Zeitschrift für Erziehungswissenschaft (2026) 29:73–90
https://doi.org/10.1007/s11618-025-01344-w

SCHWERPUNKT

Assessing the information quality of online sources used by first year students for solving generic critical online reasoning tasks

Tobias Scherer · Alice Laufer · Marcus Maurer · Christian Schemer

Received: 23 September 2024 / Revised: 10 March 2025 / Accepted: 29 May 2025 / Published online: 14 October 2025

Abstract University students are increasingly using online media for learning purposes. However, the information presented there varies greatly in terms of quality. Previous studies have often examined the quality of online sources based on a simplified rating but have not explicitly studied the specific content of those sources or taken a multidimensional view of quality. The present study goes beyond previous research and, for the first time, explicitly examined the information quality of website content used by university students. Our results show that students use a variety of different online sources. Differences in quality become apparent regarding the *completeness, accuracy,* and *balance* of the content. This suggests that students are not always able to evaluate online content appropriately. Differences in quality based on the individual types of sources were identified, but not for all types of sources. Instead of the previously used source heuristics, a more detailed approach is therefore appropriate, examining the content of a website in detail, in order to make statements about the information quality of an online source.

✉ Tobias Scherer, M.A. · Prof. Dr. Marcus Maurer
Institut für Publizistik, Politische Kommunikation, Johannes Gutenberg-Universität Mainz, Jakob-Welder-Weg 12, 55128 Mainz, Germany
E-Mail: tobias.scherer@uni-mainz.de

Prof. Dr. Marcus Maurer
E-Mail: mmaurer@uni-mainz.de

Dr. Alice Laufer · Prof. Dr. Christian Schemer
Institut für Publizistik, DySCo Research Group, Johannes Gutenberg-Universität Mainz, Jakob-Welder-Weg 12, 55128 Mainz, Germany

Dr. Alice Laufer
E-Mail: alaufer@uni-mainz.de

Prof. Dr. Christian Schemer
E-Mail: schemer@uni-mainz.de

Keywords Information quality · Online information landscape · Critical Online Reasoning

Die Bewertung der Informationsqualität von Online-Quellen, die von Studienanfängern zur Lösung von generischen Critical Online Reasoning-Aufgaben genutzt werden

Zusammenfassung Studierende nutzen häufig Online-Medien zu Lernzwecken. Die dort präsentierten Informationen unterscheiden sich allerdings stark hinsichtlich ihrer Qualität. Bisherige Studien haben Online-Quellen oft anhand grober Qualitäts-Ratings untersucht, sich dabei jedoch nicht auf die Inhalte der Quellen fokussiert. Mit Hilfe einer quantitativen Inhaltsanalyse hat diese Studie die bisherige Quellen-Heuristik erweitert und erstmals die Informationsqualität der von Studierenden genutzten Webseiten analysiert. Unsere Ergebnisse zeigen, dass Studierende eine Vielzahl an unterschiedlichen Online-Quellen benutzen. Im Hinblick auf *Vollständigkeit*, *Korrektheit* und *Ausgewogenheit* der Inhalte werden Qualitätsunterschiede ersichtlich. Das spricht dafür, dass Studierende nicht immer in der Lage sind, Online-Inhalte entsprechend zu evaluieren. Signifikante Qualitätsunterschiede ließen sich nur für einzelne Webseiten-Arten beobachten. Statt der bisher verwendeten Quellenheuristik ist somit ein detaillierteres Vorgehen angemessen, welches die Inhalte einer Webseite genau untersucht, um Aussagen über die Informationsqualität einer Online-Quelle zu treffen.

Schlüsselwörter Informationsqualität · Online-Informationslandschaft · Critical Online Reasoning

1 Introduction

Recent studies indicate that university students primarily turn to online sources such as news media, social media, or blogs when completing study related tasks (Bartlett et al. 2020; Weber et al. 2018). Since the information in these sources is often not subject to any scientific control, and research on the quality of news and social media suggests that their content can be biased or inaccurate (e.g., Garz 2014; Reinemann et al. 2024), it can be assumed that the use of such sources may have negative effects (Madge et al. 2009; Nagel et al. 2020; Schemer et al. 2022). Previous studies on the use of online sources for student learning have focused on subjective ratings of websites, based on credibility research (Flanagin and Metzger 2007). However, this is a subjective attribution, which is rather caused by user characteristics than by actual website content and its information quality (Hoorn and van Wijngaarden 2010). More detailed studies have categorized sources into groups using an ordinal classification, ranging from high information quality for scientific publications to moderate quality for news media websites, or low information quality for social media content (Nagel et al. 2020). However, also this heuristic can be misleading since it fails to consider differences within types of sources. For example, it

overlooks that even academic publications can lack quality (Bohannon 2013). The present research addresses these shortcomings by explicitly studying the quality of information content used by first-year university students when solving an information problem. Unlike prior studies, we coded website quality based on established dimensions of information quality as used in communication science (e.g., Schatz and Schulz 1992), providing a more in-depth understanding of the online content that university students access when solving generic information problems.

2 Theoretical background

2.1 University students' information processing

Information processing and problem solving depend on individual factors such as cognitive capacity (Brünken et al. 2019) or prior knowledge (Renkl 2020). Many models exist that try to visualize and segmentize those processes in order to examine differences, e.g. the *Heuristic-Systematic Model* (Chen and Chaiken 1999) or the *Information Problem-Solving Model regarding Internet usage* (IPS-I) (Brand-Gruwel et al. 2009). As a way to measure potential differences in the ability to critically evaluate online information when solving an information problem, we refer to the so-called construct of Critical Online Reasoning (COR) (Molerov et al. 2020). COR was developed based on related constructs such as multiple source use and understanding (Braasch et al. 2018; Goldman and Brand-Gruwel 2018), internet credibility (Wierzbicki 2018) or the IPS-I (Brand-Gruwel et al. 2009) and has already been empirically tested (Nagel et al. 2022). It describes the skills that students need to solve an online information problem. They can be divided into three interrelated processes. First, searching and initial assessment of online information, second, evaluation of sources based on website characteristics and the corresponding selection of sources, and third, summarizing the information to form a convincing argument to solve the initial information problem (Molerov et al. 2020; Nagel et al. 2020). This approach usually involves synthesizing different pieces of information and critically handling and weighing up potentially conflicting information (Goldman and Brand-Gruwel 2018). The information problems to be solved can be divided into two subject areas: The generic COR (GEN-COR) addresses a general, everyday information problem, whereas the domain-specific DOM-COR focuses on domain-specific problems of students, e.g., in economics or medicine. This article centers on websites used to solve generic information problems. As our study is exclusively dedicated to the information quality of the websites used, the COR abilities of the participants are not analyzed.

2.2 Media use of first-year university students

The internet has become one of the most utilized resources for information search in everyday life and for school and learning purposes (Nagel et al. 2020, Walraven et al. 2009). Higher education students utilize various online sources, such as news websites and social media (Maurer et al. 2019b; Madge et al. 2009), although the

lack of quality control can impede or undermine students' learning (Rieh and Belkin 1998). Since anyone can publish and edit information on the internet, bypassing traditional gatekeepers, such as journalists or scientific reviewers (Hoorn and van Wijngaarden 2010; Olteanu et al. 2013), this information can be biased or false (Kelly 2019; Nagel et al. 2020) and its use can lead to misconceptions (Maurer and Reinemann 2006; Schemer et al. 2022).

Additionally, first-year university students often lack the necessary skills to critically evaluate information quality (Brand-Gruwel et al. 2017; Weber et al. 2018). Furthermore, students perform poorly in evaluating the trustworthiness and accuracy of the media content they consume (Brox 2012). Because of the vast amount of information available (Hughes 2013), they tend to rely on sources they consider well-known and reliable in their online research, such as Wikipedia (Nagel et al. 2020). They also tend to rely on familiar brand images of online sources (Urban and Schweiger 2014). However, it remains to be seen whether the reliance on already known sources instead of in-depth research results in exposure to websites with superior information quality.

3 Website quality evaluation criteria and differences in quality

Previous studies used different approaches to examine the quality of online information, for example rating sources as "good" or "bad" based on the type of source (Nagel et al. 2020; Zlatkin-Troitschanskaia et al. 2021). However, it is controversial which aspects constitute information quality at all. Hoorn and van Wijngaarden (2010) classified information quality as an umbrella term that encompassed aspects such as correctness and readability. Other studies have relied on dimensions, such as believability, accuracy, and bias to assess information quality (Johnson and Kaye 1998). Flanagin and Metzger (2000) measured information quality based on believability, accuracy, trustworthiness, bias, and completeness. However, similar to Johnson and Kaye (1998), they used a survey approach encompassing subjective quality assessments instead of an intersubjective method to evaluate website content.

In communication science, however, there have long been established models for measuring the content quality of media coverage (McQuail 1992; Schatz and Schulz 1992). Although these were originally developed to especially measure the quality of news media coverage (news on TV or in newspapers), they can easily be transferred to other sources such as social media (Döring et al. 2023) or online encyclopedias (Döring et al. 2022). Even though the theoretical framework for measuring media quality is complex (Schatz and Schulz 1992), three dimensions stand out: completeness, accuracy, and balance.

Completeness is generally understood as the availability of all relevant information on a particular topic (McQuail 1992). Of course, it is hardly possible to include all relevant information in one article. The question is therefore whether an article or a website contains any relevant information at all and how much of the relevant information is included. In order to measure completeness, a list of all relevant information must be available in advance, which can then be used to determine

the proportion of information contained in an article in an input-output comparison (Tankard and Royal 2005). For online media, Saleh et al. (2022) showed that websites often exhibit incomplete information. Additionally, Basavakumar et al. (2019) found considerable differences between types of websites, with websites of non-profit organizations and news media most frequently exhibiting complete information, while commercial websites performed worse. There are also large differences in completeness for social media posts (Döring et al. 2023) and Wikipedia articles (Döring et al. 2022).

Accuracy is usually assessed by measuring whether media content correctly depicts reality (Schatz and Schulz 1992). Of course, whether media information is accurate can only be verified under certain conditions, for example by comparing it with external reality indicators, such as official statistics (Maurer et al. 2019a). To avoid this problem, some studies also consider the inclusion of (scientific) evidence as part of *accuracy* (Döring et al. 2023, 2022). News media coverage generally shows a relatively high level of accuracy. McNally et al. (2012) showed a rather high level of *accuracy* without major differences between website types, such as academic sources, commercial websites, and websites of non-profit organizations.

However, the content of news media can also be incorrect under certain circumstances (Allen et al. 2024). Previous studies on the content of websites that were classified as low-quality sources, such as YouTube (Nagel et al. 2020), found information to be mostly low in *accuracy* (Springer et al. 2020). However, Madathil et al. (2015) showed that many videos on YouTube conveyed correct information, e.g., science tutorials or TED conference presentations. Döring et al. (2023) supported these findings for content on social media. However, it is difficult to generalize findings on specific sources, e.g., YouTube channels, to the overall platform, i.e., YouTube. In a similar vein, Denniss et al. (2023) showed that although commercial websites, for example, had a lower level of *accuracy* than government websites or news media, this finding varied with the topic of the website.

Balance refers to the adequate distribution of perspectives, evaluations, or arguments in a text (Schatz and Schulz 1992). Although most news media should strive for balanced reporting, they pursue more or less pronounced editorial lines, which lead to media coverage containing a (political) bias that varies in strength depending on the medium (e.g., Maurer et al. 2024; Stark et al. 2021). Since many other sources of online information are not expected to provide balanced information, it can be assumed that the websites or social media accounts of political parties, interest groups or companies exhibit an even stronger bias. Consequently, research provides evidence of differences between and within platforms, e.g., social media posts differing in balance depending on the platform and author of the respective post (Döring et al. 2023). Moreover, biased content was even found in scientific studies (Hefferon and Miller 2020). Therefore, these inconsistency in findings call for a thorough examination of information content of a website. Generalizations across websites, platforms, or channels are insufficient to assess the quality of the information landscape online. Based on this reasoning we pose the following questions:

RQ1: Which types of online information sources do university students use to solve generic COR-tasks?

RQ2: Do these online information sources differ in terms of accuracy, completeness and balance?

RQ3: Are there differences in terms of accuracy, completeness and balance for different types of sources?

4 Methods

4.1 GEN-COR tasks and survey

Participants were 473 first-year university students recruited at different German universities. Four distinct tasks were developed by other participating projects as part of the CORE project (Zlatkin-Troitschanskaia et al. 2023). The tasks were randomly assigned to participants, ensuring that each participant completes at least one GEN-COR task. Each task presented a basic information search problem. For example, one of the tasks dealt with the issue of potentially dangerous radiation from a cell phone tower. In this "cell phone task", the subjects were in the fictional situation of looking for a flat with a friend. However, the friend did not want to move into a potential flat because there was a cell phone tower in the neighborhood. The respondents were therefore asked to write a statement, based on the sources they had researched, on whether such cell phone antennas could actually be a health hazard. They were not limited in terms of which websites they could use, and it was not predetermined that specific types of websites should be prioritized over others. All tasks were completed by participants using a virtual machine that tracked any browser activity. Additionally, after a set period, a screenshot of the active browser tab was captured for use in subsequent evaluations.

Participants had 20 min to solve each task, including both the online information search and the written response. After completing each task, participants were asked to paste at least two URLs of the sources that were most important to them for solving the task at hand. During the four tasks, the 473 participants mentioned 343 unique websites that they found especially important. These websites are the basis of our analysis.

5 Coding procedure

We utilized an existing online annotation tool that was adjusted to the present categories. This tool used the screenshots of the websites as a reference for coding, ensuring that the web pages remained in their original state. We conducted a quantitative content analysis to analyze the websites. This method allowed us to record the characteristics of large quantities of websites as objectively as possible. We developed a codebook as a standardized measurement instrument that defined and

helped to assess the individual website characteristics. Characteristics were thus assigned specific numerical codes for the respective categories. This allowed us to analyze the content of every website in the same manner and in a systematic and intersubjectively comprehensible way (Krippendorf 2018).

Typically, the entire website was considered during coding. Due to time constraints, it was assumed that participants were unlikely to read lengthy texts in full while completing tasks. As a result, for online lexicons like Wikipedia, only the article introduction was coded. In the case of online forums, only the first three forum posts were coded. Additionally, only the first five screenshots were used for coding given that students only scrolled lengthy pages and most frequently read only the first sections of lengthy texts on a website. For scientific articles, the gist of the content is always assessed by content analyzing the abstracts.

A team of seven trained coders was responsible for the assessment, with each coder manually evaluating a website in a randomized order. During training, we carried out reliability tests to check whether the coding was reliable. The reliability tests for each category were overall satisfactory, with some categories being tested more frequently than others due to initially weaker results ($N = 40$; 26; 14). Final coefficients (Holsti) ranged from 0.80 to perfect reliability.

The codebook consisted of two major parts. The first part focused on the general characteristics of the respective websites, while the second part focused on the textual elements on these websites that were related to information quality. Additional elements of a website, such as images and graphics, were coded as part of the website.

Type of source: This category refers to the website genre and included online news, scientific journal articles, entries in scientific databases, commercial websites, websites of interest groups, public authority websites, entries in online encyclopedias, and user comments. It builds on previous studies such as Nagel et al. (2020).

Completeness: This variable focused on whether the information was exhaustive and sufficient to complete the task. To achieve this, a specific coding process tailored to the different tasks was implemented. In cooperation with the developers of the respective tasks, we compiled a comprehensive list of topic elements considered relevant for the completion of each task. For the "cell phone" task, these included aspects such as "there are no known adverse health effects to date", and "other radiation exposures, e.g. from cellphones, are higher". Coding was then conducted based on this list, assessing how many of these topic elements were included on the website. Accordingly, percentages between 1 and 100 were used to code the completeness of the relevant information. For example, if two out of six required topic elements occurred on a website, its completeness score was rated 33%.

Accuracy: In a similar vein, we utilized a list of elements representing features of accurate information. This list contained keywords derived from sample solutions for the tasks. Here, we checked for factual correctness by determining as to whether the information on the website aligned with or contradicted the correct information elements from the sample solution. Insufficient accuracy was thus observed

if accurate elements were missing and/or if incorrect statements were included on the website. For the "cell phone task" this included false information such as saying that proven physical and psychological effects from cell phone tower radiation exist, or that cell phone tower radiation is generally dangerous. Coding was done on a scale ranging from 1 ("exclusively incorrect information") to 5 ("exclusively correct information"). Websites that did not contain any content related to the task were coded 0.

We also assessed whether the claims made on the website were supported by evidence, including specific references to sources backing the evidence. Specific references can include citations or links to primary sources. For the assessment of the evidence coders used a scale ranging from 1 ("the article only contains evidence without specific references") to 5 ("the article only contains evidence with specific references"). The absence of evidence was assigned a value of 0. Furthermore, we assessed the type of evidence (Murad et al. 2016). In this category, we differentiated between five types of evidence mentioned on the website: general references to (statistical) information, general reference to research, expert statements, references to specific individual studies, and reference to specific meta-analysis or systematic literature reviews on the state of research.

Balance: This variable refers to the extent to which a website takes a position on a particular topic, presents different opinions, or weighs them up against each other. To measure balance of positions, we assessed whether the information on the website favored one side or another. Although the tasks themselves did not contain any ideological framing, different characteristics of balance could still be displayed on websites. For the "cell phone task", this could be, for example, the presentation of radiation as either dangerous or harmless. Coding was done on a scale ranging from 1 ("completely one-sided positioning") to 4 ("completely balanced positioning"). If no positioning was observable, the website was coded 0. However, it is important to note that balance is not necessarily positive. For example, regarding the "cell phone" task the radiation from cell phone towers is harmless. A balanced presentation of information was therefore misleading.

6 Results

RQ1 concerns the types of online information sources that university students used to solve generic COR-tasks. The websites show a wide variety of sources, ranging from online news media (30%) followed by websites of public authorities (18%) and commercial websites (15%). User comments (2%) and entries in scientific databases (2%) made up the lowest share of unique websites. Few sources were found that could not be categorized into the classification system categories (7%).

RQ2 deals with the information quality of the websites used, i.e., completeness, accuracy, and balance. The completeness of relevant information was relatively low among all websites, although they have been identified as relevant by the students. More than one third (36%) of the websites did not display any information relevant to the tasks, with just 2% showing all of them. However, completeness also varied

between the different tasks. 52% of the websites used for the "cell phone" task did not contain any relevant information. This, for example, included websites that were somewhat thematically close to the task subject but did not cover any important information and even websites that were not at all related to the task. On the other hand, only 17% of the websites used for another task ("heat stick") contained no relevant information.

In contrast, the factual correctness of the relevant information was high. As 36% of the websites contained information that was not relevant to the tasks at hand, 55% of the websites displayed only accurate information, and 4% showed information that was mostly correct. False information was uncommon, with only 3% of websites containing mostly incorrect or completely incorrect information, concerning mostly websites of news media. However, students also found scientific studies that presented false information, for example a journal article stating that cell phone towers are a confirmed health hazard.

Regarding the occurrence of evidence, we found that more than a third of the analyzed websites (35%) did not refer to any evidence to back up the information. Of all websites that used evidence, 59% did so either exclusively or mostly without any specifications to fact-check them. This held especially true for many news media articles that frequently mention scientific findings, but do not provide a link to the original research. When examining the type of evidence, we observed that most websites (25%) only used non-scientific sources, followed closely by 20% of websites that used specific scientific studies as evidence. The use of expert statements (9%) and a general use of scientific studies, such as "studies show, that (...)" (9%), were equally common as evidence. Only 2% of websites used a meta-analysis or a systematic overview of scientific results as evidence.

Regarding balance, only a small number of websites (19%) exhibited either an exclusively balanced or partially balanced positioning towards the presented information, for example, both concerns about cell phone tower radiation and refutations of risks. More than half of the websites (56%) presented information in a clearly or partly one-sided way. On the other hand, 25% did not showcase any kind of positioning. This included websites which just contained lists of information, without further elaborations.

Finally, RQ3 concerns the differences in accuracy, completeness and balance in different types of information sources. For the categories, completeness of relevant information, factual correctness, and occurrence of evidence we conducted ANOVAs to analyze source differences. We also used the assignment of the websites to the respective tasks as covariates to adjust for potential differences between tasks.

Regarding completeness, we found that news media exhibited the highest amount of complete information ($M = 32.77$; $SD = 30.79$) and online encyclopedias the least ($M = 18.35$; $SD = 26.86$). Although the amount of complete information differed statistically significant for the type of source ($F\ (6, 128) = 3.53$, $p < 0.001$), subsequent post-hoc tests did not show any statistically significant differences between single website types. The difference between news websites and commercial websites only approached significance ($p = 0.06$) (see Table 1). For factual correctness, we found that websites of public authorities ($M = 4.89$; $SD = 0.40$) used slightly more correct information than the other types of sources. However, factual correctness was overall

high and did not vary significantly as a function of type of source (F (6, 225) = 0.93; p = 0.67) (see Table 1).

The use of evidence significantly varied between the different types of information sources ($\chi^2(6) = 20.31$, $p = 0.002$). For example, scientific sources (80%) and public authority (74%) websites used evidence more frequently than commercial websites (52%). Furthermore, we found differences in the use of specific evidence (citations, links, etc.) between website genres. It was used the most by scientific sources ($M = 4.13$; $SD = 1.59$), online encyclopedias ($M = 3.47$; $SD = 1.81$), and public authority websites ($M = 3.22$; $SD = 1.71$). News media ($M = 1.99$; $SD = 1.54$) and commercial websites ($M = 1.93$; $SD = 1.64$) used specific evidence the least. We also found statistically significant differences in this case (F (6, 216) = 6.54; $p < 0.001$), as scientific articles included more specific evidence than news media ($p < 0.001$) and commercial websites ($p < 0.001$). Public authority websites also showed a higher use of specific evidence compared to news ($p = 0.002$) and commercial websites ($p = 0.031$). News media also used fewer specific evidence than online encyclopedias ($p = 0.04$) (see Table 2).

For type of evidence, we found that, among all sources that used evidence overall, scientific articles used specific studies, including meta-analyses, the most (75%). On the other hand, only 21% of news media websites used specific studies as evidence, instead relying mostly on expert statements (33%). Overall, Chi-Square test showed a significant statistical correlation of website genre and type of evidence ($\chi^2(24) = 58.67$, $p < 0.001$) (see Table 2).

Regarding balance of positions, we found that news media exhibited the highest percentage of websites with a balanced positioning (35%). On the other hand, only 10% of scientific articles positioned themselves in a balanced way. Online encyclopedias were the type of source for which the positioning was rarely observable (61%). A Chi-Square test ($\chi^2(12) = 48.13$, $p < 0.001$) indicates that these differences are statistically significant across different types of sources (see Table 3).

Table 1 Completeness of the relevant information and factual correctness by type of source

Type of source	Completeness of the relevant information[a]		Factual correctness[b]	
	M	SD	M	SD
News media	32.77	30.79	4.51	1.07
Websites of public authorities	29.06	31.76	4.89	0.40
Commercial websites	22.37	21.74	4.64	0.99
Websites of interest groups	30.51	31.62	4.69	0.93
Online encyclopedias	18.35	26.86	4.69	0.75
Scientific articles	20.10	26.46	4.43	1.45
Other sources	18.03	21.96	4.38	1.20

[a] Scale from 0 to 100, with 100 representing full completeness, and 0 representing no relevant information for the respective task ($N = 343$)

[b] Scale from 1 to 5, with 5 representing exclusively correct information and 1 representing exclusively incorrect information ($N = 232$)

Table 2 Type of evidence by type of source

Type of source	News media	Websites of public authorities	Commercial websites	Websites of interest groups	Online encyclopedias	Scientific articles	Other sources
Type of evidence	%	%	%	%	%	%	%
General reference to information	34	39	48	44	47	25	37
General reference to research	13	22	15	13	7	0	21
Expert statements	32	0	4	17	13	0	5
Specific individual studies	21	30	33	22	33	69	37
Specific meta-analysis or literature reviews	0	9	0	4	0	6	0
Sum	100	100	100	100	100	100	100

$N = 223$

Table 3 Balance of positioning by type of source

Type of source	News media	Websites of public authorities	Commercial websites	Websites of interest groups	Online encyclopedias	Scientific articles	Other sources
Balance of positioning	%	%	%	%	%	%	%
Balanced	35	13	12	13	12	10	18
Unbalanced	52	56	67	64	27	70	50
No positioning	13	31	21	23	61	20	32
Sum	100	100	100	100	100	100	100

$N = 343$

7 Discussion

Overall, the results confirmed certain aspects of previous research on the information quality of websites, while also refuting some existing assumptions. The classification of scientific publications as high-quality sources (Nagel et al. 2020) was validated, specifically in how websites utilize evidence. These sources tended to provide more evidence supporting task-related information compared to other websites. Scientific articles were also found to use the highest quality evidence among all website genres analyzed. However, there are also few scientific articles that can provide inaccurate information and thus can mislead university students.

In general, most websites displayed signs of low information quality by either lacking evidence entirely or failing to disclose specific and verifiable sources of evidence. Our finding that there are no major differences in accuracy between website genres supports previous research (McNally et al. 2012).

The results also confirmed previous research related to university students' deficient information quality evaluation skills. They utilized a significant number of websites that did not contain relevant information to solve the assigned tasks. This finding supports the assumption that students lack skills in evaluating online content (Walraven et al. 2009). The limited selection of scientific sources by the first-year university students further confirmed their unfamiliarity with such sources (Weber et al. 2018). The task structure may also have influenced source selection. As the tasks only considered domain-unrelated issues of everyday life, scientific studies were not necessarily needed to solve the tasks, thus not prompting students to search for these sources.

However, the results of our study also differed from previous research. We found that the use of online encyclopedias like Wikipedia was less frequent than prior research suggests (Nagel et al. 2020). One account to explain this difference is that the respective task played a role. Given that participants had to solve generic information problems, online encyclopedias may not have been the first search results they encountered, and general news outlets may have been more up to date for the problem at hand.

7.1 Implications

Regarding previous research, some of our findings require further discussion. While scientific sources appeared to show high information quality, it is problematic to automatically assume all scientific sources to be superior in information quality. This simplification fails to account for the variation in information quality within scientific sources. In one instance, participants relied on a scientific article that presented evidence from studies that public authorities considered low-quality, lacking statistical control variables or exhibiting statistical confounding. Our results for other categories, along with the high standard deviations, indicate that differences between genres can obscure differences within genres. This underscores the importance of an in-depth analysis of information quality of the specific source content and cautions against making blanket statements about the quality of a website genre across categories.

Even though the completeness of single websites was low, it does not necessarily mean that the websites were inappropriate for the task at hand. Participants were required to indicate at least two relevant websites for their completion of the task. Therefore, it is possible that while each of the two websites may have had incomplete information, they may have complemented each other. Analyzing the two sources cited by the students as relevant in a complementary manner could offer more insight into the process of source selection and information navigation.

It must also be taken into account that, even though most websites presented information in a one-sided way, this is not necessarily an indication of low information quality. If there is unambiguously correct or false information, websites may be classified as one-sided if they only include accurate claims. This presentation of information on websites is desirable, as seen in our results regarding the balance of positioning of scientific studies. Websites that present both correct and false information may be more balanced, but this fact does not necessarily indicate higher information quality when false information is not qualified as inaccurate. This can even lead to a false equivalence of the arguments if incorrect information is presented to the same extent as correct information and is not categorized accordingly (Grimes 2019). Therefore, researchers need to account for whether a given task commands a response that can be qualified as true or false or whether a task requires students to present a balanced response deliberating on pro and con arguments. These task requirements should be taken into consideration when assessing information quality.

Overall, our results suggest that methods of classifying websites in previous research may be insufficient for assessing information quality in detail. For some dimensions of information quality, the classification of a few different types of sources as "good" or "bad" outlets may seem adequate, e.g. scientific articles for the occurrence of evidence. However, huge within-category differences cast doubt on the practice of coarse information quality ratings by which individual outlets are lumped together in an overarching category. This makes it difficult to generalize statements about the information quality of different types of online sources. Ultimately, the variation between single sources within a website genre can be greater than the differences between website genres, emphasizing the necessity for a more nuanced approach to evaluating information quality.

7.2 Limitations

Although the present study provided new insights, it is not without limitations. First, to assess the information quality of websites, they must always be considered in the context of the respective task. In this study, we only focused on generic information problems and neglected domain-specific tasks. This approach may be feasible for freshmen in the first semester but results in an incomplete picture for sophomores or more advanced students. Future studies should therefore also study information quality in domain-specific tasks.

Second, we only analyzed the information quality of websites in the context of four generic tasks with relatively few cases. It is therefore unclear to what extent our results can be generalized to other contexts that differ in the nature of the information problem or the task structure.

Third, as we only analyzed websites that the students indicated as relevant, we did not evaluate the information quality of discarded websites. Future studies should consider this and try to examine search behavior inside the information landscape as well. As students used comparatively few unique websites, it is of interest to see what factors influenced the selection of these sources. Future studies could rely on log data to examine these behaviors. For example, follow-up research can examine whether the selection of sources is affected by the placement of the search results. This could also provide a more coherent picture of the overall information landscape by examining which websites were generally visited. Additionally, the judgment that websites were relevant may have been influenced by the position participants wanted to take in their answer. Further research could aim to evaluate students' answers and the argumentative reasoning behind their source selection and combine these aspects with an analysis of information quality of the websites.

Finally, information quality is a complex construct consisting of different dimensions. The present study was unable to consider all of them. For example, in addition to the variables we assessed it is also relevant whether the websites communicated uncertainties regarding their information. The sources may also differ with respect to how factually they report on a topic, or whether they relied on statistics or case studies to back the evidence that is presented. These additional categories could be considered in more detail. Despite these limitations this study provides a solid foundation for further research into the role of information quality of websites when university students learn from online sources.

Funding The Research Unit FOR 5405 is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project number 462702138.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest T. Scherer, A. Laufer, Dr., M. Maurer and C. Schemer declare that they have no competing interests.

Ethical standards For this article no studies with human participants or animals were performed by any of the authors. All studies mentioned were in accordance with the ethical standards indicated in each case.

References

Allen, J., Watts, D. J., & Rand, D. G. (2024). Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science*, *384*(6699), eadk3451.

Bartlett, J. C., Frissen, I., & Beheshti, J. (2020). Comparing academic and everyday-life information seeking behavior among millennial students. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (pp. 402–406).

Basavakumar, D., Flegg, M., Eccles, J., & Ghezzi, P. (2019). Accuracy, completeness and accessibility of online information on fibromyalgia. *Rheumatology International*, *39*, 735–742.

Bohannon, J. (2013). Who's afraid of peer review? *Science*, *342*, 60–65.

Braasch, J. L. G., Bråten, I., & McCrudden, M. T. (Eds.). (2018). *Handbook of multiple source use*. Routledge.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, *53*(4), 1207–1217.

Brand-Gruwel, S., Kammerer, Y., Van Meeuwen, L., & Van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *Journal of Computer Assisted Learning*, *33*(3), 234–251.

Brox, H. (2012). The elephant in the room: a place for Wikipedia in higher education? *Nordlit*, *30*, 143–155.

Brünken, R., Münzer, S., & Spinath, B. (2019). *Pädagogische Psychologie – Lernen und Lehren*. Hogrefe.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). Guilford.

Denniss, E., Lindberg, R., & McNaughton, S. A. (2023). Quality and accuracy of online nutrition-related information: a systematic review of content analysis studies. *Public health nutrition*, *26*(7), 1345–1357.

Döring, N., Lehmann, S., & Schumann-Doermer, C. (2022). Verhütung in der deutschsprachigen Wikipedia: Eine Inhalts- und Qualitätsanalyse. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *65*(6), 706–717.

Döring, N., Lehmann, S., & Schumann-Doermer, C. (2023). Verhütung auf YouTube, Instagram und TikTok. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *66*(9), 990–999.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism and Mass Communication Quarterly*, *77*(3), 515–540.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, *9*(2), 319–342.

Garz, M. (2014). Good news and bad news. Evidence of media bias in unemployment reports. *Public Choice*, *161*, 499–515.

Goldman, S. R., & Brand-Gruwel, S. (2018). Learning from multiple sources in a digital society. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 86–95). Routledge.

Grimes, D. R. (2019). A dangerous balancing act: on matters of science, a well-meaning desire to present all views equally can be an trojan horse for damaging falsehoods. *EMBO reports*, *20*(8), e48706.

Hefferon, K. L., & Miller, H. I. (2020). Flawed scientific studies block progress and sow confusion. *GM Crops & Food*, *11*(3), 125–129.

Hoorn, J. F., & van Wijngaarden, T. D. (2010). Web intelligence for the assessment of information quality: credibility, correctness, and readability. In *Web intelligence and intelligent agents* (pp. 305–324). In-Tech.

Hughes, H. (2013). International students using online information resources to learn: complex experience and learning needs. *Journal of Further and Higher Education*, *37*(1), 126–146.

Johnson, T. J., & Kaye, B. K. (1998). Cruising is believing?: comparing Internet and traditional sources on media credibility measures. *Journalism and Mass Communication Quarterly*, *75*(2), 325–340.

Kelly, D. (2019). Evaluating the news:(Mis) perceptions of objectivity and credibility. *Political Behavior*, *41*(2), 445–471.

Krippendorff, K. (2018). *Content analysis: an introduction to its methodology*. SAGE.

Madathil, K. C., Rivera-Rodriguez, A. J., Greenstein, J. S., & Gramopadhye, A. K. (2015). Healthcare information on YouTube: a systematic review. *Health Informatics Journal*, *21*(3), 173–194.

Madge, C., Meek, J., Wellens, J., & Hoole, T. (2009). Facebook, social integration and informal learning at university. *Learning, Media and Technology*, *34*, 141–155.

Maurer, M., & Reinemann, C. (2006). Learning versus knowing: effects of misinformation in televised debates. *Communication Research*, *33*(6), 489–506.

Maurer, M., Jost, P., Haßler, J., & Kruschinski, S. (2019a). Auf den Spuren der Lügenpresse. Zur Richtigkeit und Ausgewogenheit der Medienberichterstattung in der "Flüchtlingskrise ". *Publizistik*, *64*(1), 15–35.

Maurer, M., Schemer, C., Zlatkin-Troitschanskaia, O., & Jitomirski, J. (2019b). Positive and negative media effects on university students' learning: preliminary findings and a research program. In O. Zlatkin-Troitschanskaia (Ed.), *Frontiers and Advances in Positive Learning in the age of Information (PLATO)* (pp. 109–119). Springer.

Maurer, M., Kruschinski, S., & Jost, P. (2024). *Fehlt da was? Perspektivenvielfalt in den öffentlich-rechtlichen Nachrichtenformaten*. Stiftung Mercator.

McNally, S.L., Donohue, M.C., Newton, K.P., Ogletree, S.P., Conner, K.K., Ingegneri, S.E., & Kagnoff, M.F. (2012). Can consumers trust web-based information about celiac disease? Accuracy, comprehensiveness, transparency, and readability of information on the internet. *Interactive Journal of Medical Research*, *1*(1), e2010.

McQuail, D. (1992). *Media performance: mass communication and the public interest*. SAGE.

Molerov, D., Zlatkin-Troitschanskaia, O., Nagel, M.-T., Brückner, S., Schmidt, S., & Shavelson, R.J. (2020). Assessing university students' critical online reasoning ability: a conceptual and assessment framework with preliminary evidence. *Frontiers in Education*, *5*, 557843.

Murad, M.H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine*, *21*(4), 125–127.

Nagel, M.-T., Schäfer, S., Zlatkin-Troitschanskaia, O., Schemer, C., Maurer, M., Molerov, D., Schmidt, S., & Brückner, S. (2020). How do university students' web search behavior, website characteristics, and the interaction of both influence students' critical Online reasoning? *Frontiers in Education*, *5*, 565062.

Nagel, M.-T., Zlatkin-Troitschanskaia, O., & Fischer, J. (2022). Validation of newly developed tasks for the assessment of generic critical online reasoning (COR) of university students and graduates. *Frontiers in Education*, *7*, 914857.

Olteanu, A., Peshterliev, S., Liu, X., & Aberer, K. (2013). Web credibility: features exploration and credibility prediction. In *Advances in Information Retrieval*. 35th European Conference on IR Research, ECIR 2013, Moscow, March 24–27, 2013. (pp. 557–568). Springer. Proceedings 35.

Reinemann, C., Maurer, M., Kruschinski, S., & Jost, P. (2024). Exaggerated and one-sided? Assessing the relevance and viewpoint diversity of German mainstream and alternative media coverage of the COVID-19-pandemic 2020-2021. *Journalism Studies*, *25*(6), 681–702.

Renkl, A. (2020). Wissenserwerb. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (pp. 3–24). Springer.

Rieh, S.Y., & Belkin, N.J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. *ASIS 98: Information access in the global information economy*. Proceedings of the 61st Annual Meeting of the American Society for Information Science. (pp. 279–289). Information Today.

Saleh, D., Fisher, J.H., Provencher, S., Liang, Z., Ryerson, C.J., & Weatherald, J. (2022). A systematic evaluation of the quality, accuracy, and reliability of internet websites about pulmonary arterial hypertension. *Annals of the American Thoracic Society*, *19*(8), 1404–1413.

Schatz, H., & Schulz, W. (1992). Qualität von Fernsehprogrammen. Kriterien und Methoden zur Beurteilung von Programmqualität im dualen Fernsehsystem. *Media Perspektiven*, *11*, 690–712.

Schemer, C., Ziegele, M., Schultz, T., Quiring, O., Jackob, N., & Jakobs, I. (2022). Political information use and its relationship to beliefs in conspiracy theories among the German public. *Journalism & Mass Communication Quarterly*, *99*(4), 908–929.

Springer, B., Bechler, U., Koller, U., Windhager, R., & Waldstein, W. (2020). Online videos provide poor information quality, reliability, and accuracy regarding rehabilitation and return to sport after anterior cruciate ligament reconstruction. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, *36*(12), 3037–3047.

Stark, B., Riedl, A., Eisenegger, M., Schneider, J., Udris, L., & Jandura, O. (2021). Qualität des politischen Nachrichtenangebots in Deutschland. Empirische Kernbefunde aus dem Projekt "Media Performance and Democracy". *Media Perspektiven*, *2021*(9), 430–449.

Tankard, J.W., & Royal, C.L. (2005). What's on the web—and what's not. *Social Science Computer Review*, *23*, 360–370.

Urban, J., & Schweiger, W. (2014). News quality from the recipients' perspective: Investigating recipients' ability to judge the normative quality of news. *Journalism Studies*, *15*(6), 821–840.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H.P. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, *52*(1), 234–246.

Weber, H., Becker, D., & Hillmert, S. (2018). Information-seeking behaviour and academic success in higher education. *Higher Education*, *77*, 57–678.

Wierzbicki, A. (2018). *Web content credibility*. Springer.

Zlatkin-Troitschanskaia, O., Hartig, J., Goldhammer, F., & Krstev, J. (2021). Students' online information use and learning progress in higher education—A critical literature review. *Studies in Higher Education*, *46*(10), 1996–2021.

Zlatkin-Troitschanskaia, O., Fischer, J., & Molerov, D. (Eds.). (2023). *Critical Online Reasoning in Higher Education (CORE)—Overview of the Research Unit* ('CORE' Working Papers, 1). Johannes Gutenberg University.