of specific types of observations [80], e.g., the effects of the specific venue in which an article was published [48]. To this end, a dummy variable is created for every venue in the data set, and for every article only the corresponding dummy variable is set to one, with all others taking the value zero.<sup>4</sup>

We include dummies for three reasons. First, we decided to not normalize citation counts by the age of a paper (e.g., [23, 51]) because this requires using a ratio as a dependent variable, which is not advisable in regression models [12]. Instead, we can include dummy

variables for every year, which explicitly accounts for an article’s age, which influences citations for obvious reasons [3, 23, 48, 69]. This solution has the added advantage that it accounts for all other effects that pertain to a specific year as well, e.g., the volume of total scientific output or total citations made by all researchers in a given year, and it does not assume that any time effects be linear.

Second, as journal articles are typically cited more frequently than conference papers [23], and there are systematic differences between different subfields of a discipline [69] as well as venues [33, 48, 69], it would further seem prudent to include dummy variables for each venue [71]. This way, they control for all time-invariant effects of all venues.

Finally, there might be effects that are fixed for every venue-year combination. We thus included interactions (i.e., multiplications) between all year and venue dummy variables. This allows us to account for all effects that might be related to all articles appearing in a venue in a specific year, capturing, e.g., the effects of changing journal or conference prestige or impact factors over time. Note that our model thus explicitly accounts for all these variables, which have often been studied in prior literature [48]. We deem their importance so uncontroversial that we did not hypothesize about them. Instead, we incorporate them in our model in a way that not only controls for their specific effects, but that also allows us to control for all other effects that are due to particularities of venues in specific years (and which prior literature typically does not account for). This includes, for instance, editorial policies or regulations regarding maximum abstract or title lengths. Our model is thus, in this way, more comprehensive than prior work in software engineering.<sup>5</sup>

<sup>3</sup>Note that because our argument is largely based on the specific legitimacy-granting power of these countries, we deliberately do not consider other countries that have English as a *de jure* or *de facto* official language.

<sup>4</sup>To avoid perfect multicollinearity in the regression, one of the dummy variables must be deliberately omitted. This is done automatically by the statistical software we use.

<sup>5</sup>Because these dummy variables contain all the information contained in the aforementioned individual dummy variables for venues only and for years only, these individual variables are automatically omitted from our regression analyses due to perfect multicollinearity without any loss of information.

We performed all statistical analyses using Stata 17.0.

## 5 RESULTS

Table 2 displays key descriptive statistics for the main variables. Correlations between them are shown in Table 3. Variance inflation factors were all far below commonly used thresholds, alleviating concerns over multicollinearity [52].

Table 4 shows the regression results.<sup>6</sup> We find a positive association between the number of authors and citations, supporting H1a. The number of keywords is only marginally related to citations, and negatively at that, providing weak evidence against H1b. We also find no support for H1c, which suggested a link between the length of an article’s title and citations. Citations were, however, significantly positively linked to the number of questions marks (supporting H1d) and dividers (supporting H1e) in abstract titles. The hypothesized negative effect of acronyms in the title on citations is also borne out in our model, supporting H1f. Regarding articles’ abstracts, we find support for H2a and H2b, since both length and propositional idea density of the abstract were significantly positively related to citations. Finally, an affiliation of the corresponding author with an institution in the core Anglosphere is also linked to citations, supporting H2c.

Our results are robust to alternative model specifications. For instance, all coefficients that are significant in our primary analysis are also significant in a negative binomial regression with robust standard errors. In addition, the coefficient for the number of keywords, which was insignificant in the main analysis, also becomes significant in a negative binomial model.

Although concerns over multiple hypothesis testing are not common in the literature on citations, we performed an additional robustness check to control the false discovery rate and computed adjusted  $p$ -values (sharpened  $q$ -values) [6, 8] for all variables of interest. The results essentially reaffirm our original findings, even strengthening the result for the number of keywords, and letting the coefficient of title length reach marginal significance.

## 6 DISCUSSION

In the following, we discuss implications as well as threats to internal and external validity.

### 6.1 Implications

Our empirical results have several important implications.

**Implication #1:** Early-stage metadata characteristics drive citations in software engineering. Our results show that citations can be at least partially explained by metadata characteristics that might function as “signals” [71] and that likely have a particularly meaningful effect at the early stages of the literature review process.

**Implication #2:** Propositional idea density matters. As the first study in software engineering, we identified a link between the idea density of articles’ abstracts and the citations they receive.

<sup>6</sup>We do not report the coefficients of the dummy variables as their interpretation is likely not helpful in our setting. Given the inclusion of venue-year dummies, we could study questions like “Is the effect of journal A in year X greater than the effect of conference B in year Y?” but we cannot, for example, make blanket comparisons between venues. A manual inspection of the dummy variables shows that many of them are significant, confirming our initial assumption that there are meaningful effects of publication venue and year that should be accounted for.

**Table 2: Descriptive statistics of all variables**

	Mean	SD	Min.	Max.
Citations	17.63	46.64	0.00	2,301.00
Authors (log)	1.08	0.52	0.00	3.50
Keywords	4.49	1.68	1.00	23.00
Length (title)	0.72	0.24	0.01	2.01
Question marks (title)	0.03	0.17	0.00	3.00
Dividers (title)	0.29	0.46	0.00	2.00
Acronyms (title)	0.23	0.51	0.00	6.00
Length (abstract)	12.18	4.18	2.43	41.37
Idea density (abstract)	0.00	1.00	-3.93	4.85
Anglosphere author	0.39	0.49	0.00	1.00

SD = Standard deviation

$N = 16,131$

**Implication #3:** The software engineering discipline behaves differently. Despite many findings confirming prior results in other disciplines, there is a marked difference regarding some of our results. For instance, whereas the relationship between the number of authors and citations is negative in marketing [67], it is positive in software engineering. While the presence of question marks in titles has a negative association with citations in biology and medicine [36], we find the opposite in software engineering. Conversely, while acronyms drive citations in medicine [35], they seem to inhibit them in software engineering.

**Implication #4:** The effects are robust. Going beyond extant literature, we controlled for venue-year fixed effects, effectively only analyzing variation between articles within a venue in a given year. We additionally used robust standard errors to account for the non-independence of observations. Despite these advanced statistical controls, the results hold.

### 6.2 Threats to validity

**6.2.1 Internal validity.** There are several potential threats to validity. While we controlled for a host of potential confounders, we are hesitant to claim proof for causality. There might be other factors that we either deliberately do not include (because we focus on the early stage of the citation process) or cannot account for and which influence both our independent variables and drive citations. The quality and importance of each individual paper is such a potentially confounding factor. Unfortunately, controlling for paper quality is notoriously difficult and thus rarely done. Initial work on research quality in empirical software engineering has attempted to operationalize quality as reported rigor and relevance, and manually scored both dimensions using rubrics that evaluate degrees of quality based on which information is disclosed in a focal article [34]. While this effort was influential and is certainly insightful, we did not apply it for four reasons. First, we propose that reviewers and editors at the different venues we study act as quality gatekeepers and apply reasonably consistent quality standards within each venue. We do thus not immediately see why fundamental article quality within a venue should vary strongly. Recall that any such venue effects are controlled for through the included dummy variables. Second, the quality measure captures reporting quality rather than necessarily actual quality of a focal

**Table 3: Correlations between all variables**

	Citations	Authors (log)	Key-words	Length (title)	Question marks (title)	Dividers (title)	Acronyms (title)	Length (abstract)	Idea density (abstract)	Anglo-sphere author
Citations	1.00									
Authors (log)	0.02	1.00								
Keywords	0.03***	-0.03***	1.00							
Length (title)	-0.02*	0.10***	0.09***	1.00						
Question marks (title)	0.01	0.04***	-0.01	0.05***	1.00					
Dividers (title)	-0.01	0.09***	-0.01	0.22***	-0.02*	1.00				
Acronyms (title)	-0.04***	0.05***	0.02*	0.03**	-0.04***	0.18***	1.00			
Length (abstract)	0.00	0.18***	0.11***	0.20***	0.05***	0.04***	-0.02**	1.00		
Idea density (abstract)	0.01	0.02**	-0.03***	-0.07***	0.06***	0.02*	-0.11***	0.03***	1.00	
Anglosphere author	0.07***	-0.08***	0.01	-0.11***	0.01	0.00	-0.05***	-0.07***	0.07***	1.00

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

**Table 4: Drivers of citations**

	Coefficient	IRR
Authors (log)	0.331*** (0.047)	1.392
Keywords	-0.016+ (0.009)	0.984
Length (title)	-0.217 (0.190)	0.805
Question marks (title)	0.414*** (0.065)	1.512
Dividers (title)	0.150* (0.063)	1.161
Acronyms (title)	-0.192*** (0.058)	0.826
Length (abstract)	0.017* (0.008)	1.017
Idea density (abstract)	0.068*** (0.016)	1.070
Anglosphere author	0.161*** (0.038)	1.174
Observations	16,131	
Venue fixed effects	Yes	
Year fixed effects	Yes	
Venue-year fixed effects	Yes	

Fixed-effects Poisson regression

Cluster-robust standard errors (by venue) in parentheses

IRR = Incidence rate ratio

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

article. While reporting quality may be important, it need not be indicative of the actual quality of the work [34]. Third, in prior work in the empirical software engineering literature [51], relevance was not found to be related to citations. While one prior analysis suggests an association between rigor and citations, other features like article length positively co-varied with rigor and were not controlled for in the analysis, suggesting that the association

may not be causal. Overall, it thus appears warranted to conclude that “rigor, relevance and normalized citations are fairly orthogonal variables” [51, p. 1466]. Fourth and critically, the measure is simply impractical to use in studies with samples as large as ours.

The texts we study do not represent the entire text of the scientific articles. While we welcome future studies that also analyze the full texts of articles, we consider ourselves in the good company of prior researchers who specifically studied titles and abstracts because these elements are the ones that set the “hook” and lure readers into reading an article in the first place [47]. In fact the abstract “is often the only part of the article that will be read” [57], and abstracts and titles are crucial surrogate information [37].

Our study also has limitations stemming from the data we used. Specifically, WoS contains missing data for some variables in a number of observations. Combined with our multiple regression approach, this means that we had to disregard such observations. Prior research suffered from similar limitations [48].

**6.2.2 External validity.** Our analysis focuses on the top 20 venues in the field of empirical software engineering over a time span of 30 years. Any statements about the validity of our results beyond this context—say, regarding other disciplines, less influential or broader software engineering venues, and other points in time including the future—would thus be speculative.

Nevertheless, to make an attempt to empirically further assess generalizability, we re-ran our analysis over the entire sample of abstracts we obtained ( $N = 255,857$ ). Recall that we deemed this sample not fully reliable for our main analysis for several reasons, such as the inclusion of venues only marginally or not at all related to software engineering, and issues in harmonizing venue titles for such a large sample, making any venue-related fixed effects somewhat less reliable. To address the latter issue in the re-analysis, we performed a simple automated harmonization of venue names (e.g., removing numerals from conference names) to reduce the overall number of distinct venue names in the sample. Since the number of venues still remained in the thousands, however, we needed to simplify our analysis to make it computationally feasible. Specifically, we retained the venue and year dummies, but omitted interactions between the two from this analysis.



Our findings in this re-analysis on the entire sample largely corroborate our findings. Specifically, all coefficients that were significant in the main sample were also significant with the same sign in the entire sample ( $p < 0.05$ ), with the exception of the number of acronyms in the paper title, which is marginally significant in the opposite direction in the entire sample ( $p < 0.1$ ). Conversely, the two insignificant (length of title) or marginally significant (number of keywords) variables in the main sample were significant in the entire sample ( $p < 0.001$ ). Notably, the number of keywords displayed the initially hypothesized positive relationship with citations in the entire sample. Overall, the findings in the entire sample suggest that most of our findings have validity in a larger software engineering context.

Finally, as other researchers cautioned, using different bibliometric databases might lead to potentially different results. In fact, while this is a limitation, we also consider it a contribution of our paper that we use a different database than that used in the working paper that is closest to our work [48].

## 7 CONCLUSION AND FUTURE WORK

Our results suggest that article metadata are related to citations. Authors might thus wish to purposefully consider how to phrase their titles and abstracts to attract a maximum number of citations [27]. Due to the potential concerns about causality, however, we also encourage future researchers to build on our work and that of others and study drivers of citations in software engineering further. Particularly, studying “later-stage” characteristics of papers, such as the actual substantive content, appears fruitful. Comprehensive models, considering large arrays of characteristics simultaneously, as well as the use of comprehensive statistical control techniques like fixed effects, might be particularly useful in this quest.

## ACKNOWLEDGMENTS

We thank Michael A. Covington for providing us with CPIDR 5.

## REFERENCES

- [1] Robert Adler, John Ewing, and Peter Taylor. 2009. Citation Statistics. *Statist. Sci.* 24, 1 (2009), 1–14. <https://doi.org/10.1214/09-STS285>
- [2] Muhammad Ovais Ahmad and Paivi Raulamo-Jurvanen. 2019. Preliminary Citation and Topic Analysis of International Conference on Agile Software Development Papers (2002-2018). In *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems (Annals of Computer Science and Information Systems)*. IEEE, 803–812. <https://doi.org/10.15439/2019F114>
- [3] Muhammad Ovais Ahmad and Paivi Raulamo-Jurvanen. 2020. Scientific Collaboration, Citation and Topic Analysis of International Conference on Agile Software Development Papers. In *Advances in Agile and User-Centred Software Engineering*, Adam Przybyłek and Miguel Ehécatl Morales-Trujillo (Eds.). Lecture Notes in Business Information Processing, Vol. 376. Springer International Publishing, Cham, 108–132. [https://doi.org/10.1007/978-3-030-37534-8\\_16](https://doi.org/10.1007/978-3-030-37534-8_16)
- [4] Dag W. Aksnes. 2003. Characteristics of highly cited papers. *Research Evaluation* 12, 3 (2003), 159–170. <https://doi.org/10.3152/147154403781776645>
- [5] Dag W. Aksnes and Arie Rip. 2009. Researchers’ perceptions of citations. *Research Policy* 38, 6 (2009), 895–905. <https://doi.org/10.1016/j.respol.2009.02.001>
- [6] Michael L. Anderson. 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J. Amer. Statist. Assoc.* 103, 484 (2008), 1481–1495. <https://doi.org/10.1198/016214508000000841>
- [7] Ian Ayres and Fredrick E. Vars. 2000. Determinants of Citations to Articles in Elite Law Reviews. *The Journal of Legal Studies* 29, S1 (2000), 427–450. <https://doi.org/10.1086/468081>
- [8] Yoav Benjamini, Abba M. Krieger, and Daniel Yekutieli. 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 3 (2006), 491–507. <https://doi.org/10.1093/biomet/93.3.491>
- [9] Clément Bosquet and Pierre-Philippe Combes. 2013. Are academics who publish more also more cited? Individual determinants of publication and citation records. *Scientometrics* 97, 3 (2013), 831–857. <https://doi.org/10.1007/s11192-013-0996-6>
- [10] Cati Brown, Tony Snodgrass, Susan J. Kemper, Ruth Herman, and Michael A. Covington. 2008. Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods* 40, 2 (2008), 540–545. <https://doi.org/10.3758/BRM.40.2.540>
- [11] R. K. Buter and A.F.J. van Raan. 2011. Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact. *Journal of Informetrics* 5, 4 (2011), 608–617. <https://doi.org/10.1016/j.joi.2011.05.008>
- [12] S. Trevis Certo, John R. Busenbark, Matias Kalm, and Jeffery A. LePine. 2020. Divided We Fall: How Ratios Undermine Research in Strategic Management. *Organizational Research Methods* 23, 2 (2020), 211–237. <https://doi.org/10.1177/1094428118773455>
- [13] Tanmoy Chakraborty, Suhansanu Kumar, Pawan Goyal, Niloy Ganguly, and Animesh Mukherjee. [n.d.]. Towards a stratified learning approach to predict future citation counts. In *IEEE/ACM Joint Conference on Digital Libraries 2014*, 351–360. <https://doi.org/10.1109/JCDL.2014.6970190>
- [14] Sun-Wen Chuang, Tainyi Luor, and Hsi-Peng Lu. 2014. Assessment of institutions, scholars, and contributions on agile software development (2001–2012). *Journal of Systems and Software* 93 (2014), 84–101. <https://doi.org/10.1016/j.jss.2014.03.006>
- [15] Michael A. Covington. [n.d.]. Idea density — A potentially informative characteristic of retrieved documents. In *IEEE Southeastcon 2009*, 201–203. <https://doi.org/10.1109/SECON.2009.5174076>
- [16] Blaise Cronin and Kara Overfelt. 1994. Citation-based auditing of academic performance. *Journal of the American Society for Information Science* 45, 2 (1994), 61–72. [https://doi.org/10.1002/\(SICI\)1097-4571\(199403\)45:2%3C61::AID-ASII%3E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-4571(199403)45:2%3C61::AID-ASII%3E3.0.CO;2-F)
- [17] Alison Ferguson, Elizabeth Spencer, Hugh Craig, and Kim Colyvas. 2014. Propositional idea density in women’s written language over the lifespan: computerized analysis. *Cortex* 55 (2014), 107–121. <https://doi.org/10.1016/j.cortex.2013.05.012>
- [18] Dalibor Fiala, Pavel Král, and Martin Dostal. 2021. Are Papers Asking Questions Cited More Frequently in Computer Science? *Computers* 10, 8 (2021), 96. <https://doi.org/10.3390/computers10080096>
- [19] Boris Forthmann and Mark A. Runco. 2020. An Empirical Test of the Inter-Relationships between Various Bibliometric Creative Scholarship Indicators. *Publications* 8, 2 (2020), 34. <https://doi.org/10.3390/publications8020034>
- [20] William Gardner, Edward P. Mulvey, and Esther C. Shaw. 1995. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin* 118, 3 (1995), 392–404. <https://doi.org/10.1037/0033-2909.118.3.392>
- [21] Vahid Garousi and João M. Fernandes. 2016. Highly-cited papers in software engineering: The top-100. *Information and Software Technology* 71 (2016), 108–128. <https://doi.org/10.1016/j.infsof.2015.11.003>
- [22] Vahid Garousi and João M. Fernandes. 2017. Quantity versus impact of software engineering papers: a quantitative study. *Scientometrics* 112, 2 (2017), 963–1006. <https://doi.org/10.1007/s11192-017-2419-6>
- [23] Vahid Garousi and Mika V. Mäntylä. 2016. Citations, research topics and active countries in software engineering: A bibliometrics study. *Computer Science Review* 19 (2016), 56–77. <https://doi.org/10.1016/j.cosrev.2015.12.002>
- [24] Vahid Garousi and Tan Varma. 2010. A bibliometric assessment of canadian software engineering scholars and institutions (1996-2006). *Computer and Information Science* 3, 2 (2010), 19–29.
- [25] Lorenz Graf-Vlachy. 2021. Is the readability of abstracts decreasing in management research? *Review of Managerial Science* (2021). <https://doi.org/10.1007/s11846-021-00468-7>
- [26] Lorenz Graf-Vlachy. 2021. The readability of information systems research over three decades: An analysis of the Senior Scholars’ Basket of Journals from 1990 to 2020. In *Pacific Asia Conference on Information Systems Proceedings*, 1–7.
- [27] Maria J. Grant. 2013. What makes a good title? *Health Information and Libraries Journal* 30, 4 (2013), 259–260. <https://doi.org/10.1111/hir.12049>
- [28] Hardy, Melissa, A. 1993. *Regression with Dummy Variables*. Quantitative Applications in the Social Sciences, Vol. 07-093. Sage, Newbury Park, CA.
- [29] H. Hertz. 1995. Does it pay to cooperate? A bibliometric case study in molecular biology. *Scientometrics* 33, 1 (1995), 117–122. <https://doi.org/10.1007/bf02020777>
- [30] Edward Hill, Jane Alty, Larissa Bartlett, Lyn Goldberg, Mira Park, Soonja Yeom, and James Vickers. 2021. Automated analysis of propositional idea density in older adults. *Cortex* 145 (2021), 264–272. <https://doi.org/10.1016/j.cortex.2021.09.018>
- [31] J. E. Hirsch. 2005. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102, 46 (2005), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- [32] Oliver Hummel, Alexander Gerhart, and Bernhard Schäfer. 2012. Analyzing Citation Frequencies of Leading Software Engineering Scholars. *Computer and Information Science* 6, 1 (2012), 1–14. <https://doi.org/10.5539/cis.v6n1p1>
- [33] Alfonso Ibáñez, Concha Bielza, and Pedro Larrañaga. 2013. Relationship among research collaboration, number of documents and number of citations: a case

- study in Spanish computer science production in 2000–2009. *Scientometrics* 95, 2 (2013), 689–716. <https://doi.org/10.1007/s11192-012-0883-6>
- [34] Martin Ivarsson and Tony Gorschek. 2011. A method for evaluating rigor and industrial relevance of technology evaluations. *Empirical Software Engineering* 16, 3 (2011), 365–395. <https://doi.org/10.1007/s10664-010-9146-4>
- [35] Thomas S. Jacques and Neil J. Sebire. 2010. The impact of article titles on citation hits: an analysis of general and specialist medical journals. *Journal of the Royal Society of Medicine Short Reports* 1, 1 (2010), 2. <https://doi.org/10.1258/shorts.2009.100020>
- [36] Hamid R. Jamali and Mahsa Nikzad. 2011. Article title type and its relation with the number of downloads and citations. *Scientometrics* 88, 2 (2011), 653–661. <https://doi.org/10.1007/s11192-011-0412-z>
- [37] Joseph W. Janes. 1991. Relevance judgments and the incremental presentation of document representations. *Information Processing & Management* 27, 6 (1991), 629–646. [https://doi.org/10.1016/0306-4573\(91\)90004-6](https://doi.org/10.1016/0306-4573(91)90004-6)
- [38] Walter Kintsch. 1974. *The representation of meaning in memory*. Erlbaum, Hillsdale.
- [39] Walter Kintsch. 1998. *Comprehension*. Cambridge University Press, Cambridge, UK.
- [40] Walter Kintsch and Janice Keenan. 1973. Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology* 5, 3 (1973), 257–274.
- [41] Barbara Kitchenham. 2010. What's up with software metrics? – A preliminary mapping study. *Journal of Systems and Software* 83, 1 (2010), 37–51. <https://doi.org/10.1016/j.jss.2009.06.041>
- [42] Jeffrey Kuiken, Anne Schuth, Martijn Spitters, and Maarten Marx. 2017. Effective Headlines of Newspaper Articles in a Digital Environment. *Digital Journalism* 5, 10 (2017), 1300–1314. <https://doi.org/10.1080/21670811.2017.1279978>
- [43] Linda Lai and Audun Farbot. 2014. What makes you click? The effect of question headlines on readership in computer-mediated communication. *Social Influence* 9, 4 (2014), 289–299. <https://doi.org/10.1080/15534510.2013.847859>
- [44] Vincent Larivière and Yves Gingras. 2010. The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology* 61, 2 (2010), 424–427. <https://doi.org/10.1002/asi.21232>
- [45] Roosa Leimu and Julia Koricheva. 2005. What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution* 20, 1 (2005), 28–32. <https://doi.org/10.1016/j.tree.2004.10.010>
- [46] Kåre Letrud and Sigbjørn Hernes. 2019. Affirmative citation bias in scientific myth debunking: A three-in-one case study. *PLOS ONE* 14, 9 (2019), e0222213. <https://doi.org/10.1371/journal.pone.0222213>
- [47] Michael A. Mabe and Mayur Amin. 2002. Dr Jekyll and Dr Hyde: author–reader asymmetries in scholarly publishing. *Aslib Proceedings* 54, 3 (2002), 149–157. <https://doi.org/10.1108/00012530210441692>
- [48] Mika Mäntylä and Vahid Garousi. 2019. Citations in Software Engineering - Paper-related, Journal-related, and Author-related Factors. <http://arxiv.org/pdf/1908.04122v2>
- [49] Henk F. Moed. 2005. *Citation analysis in research evaluation*. Information Science and Knowledge Management, Vol. 9. Springer, Dordrecht.
- [50] David Moher, Florian Naudet, Ioana A. Cristea, Frank Miedema, John P. A. Ioannidis, and Steven N. Goodman. 2018. Assessing scientists for hiring, promotion, and tenure. *PLOS Biology* 16, 3 (2018), e2004089. <https://doi.org/10.1371/journal.pbio.2004089>
- [51] Jefferson Seide Molléri, Kai Petersen, and Emilia Mendes. 2018. Towards understanding the relation between citations and research quality in software engineering studies. *Scientometrics* 117, 3 (2018), 1453–1478. <https://doi.org/10.1007/s11192-018-2907-3>
- [52] Robert M. O'Brien. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity* 41, 5 (2007), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- [53] Natsuo Onodera and Fuyuki Yoshikane. 2015. Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology* 66, 4 (2015), 739–764. <https://doi.org/10.1002/asi.23209>
- [54] Natsuo Onodera and Fuyuki Yoshikane. 2021. Extrinsic Factors Affecting Citation Frequencies of Research Articles. In *Predicting the Dynamics of Research Impact*, Yannis Manolopoulos and Thanasis Vergoulis (Eds.). Springer International Publishing, Cham, 23–49. [https://doi.org/10.1007/978-3-030-86668-6\\_12](https://doi.org/10.1007/978-3-030-86668-6_12)
- [55] André Andrian Padial, João Carlos Nabout, Tadeu Siqueira, Luis Mauricio Bini, and José Alexandre Felizola Diniz-Filho. 2010. Weak evidence for determinants of citation frequency in ecological articles. *Scientometrics* 85, 1 (2010), 1–12. <https://doi.org/10.1007/s11192-010-0231-7>
- [56] Tai-Quan Peng and Jonathan J.H. Zhu. 2012. Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology* 63, 9 (2012), 1789–1803. <https://doi.org/10.1002/asi.22649>
- [57] Roy M. Pitkin, Mary Ann Branagan, and Leon F. Burmeister. 1999. Accuracy of data in abstracts of published research articles. *JAMA* 281, 12 (1999), 1110–1111. <https://doi.org/10.1001/jama.281.12.1110>
- [58] Karl R. Popper. 2005. *Logik der Forschung* (11 ed.). Gesammelte Werke in deutscher Sprache, Vol. 3. Mohr Siebeck, Tübingen.
- [59] Kathryn P. Riley, David A. Snowdon, Mark F. Desrosiers, and William R. Markesbery. 2005. Early life linguistic ability, late life cognitive function, and neuropathology: findings from the Nun Study. *Neurobiology of Aging* 26, 3 (2005), 341–347. <https://doi.org/10.1016/j.neurobiolaging.2004.06.019>
- [60] Fatemeh Rostami, Asghar Mohammadpoorals, and Mohammad Hajizadeh. 2014. The effect of characteristics of title on citation rates of articles. *Scientometrics* 98, 3 (2014), 2007–2010. <https://doi.org/10.1007/s11192-013-1118-1>
- [61] Mark Runco, Burak Turkman, Selcuk Acar, and Mustafa Nural. 2017. Idea Density and the Creativity of Written Works. *Journal of Genius and Eminence* 2, 1 (2017), 26–31. <https://doi.org/10.18536/jge.2017.04.02.01.03>
- [62] Marco Seeber, Mattia Cattaneo, Michele Meoli, and Paolo Malighetti. 2019. Self-citations as strategic response to the use of metrics for career decisions. *Research Policy* 48, 2 (2019), 478–491. <https://doi.org/10.1016/j.respol.2017.12.004>
- [63] David A. Snowdon, Susan J. Kemper, James A. Mortimer, Lydia H. Greiner, David R. Wekstein, and William R. Markesbery. 1996. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA* 275, 7 (1996), 528–532.
- [64] Minh So, Jiyoung Kim, Sangki Choi, and Han Woo Park. 2015. Factors affecting citation networks in science and technology: focused on non-quality factors. *Quality & Quantity* 49, 4 (2015), 1513–1530. <https://doi.org/10.1007/s11135-014-0110-z>
- [65] Stefan Stremersch, Nuno Camacho, Sofie Vanneste, and Isabel Verniers. 2015. Unraveling scientific impact: Citation types in marketing journals. *International Journal of Research in Marketing* 32, 1 (2015), 64–77. <https://doi.org/10.1016/j.ijresmar.2014.09.004>
- [66] Stefan Stremersch and Peter C. Verhoef. 2005. Globalization of Authorship in the Marketing Discipline: Does It Help or Hinder the Field? *Marketing Science* 24, 4 (2005), 585–594. <https://doi.org/10.1287/mksc.1050.0152>
- [67] Stefan Stremersch, Isabel Verniers, and Peter C. Verhoef. 2007. The Quest for Citations: Drivers of Article Impact. *Journal of Marketing* 71, 3 (2007), 171–193. <https://doi.org/10.1509/jmk.71.3.171>
- [68] Sinisa Subotic and Bhaskar Mukherjee. 2014. Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science* 40, 1 (2014), 115–124. <https://doi.org/10.1177/0165551513511393>
- [69] Iman Tahamtan, Askar Safipour Afshar, and Khadijeh Ahamdzadeh. 2016. Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics* 107, 3 (2016), 1195–1225. <https://doi.org/10.1007/s11192-016-1889-2>
- [70] A. Turner and E. Greene. 1977. *The Construction and Use of a Propositional Text Base: Technical Report 63*. Boulder, CO.
- [71] Hendrik P. van Dalen and Kène Henkens. 2005. Signals in science - On the importance of signaling in gaining attention in science. *Scientometrics* 64, 2 (2005), 209–233. <https://doi.org/10.1007/s11192-005-0248-5>
- [72] Anthony F. J. van Raan. 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics* 67, 3 (2006), 491–502. <https://doi.org/10.1556/Scient.67.2006.3.10>
- [73] Maarten van Wesel, Sally Wyatt, and Jeroen ten Haaf. 2014. What a difference a colon makes: how superficial factors influence subsequent citation. *Scientometrics* 98, 3 (2014), 1601–1615. <https://doi.org/10.1007/s11192-013-1154-x>
- [74] Mingyang Wang, Guang Yu, and Daren Yu. 2011. Mining typical features for highly cited papers. *Scientometrics* 87, 3 (2011), 695–706. <https://doi.org/10.1007/s11192-011-0366-1>
- [75] Ben Wellings and Andrew Mycock (Eds.). 2019. *The anglosphere: Continuity, dissonance and location*. Proceedings of the British Academy, 068-1202, Vol. 226. Oxford University Press, Oxford.
- [76] Claes Wohlin. 2005. An analysis of the most cited articles in software engineering journals - 1999. *Information and Software Technology* 47, 15 (2005), 957–964. <https://doi.org/10.1016/j.infsof.2005.09.002>
- [77] Claes Wohlin. 2007. An analysis of the most cited articles in software engineering journals - 2000. *Information and Software Technology* 49, 1 (2007), 2–11. <https://doi.org/10.1016/j.infsof.2006.08.004>
- [78] Claes Wohlin. 2008. An analysis of the most cited articles in software engineering journals - 2001. *Information and Software Technology* 50, 1-2 (2008), 3–9. <https://doi.org/10.1016/j.infsof.2007.10.002>
- [79] Jeffrey M. Wooldridge. 1999. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 90, 1 (1999), 77–97. [https://doi.org/10.1016/S0304-4076\(98\)00033-5](https://doi.org/10.1016/S0304-4076(98)00033-5)
- [80] Jeffrey M. Wooldridge. 2010. *Econometric analysis of cross section and panel data* (2nd ed.). MIT, Cambridge, Mass. and London.