Bibliometric Indicators and Relevance Criteria – An Online Experiment

Jacqueline Sachse [a]

[a] *Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany*

Abstract

Previous studies on relevance criteria neglected the role of bibliometric indicators, although their use in search engine results pages is widespread. This paper reports on an online search experiment, considering both the traditional citation count and altmetrics, and illustrates its potential to examine the role of indicators within the relevance evaluation process as well as the relation of indicators and relevance criteria. It provides an in-depth analysis of the relevance evaluation behavior of two participants chosen as examples against the background of their disciplines, social and life sciences.

Keywords

Relevance evaluation, relevance criteria, bibliometric indicators, altmetrics, scientific search

1. Introduction

Search engines play an essential role for researchers in finding scientific literature. There are numerous studies on relevance criteria or relevance clues, i. e. the aspects that impact researchers' relevance assessment of a document. Schamber [45] presents a table of eighty relevance factors suggested from the literature, noticing a high overlap in spite of the variety of research approaches and information environments. However, most of the studies reviewed in [45] focus on criteria use by expert relevance judges as part of information retrieval effectiveness studies, not on real users. This is different in [7]. Barry & Schamber compare two of their previous studies on researchers and students from social sciences and humanities [6] and users of weather information [44]. They find a high overlap of used relevance criteria despite differences in types of users and documents, thus suggesting that a common set of relevance criteria exists independently of users and information systems. Cool et al. [13] studied computer science students and humanities scholars. In line with [7], they found an overlap between the two groups, but also significant differences in dependence of the user's situation (knowledge, goals etc). Among more recent research, [33] studied the relevance criteria applied by PubMed users. They found that a topical match was the most important selection criterion, followed by the year of publication, a preference of reviews, and the journal's quality or reputation. Further, [43] provides an extensive overview supporting the hypothesis of a common set of relevance criteria. The studies covered in the reviews and mentioned here have in common that they address the criteria, but to a lesser extent which parts of a document or document representation are used to assess these criteria. This is also the case for the branch of literature dealing with models and theory of relevance, such as the literature reviewed in [35; 42; 43]. An exception is [48]. Wang & Soergel developed a document selection model, in which document information elements (title, authors, journal etc.) are processed to judge relevance criteria (topicality, quality, novelty, …), which are in turn applied to assess document values (e. g. functional or social values) to form the basis for the selection decision.

Most of the relevance criteria studies are from the 1990's and early 2000's. Since then interfaces have changed, and while these studies layed the foundation for our understanding of criteria used for relevance evaluation, it is unclear if and to what extent interface changes also encourage changes in search behavior and weighting of relevance criteria. One of these changes is that nowadays search

BIR 2021: 11th International Workshop on Bibliometric-enhanced Information Retrieval at ECIR 2021, April 1, 2021, online
EMAIL: jacqueline.sachse@hu-berlin.de
ORCID: 0000-0003-2587-4305

results are often enriched with bibliometric indicators. Most common is the citation count, used for instance by Google Scholar, Web of Science and Scopus. In consideration of the important role of citation-based measures in research assessment and long history as indicators of scientific impact, the question rises whether the number of citations influences the human relevance evaluation process. How much of an impact and what kind of impact do citation counts have on researcher's assessments of relevance?

Some search engines also include indicators that are not derived from traditional scholarly communication. Dimensions and 1Findr show altmetrics on their search engine results pages (SERP). Altmetrics count mentions of scholarly work in social media, mainstream media and other online platforms. Due to the heterogeneity of altmetric data sources a common definition is lacking [24]. Altmetrics do neither have citations' tradition nor familiarity among researchers, and, also due to their heterogeneity, their meaning may be considered even more vague, thus impeding their interpretability. In view of these differences between citations and altmetrics the question is whether both indicators fulfill different roles for relevance evaluation or whether their common denominator – the fact that they are both numbers – makes them have the same kind of impact on relevance evaluation. An online ranking experiment by Lemke et al. [30] showed that researchers preferred bibliometric indicators over usage metrics and altmetrics when deciding what to read. However, in the experiment participants ranked fictitious publications solely based on six indicators (citation counts, journal impact factor, h-index, download count, tweets, Mendeley readers) without any other information provided. Although the results reveal an interesting trend, they do not allow conclusions about the prevalence of indicator use in a natural search setting with many other bibliographic information available. In this regard, this study aims at answering the following research questions:

RQ1: What role do indicators play within the relevance evaluation process?
RQ2: Which relevance criteria are related to the use of indicators?
RQ3: To what extent does behavior vary between different indicators?

In order to approach these questions, an online experiment was conducted. As a starting point this paper focuses on RQ1 and RQ2 and presents an analysis of two participants from the social and life sciences as example cases. Since not all data is available for analysis yet, RQ3 will be part of future work. In the next section the methodology of the experiment will be elaborated including a detailed description of the stimuli as well as a coding scheme of relevance criteria, factors and properties to be used in the analysis of participants' evaluation process. The third section summarizes preliminary results based on the analysis of two individual cases in order to illustrate the potential of the study design with a special focus on the effect of structural factors such as disciplinary background in the selection process. The paper concludes with a brief overview of such analysis potential and provides an outlook on further analyses to be conducted.

2. Methodology

2.1. Stimulus

In an experimental within-subject design either an altmetric score, a citation count or no indicator was attached to the search results of a fictitious academic SERP. Participants performed searches for given scenarios, one per experimental condition (citations, altmetrics, no indicator). A SERP contains ten results with links to a detail page providing the abstract. The search results were retrieved from Google Scholar. The original ranking was not altered. No second result pages are provided and it is not possible to re-query or re-sort the results. An indicator is attached to each entry of the SERP (cf. Figure 1), except in the no-indicator condition. These indicators do not represent the article's actual indicator values. To compare if high-score results receive more, less or equal attention in dependence of the experimental condition, the ranking bias – the tendency to favor documents early in the result list – needs to be controlled. A strong influence of the ranking bias has been shown especially for Web search [17; 18; 27; 28; 29; 32; 39; 41; 46], but a similar order effect on expert relevance judgments is observed according to [43; 51]. The ranking bias is controlled by placing high scores on second, fifth and seventh ranks and let the other results show a comparably low score.

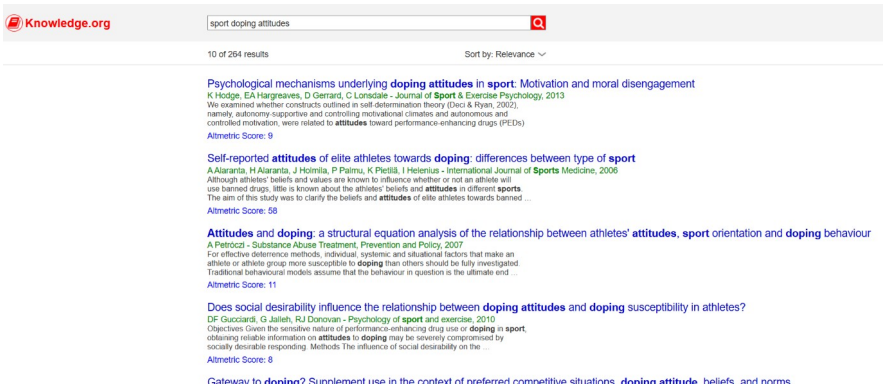

Figure 1: Search engine results page

2.2. Participants

Previous research showed that a user's disciplinary background influences the weighting of relevance criteria [43]. The user's discipline can be expected to be an influential factor especially in the context of indicators, because traditional citation analysis is less applicable in the social sciences and humanities due to an insufficient journal coverage, a stronger book culture and a higher tendency to publish in native languages [3; 11; 19; 20; 24; 36; 38; 40]. Furthermore, researchers in the social sciences and humanities tend to be more critical towards bibliometric indicators [19; 21; 22; 26; 31; 38]. In contrast, the natural and life sciences generally show a more positive attitude and stronger usage of indicators [1; 2; 5; 10; 14]. Participants were selected in view of these disciplinary differences and recruited via personal e-mail invitation sent to members of Berlin-based research facilities. 50 participants (22 females) took part in this study. 25 participants had a background in medicine-related fields, and 25 in the social sciences, especially sociology. 31 participants had a doctoral degree, 11 were doctoral students. The study sessions were conducted in German or English.

2.3. Procedure

The study was conducted online using a video conference tool. After a training task, participants performed three search tasks while sharing their screen with the investigator. They were instructed to choose up to three documents from the SERP and to prioritize these according to their relevance. Participants were encouraged to think-aloud. Since domain knowledge impacts search behavior [12; 15; 43; 47], after each task participants were asked to assess their topical familiarity on a 5-point Likert scale. Likewise, they assessed the perceived level of difficulty in evaluating the results' relevance in order to examine if indicators facilitate the selection process. After the tasks were completed an interview was conducted. Participants were encouraged to reflect their criteria use by asking them about their usual search behavior. Further, they were asked about their use of indicators in this search and in general as well as their attitude towards indicators. To examine whether there is a relation between social media use, attitudes towards and use of altmetrics, participants were also asked about their usage of online services. The average duration time of a session was 90 minutes. Each session was recorded and transcribed, including think-aloud and click behavior.

2.4. Tasks

Broad literature search is usually performed at an early stage of the research process to gain an overview of a new topic [4; 9; 16; 25; 34; 37; 49]. Hence, every task asked participants to become acquainted with a topic that is broad and easy enough to grasp for novices and only loosely related to participants' disciplines. Topics were vegetarianism from a sociological perspective, attitudes of athletes towards doping, and use of big-data applications in health care. To determine the order of tasks, experimental conditions and their combinations, a Graeco-Latin square design was used, leading to 36 different study configurations to which participants were randomly assigned.

2.5. Coding scheme

To address RQ1 and RQ2, the transcripts are coded using MaxQDA with regard to a model by Behnert [8], which takes up the essential differentiation between relevance criteria and document information elements of the document selection model by Wang & Soergel [48]. The basic idea is that users evaluate the relevance of an article with respect to *relevance criteria* by means of *relevance properties*. The evaluation process is further influenced by *relevance factors*. For instance, a user assesses the topicality, method and reputation of an article (*criteria*) by examining the title, the abstract and the name of the journal (*properties*). The users' choice to focus on these criteria and the way they assess and weigh them against one another, is influenced by e. g. their personal knowledge level and the article's rank within the SERP (*factors*). To create a useful top-down coding scheme with regard to RQ1 and RQ2, the reported relevance criteria found in previous research, especially [6; 7; 43; 48], were merged and grouped according to their underlying concepts, and finally sorted into the threefold model of relevance criteria, factors and properties. Lastly, the properties *altmetric score* and *citations* were added (cf. Table 1). This top-down coding scheme is open to bottom-up extensions during the coding process.

Table 1
Coding scheme of the relevance evaluation behavior

Relevance criteria	Relevance factors	Relevance properties
quality	knowledge level	abstract
validity	content novelty	authors
accuracy	document novelty	journal
source reputation	source novelty	publication date
topic	understanding	title
depth	mental effort	language
scope	personal credence	snippet
recency	affectiveness	search terms
clarity	interface	altmetric score
type	rank	citations
method	time constraints	
theory	task constraints	
empirical	task difficulty	
accessibility		
affordability		
importance		

3. Results

The transcription and coding process is still ongoing. For the purpose of demonstrating the potential of the described experimental study design, this paper reports on preliminary analyses of two participants, chosen as examples against the background of their discipline. One is a social scientist (P17), the other a life scientist (P28). Both participants are male, 37 years old and hold a doctoral degree.

Table 2 displays the frequency of the codes applied on P17's transcript, thus showing which aspects he focuses on during relevance evaluation. Method, scope and topic are the criteria most often mentioned. He relies heavily on information provided by the abstract, but also gains information from the title and the journal. Moreover, it is important to him to find new content, which strongly depends on his own knowledge level. Another influencing factor is his affectiveness, which is defined here as the extent to which the user exhibits an affective or emotional response to the provided information. In the case of P17, this specifically means that he finds articles relevant if they meet his personal interests, if he "likes" them. In this case he would also take into account if there was no exact topical

match, as the following remark illustrates: "I would probably read it, because I find it interesting. […] But more out of some interest, and less topic specific."

Table 2
Relevance criteria, factors and properties used by P17

Relevance criteria	Relevance factors	Relevance properties
method (25)	content novelty (11)	abstract (46)
scope (22)	knowledge level (8)	title (19)
topic (14)	affectiveness (7)	journal (18)
empirical (7)	mental effort (2)	publication date (10)
recency (6), theory (6)	personal credence (2)	language (1), search terms (1)
clarity (4)		
depth (3), validity (3)		
quality (2)		
type (1)		

The relevance evaluation process of P17 follows recurring patterns. They emerge when not only the frequency of the codes, but their co-occurrence and chronological sequence are considered. P17 generally starts at the SERP with filtering for the method and scope of an article by looking for certain keywords in the title, but also the journal. If he then decides to click on a result, he starts reading the abstract to get a more detailed impression of the articles' method, scope and topic. To P17, it increases an article's relevance if empirical data is provided or if some aspect or idea is described that is perceived as new. Sometimes he also considers the article's recency or puts the content novelty in context of the publication date. He never mentioned the indicators, neither the number of citations nor the altmetric score, nor did he express noticing the interface changes between the experimental conditions.

The life scientist P28 focuses primarily on scope (cf. Table 3), thus a topic-related criterion as can be expected. Scope alone, however, is not sufficient to him. Type and validity and to a lesser extent method are frequently mentioned criteria. In the case of P28, type and validity are referring to a strong preference of P28 for reviews over original articles and a strong focus on the sample size of the examined studies. There are only a few times factors are influencing P28's relevance perception. He uses a quite diverse set of properties to gather information about the relevance criteria. Apart from the abstract, P28 interacts frequently with the title. Sometimes he considers the citations, the journal or the publication date.

Table 3
Relevance criteria, factors and properties used by P28

Relevance criteria	Relevance factors	Relevance properties
scope (23)	task difficulty (3)	abstract (29)
type (14), validity (14)	personal credence (1)	title (24)
method (11)		citations (4), journal (4),
topic (8)		publication date (4)
recency (4)		search terms (3), snippet (3)
empirical (2), topicality (2)		
clarity (1), importance (1),		
source reputation (1)		

Regarding P28's typical relevance evaluation process, taking the code co-occurrence and chronological sequence into account, we find that P28 first scans the SERP for articles matching his thematic scope, but is at the same time eager to find reviews, which report on numerous studies and data sets, or broad studies. He mostly uses the title, but sometimes also the snippet and citations to identify articles of interest. After this filtering, he reads the selected abstracts focusing strongly on the method and validity of a study, especially the size of the data set. Indicators, specifically the citation

count, are an explicit part of P28's search strategy. For instance, he uses citations to estimate whether an article is a review, assuming that reviews are highly cited. The search task also required participants to prioritize selected results according to their relevance. During this ordering process, he started comparing the citations. However, although he noticed that one article he considered least relevant (among the selected ones) was actually more often cited than the others, this did not change his own prioritization. During P28's last search task, there were no indicators displayed. He took notice of this circumstance after a few minutes, and this made him reflect about what he uses citations for. In P28's opinion, citations indicate the importance of an article within the scientific community: "Here [I] cannot rely on how many people have cited this and by that found it important at some point. […] Of course the more citations [an article has], the more central [it is]. By now I am looking for central things, and this help is not there now."

Comparing the relevance evaluation behavior of the two researchers from different fields, it is apparent that both participants' first concern is to look for topic matches within the search results. This can be expected and is in line with previous research stating that topic- and content-oriented criteria are among the most determining relevance criteria [33; 43; 48; 50]. Apart from this, relevance evaluation behavior is more diverse: The social scientist P17 focuses strongly on the content of an article and what kind of discourse it belongs to, whereas the interest of the life scientist P28 lies mostly in a study's validity. There are less relevance factors influencing P28's relevance perception. Instead, he uses a more diverse set of relevance properties than P17. P17 spends most time reading the abstracts, while P28 engages more with prefiltering on the SERP before clicking and reading the abstracts. The results further support the assumption that users' knowledge level plays a role in the relevance evaluation process as indicated by previous research [12; 15; 43; 47]. When working on tasks where participants self-assessed a high topical familiarity, they adjust their behavior differently. P17 uses his knowledge to better locate an article within a scientific discourse or theoretical tradition, whereas P28's familiarity seems to refer mainly to the journals and thus he uses the criterion source reputation to perform a finer filtering on the SERP level. Furthermore, participants applied different weights to their own opinion. P17's relevance evaluation was often influenced by the extent to which he exhibited an affective response, and he chose not to select articles when he did not agree with information presented there. P28 mentioned his own opinion just once. Nevertheless, in that case, although he strongly agreed to one of the findings, he did not select the respective article, because he felt the scope was not close enough to the search task. Lastly, regarding their use of indicators, while P17 did not pay any attention to indicators, P28 described the number of citations as a useful indicator of importance within the community. Hence, especially high citations attract his attention, but they are not important enough to override a not-relevant assessment of a highly-cited article based on scope, validity and other criteria.

4. Conclusion

This paper illustrated the potential of an experimental study design to examine the role of indicators within the relevance evaluation process (RQ1) as well as the relevance criteria that are related to the use of indicators (RQ2). It provided an in-depth analysis of the relevance evaluation behavior of two participants from the social and the life sciences as example cases. Overall, even though the analyses and results being yet preliminary in nature, they illustrate the benefits of, firstly, using a relevance model that distinguishes between relevance criteria, factors and properties, and secondly, an analysis focusing on patterns in the sequentiality and co-occurrence of relevance codes.

Results support the hypothesis that different usage of relevance criteria, factors and properties are related to the epistemic cultures of scientific disciplines. Typically, life sciences are more data oriented and thus focus more strongly on indicators than the social sciences which prefer more qualitative approaches. The analysis of all 50 participants will show if this hypothesis holds. In addition, the analysis of the post-search interview material, where participants expressed their attitude towards indicators in general and with regard to their search behavior, will be used to produce a more comprehensive picture of the relation of relevance criteria and indicators. Another approach to study the impact of indicators on relevance evaluation is not by qualitatively analyzing individuals, but with a quantitative analysis on the level of the experimental conditions (citations, altmetrics, no indicators) across participants, to see whether results with higher indicator values are more often clicked or

mentioned than results with low values, or if they are more often selected as relevant. Another focus will be on the differences between citations and altmetrics (RQ3), in terms of both their use for relevance evaluation and researchers attitudes towards their importance as impact indicators.

By addressing these questions, I hope to contribute to the extension of our understanding of relevance criteria, especially since indicators have become a crucial, but yet understudied part of scientific search interfaces.

5. Acknowledgments

This research has been funded under the research group grant “Reflexive Metrics” (FKZ 01PQ17002) by the German Federal Ministry of Education and Research (BMBF).

6. References

[1] D. W. Aksnes, Citation rates and perceptions of scientific contribution, Journal of the American Society for Information Science and Technology 57(2) (2006) 169–185. doi:10.1002/asi.20262

[2] D. W. Aksnes, A. Rip, Researchers’ perceptions of citations, Research Policy 38(6) (2009) 895–905. doi:10.1016/j.respol.2009.02.001

[3] D. W. Aksnes, G. Sivertsen, A criteria-based assessment of the coverage of Scopus and Web of Science, in: Proceedings of the 23rd International Conference on Science and Technology Indicators, STI 2018, pp. 707–716. URL: https://openaccess.leidenuniv.nl/handle/1887/65211

[4] S. Antonijević, E. S. Cahoy, Personal Library Curation: An Ethnographic Study of Scholars’ Information Practices, Portal: Libraries and the Academy 14(2) (2014) 287–306. doi:10.1353/pla.2014.0010

[5] H. H. Aung, H. Zheng, M. Erdt, A. S. Aw, S. J. Sin, Y. Theng, Investigating familiarity and usage of traditional metrics and altmetrics, Journal of the Association for Information Science and Technology 70(8) (2019) 872–887. doi:10.1002/asi.24162

[6] C. L. Barry, User-defined relevance criteria: An exploratory study, Journal of the American Society for Information Science 45(3) (1994) 149–159. doi:10.1002/(SICI)1097-4571(199404)45:3<149::AID-ASI5>3.0.CO;2-J

[7] C. L. Barry, L. Schamber, Users’ criteria for relevance evaluation: A cross-situational comparison, Information Processing & Management 34(2–3) (1998) 219–236. doi:10.1016/S0306-4573(97)00078-2

[8] C. Behnert, Kriterien und Einflussfaktoren bei der Relevanzbewertung von Surrogaten in akademischen Informationssystemen, Information – Wissenschaft & Praxis 70(1) (2019) 24–32. doi:10.1515/iwp-2019-0002

[9] I. Bøyum, S. Aabø, The information practices of Business PhD students, New Library World 116(3/4) (2015) 187–200. doi:10.1108/NLW-06-2014-0073

[10] G. Buela-Casal, I. Zych, What do the scientists think about the impact factor?, Scientometrics 92(2) (2012) 281–292. doi:10.1007/s11192-012-0676-y

[11] D. A. Chavarro Bohórquez, Universalism and Particularism: Explaining the Emergence and Growth of Regional Journal Indexing Systems, Ph.D. thesis, University of Sussex, GB, 2017.

[12] M. J. Cole, X. Zhang, C. Liu, N. J. Belkin, J. Gwizdka, Knowledge effects on document selection in search results pages, in: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, SIGIR 2011, pp. 1219–1220. doi:10.1145/2009916.2010128

[13] C. Cool, N. J. Belkin, O. Frieder, P. Kantor, Characteristics of texts affecting relevance judgments, in: Proceedings of the 14th National Online Meeting, 1993, pp. 77–84.

[14] G. E. Derrick, J. Gillespie, "A number you just can’t get away from”: Characteristics of Adoption and the Social Construction of Metric Use by Researchers, in: S. Hinze, A. Lottmann (Eds.), Translational twists and turns: Science as a socio-economic endeavour, Berlin, Germany, 2013, pp. 104–116. URL: https://eprints.lancs.ac.uk/id/eprint/88786/

[15] J. Dinet, J. M. C. Bastien, M. Kitajima, What, where and how are young people looking for in a search engine results page? Impact of typographical cues and prior domain knowledge, in:

Proceedings of the 22nd Conference on l'Interaction Homme-Machine, 2010, pp. 105–112. doi:10.1145/1941007.1941022
[16] J. T. Du, N. Evans, Academic Users' Information Searching on Research Topics: Characteristics of Research Tasks and Search Strategies, The Journal of Academic Librarianship 37(4) (2011) 299–306. doi:10.1016/j.acalib.2011.04.003
[17] L. A. Granka, T. Joachims, G. Gay, Eye-tracking analysis of user behavior in WWW search, in: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR 2004, pp. 478–479. doi:10.1145/1008992.1009079
[18] Z. Guan, E. Cutrell, An eye tracking study of the effect of target rank on web search, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, SIGCHI 2007, pp. 417–420. doi:10.1145/1240624.1240691
[19] G. Haddow, B. Hammarfelt, Quality, impact, and quantification: Indicators and metrics use by social scientists, Journal of the Association for Information Science and Technology 70(1) (2019) 16–26. doi:10.1002/asi.24097
[20] B. Hammarfelt, Beyond Coverage: Toward a Bibliometrics for the Humanities, in: M. Ochsner, S. E. Hug, H.-D. Daniel (Eds.), Research Assessment in the Humanities: Towards Criteria and Procedures, Springer, Cham, 2016, pp. 115–131. doi:10.1007/978-3-319-29016-4_10
[21] B. Hammarfelt, S. de Rijcke, Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University, Research Evaluation 24(1) (2015) 63–77. doi:10.1093/reseval/rvu029
[22] B. Hammarfelt, G. Haddow, Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators, Journal of the Association for Information Science and Technology 69(7) (2018) 924–935. doi:10.1002/asi.24043
[23] S. Haustein, Grand challenges in altmetrics: Heterogeneity, data quality and dependencies, Scientometrics 108(1) (2016) 413–423. doi:10.1007/s11192-016-1910-9
[24] D. Hicks, The Four Literatures of Social Science, in: H. F. Moed, W. Glänzel, U. Schmoch (Eds.), Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems, Springer Netherlands, 2005, pp. 473–496. doi:10.1007/1-4020-2755-9_22
[25] C.-T. Hsin, Y.-H. Cheng, C.-C Tsai, Searching and sourcing online academic literature: Comparisons of doctoral students and junior faculty in education, Online Information Review 40(7) (2016) 979–997. doi:10.1108/OIR-11-2015-0354
[26] S. E. Hug, M. Ochsner, H.-D. Daniel, Criteria for assessing research quality in the humanities: A Delphi study among scholars of English literature, German literature and art history, Research Evaluation 22(5) (2013) 369–383. doi:10.1093/reseval/rvt008
[27] T. Joachims, L. Granka, B. Pan, H. Hembrooke, G. Gay, Accurately Interpreting Clickthrough Data as Implicit Feedback, ACM SIGIR Forum 51 (2005) 4–11. doi:10.1145/3130332.3130334
[28] M. T. Keane, M. O'Brien, B. Smyth, Are people biased in their use of search engines?, Communications of the ACM 51(2) (2008) 49–52. doi:10.1145/1314215.1314224
[29] K. Klöckner, N. Wirschum, A. Jameson, Depth- and breadth-first processing of search result lists, in: CHI '04 Extended Abstracts on Human Factors in Computing Systems, CHI 2004, p. 1539. doi:10.1145/985921.986115
[30] S. Lemke, A. Mazarakis, I. Peters, Conjoint analysis of researchers' hidden preferences for bibliometrics, altmetrics, and usage metrics, Journal of the Association for Information Science and Technology January (2021). doi:10.1002/asi.24445
[31] S. Lemke, M. Mehrazar, A. Mazarakis, I. Peters, "When You Use Social Media You Are Not Working": Barriers for the Use of Metrics in Social Sciences, Frontiers in Research Metrics and Analytics 3 Art. 39 (2019). doi:10.3389/frma.2018.00039
[32] L. Lorigo, M. Haridasan, H. Brynjarsdóttir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, B. Pan, Eye tracking and online search: Lessons learned and challenges ahead, Journal of the American Society for Information Science and Technology 59(7) (2008) 1041–1052. doi:10.1002/asi.20794
[33] M. Macedo-Rouet, J.-F. Rouet, C. Ros, N. Vibert, How do scientists select articles in the PubMed database? An empirical study of criteria and strategies, European Review of Applied Psychology 62(2) (2012) 63–72. doi:10.1016/j.erap.2012.01.003

[34] L. I. Meho, H. R. Tibbo, Modeling the information-seeking behavior of social scientists: Ellis's study revisited, Journal of the American Society for Information Science and Technology 54(6) (2003) 570–587. doi:10.1002/asi.10244
[35] S. Mizzaro, Relevance: The whole history, Journal of the American Society for Information Science 48(9) (1998) 810–832. doi:10.1002/(SICI)1097-4571(199709)48:9<810::AID-ASI6>3.0.CO;2-U
[36] P. Mongeon, A. Paul-Hus, The journal coverage of Web of Science and Scopus: A comparative analysis, Scientometrics 106(1) (2016) 213–228. doi:10.1007/s11192-015-1765-5
[37] J. Murphy, Information-seeking habits of environmental scientists: A study of interdisciplinary scientists at the U.S. Environmental Protection Agency in Research Triangle Park, North Carolina, Master's thesis, University of North Carolina, N.C., US, 2003.
[38] A. J. Nederhof, Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A review, Scientometrics 66(1) (2006) 81–100. doi:10.1007/s11192-006-0007-2
[39] M. O'Brien, M. T. Keane, Modeling Result-List Searching in the World Wide Web: The Role of Relevance Topologies and Trust Bias, in: Proceedings of the 28th Annual Conference of the Cognitive Science Society, 2006, pp. 1881–1886.
[40] T. L. B. Ossenblok, T. C. E. Engels, G. Sivertsen, The representation of the social sciences and humanities in the Web of Science. A comparison of publication patterns and incentive structures in Flanders and Norway (2005–2009), Research Evaluation 21(4) (2012) 280–290. doi:0.1093/reseval/rvs019
[41] B. Pan, H. Hembrooke, T. Joachims, L. Lorigo, G. Gay, L. Granka, In Google We Trust: Users' Decisions on Rank, Position, and Relevance, Journal of Computer-Mediated Communication 12(3) (2007) 801–823. doi:10.1111/j.1083-6101.2007.00351.x
[42] T. Saracevic, Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance, Journal of the American Society for Information Science and Technology 58(13) (2007) 1915–1933. doi:10.1002/asi.20682
[43] T. Saracevic, The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really?, Synthesis Lectures on Information Concepts, Retrieval, and Services 8(3) (2016), i–109. doi:10.2200/S00723ED1V01Y201607ICR050
[44] L. Schamber, Users' Criteria for Evaluation in a Multimedia Environment, in: Proceedings of the ASIS Annual Meeting, 28, 1991, pp. 126–133.
[45] L. Schamber, Relevance and Information Behavior, Annual Review of Information Science and Technology (ARIST) 29 (1994) 3–48.
[46] S. Schultheiß, S. Sünkler, D. Lewandoski, We still trust in Google, but less than 10 years ago: An eye-tracking study, Information Research 23(3) paper 799 (2018). URL: http://informationr.net/ir/23-3/paper799.html
[47] A. Sihvonen, P. Vakkari, Subject knowledge improves interactive query expansion assisted by a thesaurus, Journal of Documentation 60(6) (2004) 673–690. doi:10.1108/00220410410568151
[48] P. Wang, D. Soergel, A cognitive model of document use during a research project. Study I. Document selection, Journal of the American Society for Information Science 49(2) (1998) 115–133. doi:10.1002/(SICI)1097-4571(199802)49:2<115::AID-ASI3>3.0.CO;2-T
[49] L. Westbrook, Information Needs and Experiences of Scholars in Women's Studies: Problems and Solutions, College & Research Libraries 64(3) (2003) 192–209. doi:10.5860/crl.64.3.192
[50] Y. Xu, Z. Chen, Relevance judgment: What do information users consider beyond topicality?, Journal of the American Society for Information Science and Technology 57(7) (2006) 961–973. doi:10.1002/asi.20361
[51] Y. Xu, D. Wang, Order effect in relevance judgment, Journal of the American Society for Information Science and Technology 59(8) (2008) 1264–1275. doi:10.1002/asi.20826