How Do University Students' Web Search Behavior, Website Characteristics, and the Interaction of Both Influence Students' Critical Online Reasoning?

Marie-Theres Nagel[1]*, Svenja Schäfer[2], Olga Zlatkin-Troitschanskaia[1], Christian Schemer[3], Marcus Maurer[3], Dimitri Molerov[1], Susanne Schmidt[1] and Sebastian Brückner[1]

[1] Department of Business and Economics Education, Johannes Gutenberg University Mainz, Mainz, Germany, [2] Department of Communication, University of Vienna, Vienna, Austria, [3] Department of Communication, Johannes Gutenberg University Mainz, Mainz, Germany

The Internet has become one of the main sources of information for university students' learning. Since anyone can disseminate content online, however, the Internet is full of irrelevant, biased, or even false information. Thus, students' ability to use online information in a critical-reflective manner is of crucial importance. In our study, we used a framework for the assessment of students' *critical online reasoning (COR)* to measure university students' ability to critically use information from online sources and to reason on contentious issues based on online information. In addition to analyzing students' COR by evaluating their open-ended short answers, we also investigated the students' web search behavior and the quality of the websites they visited and used during this assessment. We analyzed both the number and type of websites as well as the quality of the information these websites provide. Finally, we investigated to what extent students' web search behavior as well as the quality of the used website contents are related to higher task performance. To investigate this question, we used five computer-based performance tasks and asked 160 students from two German universities to perform a time-restricted open web search to respond to the open-ended questions presented in the tasks. The written responses were evaluated by two independent human raters. To analyze the students' browsing history, we developed a coding manual and conducted a quantitative content analysis for a subsample of 50 students. The number of visited webpages per participant per task ranged from 1 to 9. Concerning the type of website, the participants relied especially on established news sites and Wikipedia. For instance, we found that the number of visited websites and the critical discussion of sources provided on the websites correlated positively with students' scores. The identified relationships between students' web search behavior, their performance in the CORA tasks, and the qualitative website characteristics are presented and critically discussed in terms of limitations of this study and implications for further research.

Keywords: critical online reasoning assessment, online information, web search, log file analysis, content analysis, quality of online information, higher education

INTRODUCTION

In the context of digitalization, society's overall media behavior has changed fundamentally. Digital technologies are opening up new opportunities for accessing and distributing information (Mason et al., 2010; Kruse, 2017; Tribukait et al., 2017). The Internet has become one of the main sources of information for university students' learning (Brooks, 2016; Newman and Beetham, 2017). Prior research indicates that the way students process and generally handle online information can be strongly influenced not only by personal characteristics but also by the quality of the accessed websites and their content (Tribukait et al., 2017; Braasch et al., 2018). Possible relationships between qualitative website characteristics, students' web search behavior and their judging of online information, however, have hardly been studied to date. In particular, there are hardly any studies that examine the connection between different quality criteria of websites and students' evaluation of website quality. In addition, most of the existing studies are based on students' self-reports and/or were conducted in a simulated test environment, so that their generalizability regarding students' actual web search behavior in the real online environment (Internet) remains questionable.

To bridge this gap, the study presented here aims to provide empirical insights into the complex relationship between (1) students' search behavior, (2) students' evaluation of websites, and (3) the qualitative characteristics of the websites students evaluated and used in their written responses in a free and unrestricted web search, and (4) the real online environment (Internet) where they find their sources. Therefore, the study focuses on the research question: *To what extent are students' web search behavior – regarding the number and type of accessed websites and webpages – as well as the quality of the used website contents related to students' critically reflective use of online information*?

While the multitude of online information and sources may positively affect learning processes, for instance by providing access to a wide variety of learning resources at low effort and cost (Beaudoin, 2002; Helms-Park et al., 2007; Yadav et al., 2017), online information might also have multiple negative impacts on learning (Maurer et al., 2018, 2020). First, information available on the Internet is not sufficiently structured (Kruse, 2017), so that students may, for example, feel overwhelmed by the amount of information ("information overload," Eppler and Mengis, 2004). Second, since anyone can publish content online, the Internet is full of irrelevant, biased, or even false information. As a consequence, mass media rarely offer complete information and sometimes even provide inaccurate information as they are designed to exploit mental weak points that may present judgmental traps or promote weak reasoning (Ciampaglia, 2018; Carbonell et al., 2018). This holds particularly true for social media (European Commission, 2018; Maurer et al., 2018), whereby social networks and messengers, together with online newspapers and news magazines, online videos and podcasts, are considered the least trustworthy sources of news or information. As a result, when learning with and from the Internet, students face the heightened challenge of judging the quality of the information they find online. The extent to which the advantages of the Internet prevail or the disadvantages result in students being overwhelmed or manipulated depends on their abilities concerning search behavior and critical evaluation of online information.

Especially considering students' intensive use of the Internet during their higher education studies, it is important to assess and foster their critical online reasoning skills with suitable measures. One promising approach consists of performance assessments in an open-ended format using tasks drawn from real-world judgment situations that students and graduates face in academic and professional domains as well as in their private lives (McGrew et al., 2017; Shavelson et al., 2019). We therefore used a corresponding framework for assessing students' ability to deal critically with online information, the Critical Online Reasoning Assessment (CORA), which was adapted and further developed from the American Civic Online Reasoning Assessment (Wineburg et al., 2016; Molerov et al., 2019). With this framework, we assessed university students' ability to critically evaluate information from online sources and to use this information to reason on contentious issues (Wineburg et al., 2016). To investigate the research question of this study, we used students' response data, including (i) their web browser history with log files from the CORA task processing, (ii) their written responses to the tasks, and (iii) the websites they visited and the content they used during this assessment, and performed a quantitative analysis of website characteristics.

THEORETICAL FOUNDATIONS

Students as Digital Natives?

The evaluation of information sources is crucial for successfully handling online information and learning from Internet-based inquiry (Wiley et al., 2009; Mason et al., 2010), and using online information in a critical-reflective manner is a necessary skill. Critically analyzing and evaluating digitally represented information is necessary to cope with the oversupply of unstructured information and to analyze make judgments about the information found online (Gilster, 1997; Hague and Payton, 2010; Ferrari, 2013; Kruse, 2017). In higher education, it has long been assumed that students, as the generation of digital natives, are skilled in computer use and information retrieval and thus use digital media competently (Prensky, 2001; Murray and Pérez, 2014; Blossfeld et al., 2018; Kopp et al., 2019). However, recent studies have shown that students perform poorly when it comes to correctly judging the reliability of web content (e.g., McGrew et al., 2018). Although being familiar with a variety of digital media (e.g., social networking sites, video websites; Nagler and Ebner, 2009; Jones and Healing, 2010; Thompson, 2013), students use them primarily for private entertainment or social exchange, and are not capable of applying their digital skills in higher education and critically transferring information-related skills to the learning context (Gikas and Grant, 2013; Persike and Friedrich, 2016; Blossfeld et al., 2018). Students often base their judgment of websites on irrelevant criteria such as the order of search results and authority of a search engine, the website design,

or previous experience with the websites and the information provided there, while they neglect the background of a website or the credibility of the author (McGrew et al., 2017). For instance, Wikipedia and Google were the most frequently used despite students rating them as rather unreliable and students' overall use of all web search tools was rather unsophisticated (Judd and Kennedy, 2011; Maurer et al., 2020).

The core aspect of a successful search strategy is the correct evaluation and, in particular, selection of reliable websites and content therein, as students form their opinions and make judgments on this basis. If students refer to websites with biased or misrepresented information, this inevitably leads to a lower-quality or even completely incorrect judgment. It is therefore particularly problematic that, according to first studies, students struggle to evaluate the trustworthiness of the information they encounter online and to distinguish reliable from unreliable websites (McGrew et al., 2019). The expectation that today's students generally have a digital affinity is therefore not tenable (Kennedy et al., 2008; Bullen et al., 2011). To be able to deal successfully with online information, it is urgently necessary that today's students first learn to critically question, examine and evaluate it (Mason et al., 2010; Blossfeld et al., 2018).

Critical Online Reasoning

The ability to successfully deal with online information and distinguish, for instance, reliable and trustworthy sources of information from biased and manipulative ones (Wineburg et al., 2016) regarded *Critical Online Reasoning (COR)*, which we define as a key facet of critical and analytic thinking while using online media (Zlatkin-Troitschanskaia et al., 2020). In contrast to other concepts related to critical thinking, COR is explicitly limited to use in the online information environment. Besides *critical thinking* (Facione, 1990), COR refers to some aspects of *digital literacy*, i.e., the ability to deal with digital information and the technology required for it in a self-determined and critical manner (Gilster, 1997; Hague and Payton, 2010; JISC, 2014), which can be placed in the broader field of media competence and communication (Gilster, 1997; Hague and Payton, 2010; McGrew et al., 2017). Since the search for information using suitable strategies is an important aspect of COR, there is a conceptual overlap with 'information problem solving' (Brand-Gruwel et al., 2009). It refers in particular to metacognition, which regulates the entire COR process, including the development of an appropriate (search) strategy to achieve the objectives, reflecting on the status of information procurement and the search process.

In this respect, COR refers in particular to three superordinate dimensions: *searching and source evaluation, critical reasoning* and *decision making* (Zlatkin-Troitschanskaia et al., 2020). *Searching and source evaluation* describes the evaluation of the information and sources found online and includes the ability to select, understand and evaluate relevant texts on a website and to judge whether a source is credible, using additional resources available online and by cross-checking with other search results. *Critical Reasoning* means to recognize and evaluate arguments and their components used in the sources found online with regard to evidentially, objectivity, validity, and consistency. *Decision-making* refers to the process of making a correct, evaluative judgment and reaching a conclusion based on reliable sources, which also includes explaining the decision in a well-structured and logically cohesive way (for a more comprehensive description of the COR construct, see Zlatkin-Troitschanskaia et al., 2020).

Students' Internet Search Behavior

When navigating the web, users can either directly access websites that might provide the information they are looking for. However, the most common way of finding a way through the overwhelming amount of available information is to use a search engine (Beisch et al., 2019). Students have to narrow down their search and to select an appropriate website out of thousands of search results presented by the search engine. In a next step, the available information on a selected website has to be assessed in view of the task or the general informational need. The process of accessing websites, either directly or through search engines, and browsing websites to examine the available information is repeated until the user is satisfied with the findings (Hölscher and Strube, 2000) and is able to construct a mental model that meets their need for information.

When it comes to evaluating this process, there are two levels need to be addressed: the websites themselves (search results) and the information (content) provided by these websites. At the website-level, both the depth and the quality of the search behavior have to be considered (Roscoe et al., 2016). Depth means that it is important that the search for information online is extensive, in terms of both the number of search inquiries and the number of visited websites. The extensive use of search engines and various sources is a typical behavior applied by experienced Internet users and has been shown to improve the solving of web-related tasks (Hölscher and Strube, 2000; White et al., 2009). In this context, Wineburg and McGrew (2016) emphasize the importance of lateral reading: They found that if professional fact checkers have to evaluate websites, they quickly open various other tabs to verify the information with other sources. Other groups like students, who focused more on single websites and their features without cross-checking the content on other websites, performed worse in the given tasks.

Moreover, it is not only important that a variety of different sources is considered, but also that the information comes from sources that can be trusted (Bråten et al., 2011). In today's digital information environment, where everyone can publish and spread information online (e.g., via Blogs, Wikis, or Social Network Sites), the Internet is also used to disseminate misinformation or other manipulative content (Zimmermann and Kohring, 2018). Thus, users need to be able to effectively evaluate the sources they use when searching for online information (Brand-Gruwel et al., 2017). Several studies confirmed that experts from various domains such as history, finance or health pay much more attention to the authors and sources of online information than novices (Stanford et al., 2002; Bråten et al., 2011), which ultimately contributes to a higher task score (Brand-Gruwel et al., 2017).

In sum, it can be assumed that web search behavior is decisive for successfully solving online information problems and tasks. More precisely, the number of search queries, the number of

visited websites as well as the type of the sources used may contribute to a higher task score. To gain a deeper insight into students' web search behavior, we investigated how it can be described regarding the number and type of visited websites.

As studies have shown, a person's information seeking behavior is influenced by various factors such as their information needs (Tombros et al., 2005). One important variable is task complexity, which can influence the search process in that, as complexity increases, searchers make more search queries and use more sources of information (Kim, 2008). In studies, the search behavior of the users was also affected by whether a task had a clear answer or was rather open-ended (Kim, 2008). Beyond the task, there are inter-individual preferences for search strategies among users, as various attempts to assign users to different search behavior profiles show (Heinström, 2002). Similarly, differences between users and between tasks can be assumed with regard to the preferred types of websites. Therefore, it can be assumed that:

H1: There are differences related to both student characteristics and CORA task characteristics in students' web search behavior in terms of the number and type of websites and webpages students used to solve these CORA tasks.

Website Characteristics

Web search behavior refers both to judging the different websites as well as to evaluating the information the websites provide (Roscoe et al., 2016). In this respect, content quality can be evaluated on different levels of website content. On a more formal level, the topic of interest should be exhaustively covered, in terms of the amount/scope of information provided as well as the variety sources referenced, for the learner to gain knowledge and a broad understanding of a topic (Gadiraju et al., 2018). For example, in the field of digital news media and political knowledge, findings show that exposure to established news sites, which provide full-length articles and usually disclose the sources they use, has positive outcomes for the gain of information (Dalrymple and Scheufele, 2007; Andersen et al., 2016) while the use of social network sites as a source for information has no (Dimitrova et al., 2011; Feezell and Ortiz, 2019) or even negative (Wolfsfeld et al., 2016) effects. Unlike established news sites, social network sites only provide news teasers and thus cover only selected aspects of a topic that are not necessarily verified (Wannemacher and Schulenberg, 2010; Guess et al., 2019), which might explain the different effects in terms of information gain.

Additional criteria that are brought up frequently when judging the quality of journalistic content usually focus on specific content features. From a normative point of view, news "should provide citizens with the basic information necessary to form and update opinions on all of the major issues of the day, including the performance of top public officials" (Zaller, 2003, p. 110). To fulfill this purpose, media content has to be objective. That means media content should be neutral, meaning without any kind of bias that is manipulative, for example by favoring political actors or taking a certain side on a controversial issue (Kelly, 2018). Closely related to neutrality is *balance*, i.e., media should cover a topic by mentioning different points of view

(Steiner et al., 2018). This is especially important for solving social conflicts in a democracy (McQuail, 1992) but also for increasing the knowledgeability of citizens (Scheufele et al., 2006). Another facet of objectivity concerns *factuality*, i.e., media content should be based on relevant and true facts that can be verified.

In sum, a neutral and balanced coverage based on facts is a prerequisite for an informed citizenry that possesses the necessary knowledge to form own opinions. Therefore, these are important indicators for the quality of news and online media (McQuail, 1992; Gladney et al., 2007; Urban and Schweiger, 2014). Considering hypothesis *H1*, that students' website preferences depend on the task, the quality of the used websites likely varies by task as well. Therefore we assume:

H2: The quality of the used websites varies substantially between the different CORA tasks.

Students' Internet Search Behavior, Website Characteristics, and COR

With regard to the number and type of websites visited, previous findings have shown that relying on several trustworthy websites and the effective use of search engines is a web search behavior that is typically applied by experts and improves scores in online information seeking tasks (Hölscher and Strube, 2000; White et al., 2009). Wineburg et al. (2016), for instance, also emphasize the importance of using a multitude of different websites while searching for information. In their study, professional fact checkers, who verified information from a website with a variety of other sources, performed much better than other groups like students, who focused more on single websites and their features without cross-checking the content on other websites. Regarding COR, it can be assumed that there is a comparable relationship with search behavior. Consequently, we assume:

H3: During students' web search, a larger number and variety of websites used by a student is positively correlated with a higher COR score, compared to using fewer websites.

Since COR includes not only the correct evaluation of websites using other sources but also the critical handling of website content and the integration of the information found into a final judgment, it also relates to the content that websites provide. As websites differ with regard to content quality, they provide different baselines for COR. Regarding political information, for instance, if media coverage was too short, not exhaustive enough and from unreliable sources like social network sites, there was no (Dalrymple and Scheufele, 2007; Andersen et al., 2016) or even a negative effect on information gain (Wolfsfeld et al., 2016). The same applies to normative quality criteria such as neutrality, balance and facticity, which were positively related to information gain (Scheufele et al., 2006). Therefore, it can be assumed that:

H4: There is a positive correlation between the quality of the media content students used to solve the CORA tasks and their COR score, i.e., higher quality corresponds to a better COR score.

MATERIALS AND METHODS

Assessment of Critical Online Reasoning

To measure critical online reasoning (COR) we used five newly developed computer-based performance tasks (hereafter referred to as CORA) which were adapted from the US-American Civic Online Reasoning assessment (Wineburg and McGrew, 2016; for details on the adaptation, development and validation of CORA, see Molerov et al., 2019). Each task requires the participants to judge (a) whether a given website or tweet is a reliable source of information on a certain topic or (b) whether a given claim is true or untrue by performing a time-restricted open web search to respond to the task questions. In CORA, which comprises tasks on four different topics (Task 1: Vegan protein sources, Task 2: Euthanasia, Task 3: Child development, Task 4: Electric mobility, Task 5: Government revenue; for an example, see **Supplementary Appendix 1**), the participants had to evaluate the strengths and weaknesses of given claims, evaluate the credibility and reliability of different sources using any resources available online, and explain their judgments.

CORA aims to measure students' *generic* COR. Thus, the five tasks were designed in a way that, although they addressed certain social or political issues, students do not need prior content knowledge to answer the CORA tasks. Rather, each task prompt asks students to use the Internet to solve these tasks and formulate their responses as written statements. In particular, the prompts for tasks 1 and 3–5 provided a link to an initial website, which the test takers were asked to evaluate. The written responses in an open-ended format (short statements) to each task were scored according to a newly developed and validated rating scheme by two to three independent (trained) raters (for details, see section "Procedure").

Procedure

To explore our hypotheses, we conducted a laboratory experiment. Prior to the survey, the students were informed that their web history would be recorded and that their participation in the following experiment was voluntary; all participants signed a declaration of consent to the use of their data for scientific purposes. Subsequently, the participants' socio-demographic data and media use behavior were measured with a standardized questionnaire (approx. 10 min). Afterward, participants were randomly assigned three out of the five CORA tasks to answer. For each task, the participants had a total time of 10 min to conduct the web search and write a short response (30 min in total).

Students were asked to use the preinstalled Firefox browser and when they closed the browser, their browser history was automatically saved using the "Browsing History View" feature for Windows. Every change to the URL, caused either by clicking on a link, entering a URL in the address line or searching with a search engine, was logged. By giving participants one-time guest access to the computers, we maintained their anonymity and ensured that their Internet search results were not affected by previous browser usage.

For the assessment of the students' performance, their written responses to the open-ended questions were scored by two independent human raters using a three-step rating scheme that was specifically developed based on the COR construct definition (McGrew et al., 2017). Using defined criteria, the raters judged whether the participants had noticed existing biases in the websites linked in the tasks and had made a well-founded judgment with regard to the question. This resulted in a score of 0, 0.5, 1, 1.5, or 2 points per answer (with 2 as the highest possible score). We then calculated the interrater reliability and averaged the scores of both raters for each participant and for each task, whereby a sufficient interrater reliability was determined, with Cohen's kappa = 0.80 ($p = 0.000$) for the overall COR score. To analyze the log file data, we conducted a quantitative content analysis as described in the next section.

To test the hypotheses, we first analyzed the data descriptively. The correlations expected in the hypotheses were tested subsequently by means of correlation analyses, chi^2-tests, one-way ANOVAs and *t*-tests. All analyses were conducted using Stata Version 15 (StataCorp, 2017) and SPSS 25 (IBM Corp, 2017).

Content Analysis of Log Files

To analyze the log files from the CORA tasks with a content analysis (Früh, 2017), we developed a corresponding coding manual (see **Supplementary Appendix 2**). The basic idea of this methodological approach is to aggregate textual or visual data into defined categories. Thereby, the coding process needs to be conducted in a systematic and replicable way (Riffe et al., 2019). The essential and characteristic instrument for this process is the coding manual, which contains detailed information about the categories that are part of the analysis. Moreover, the coding manual also provides basic information about the purpose of the study and the units of analysis that are used to code the text material in the coding process (Früh, 2017).

In our study, we introduced two units of analysis: The first was the browsing history with a log file for each participant and task. First, formal information about the *test taker ID* and the *number of the task* were coded. This coding was followed by the *number of URLs (websites)* and the number of *subpages of URLs* (webpages) that the participant visited to solve the CORA tasks. The second unit of analysis were the individual URLs the participants visited to solve the tasks. Thus, the raters followed the links provided in the log file to access the information required in the different categories of the coding manual. The coding process for this unit of analysis also started with the coding of the *test taker ID* and the *number of the task* to enable data matching. Then, after following the link and inspecting the website (and webpages), raters had to code the *type of source*. To determine the characteristics for these categories, we selected sources that varied in their degree of reliability. As reliable sources, we considered journalistic outlets of both public service broadcasters[1] as well as established private news organizations[2]. Studies confirm that these outlets provide information with a high qualitative standard (Wellbrock, 2011; Steiner et al., 2018). Moreover, other reliable sources are those

---

[1] tagesschau.de

[2] spon.de

provided by governmental institutions (e.g., Federal Agency for Civic Education). Furthermore, we also considered scientific publications as reliable sources, which are especially important for finding online information in the context of higher education (Strømsø et al., 2013).

Additionally, we also coded the responses according to whether students relied on social media for solving the tasks during the study, as they are frequently used as learning tools in higher education. Here, especially *Wikipedia* plays an important role as an information provider (Brox, 2012; Selwyn and Gorard, 2016). The non-profit online encyclopedia invites everyone to participate as a contributor by writing or editing new or existing entries. Thus, although it can be considered the largest contemporary reference resource that is freely available to everyone, these entries are not made by experts and are published without review (Knight and Pryke, 2012). Another important social media channel for higher education is Facebook (Tess, 2013), which also offers some opportunities for learning since students can connect with each other and share information (Barczyk and Duncan, 2013). However, information spread over Facebook is often unreliable or even false (Guess et al., 2019). Moreover, Facebook is a tool that is predominately used to keep in touch with friends and thus rather increases distraction when it comes to learning (Roblyer et al., 2010). Thus, both Wikipedia and Facebook but also blogs, Twitter, YouTube, Instagram, and forums were considered in the coding manual. Sources which have commercial interests like online shops of shops of organizations were considered even less reliable than social media are. Their purpose is not to provide neutral information but to convince users to buy their products and to increase financial profits.

In sum, the list of source types consisted of websites that were part of the CORA task, social media sites, research institutes, websites of governmental institutions, news sites, sites of specialist magazines, scientific publications, book uploads, online shops or sites of organizations. If none of these categories matched the source used by a student, "other" was coded. For each category, further subcategories were provided to differentiate, for example, whether social media means that the participant visited either Facebook or Wikipedia. **Figure 1** shows what types of websites students accessed according to this coding. **Figure 2** shows one example where a participant accessed the following websites to find information for one of the CORA tasks.

Each of these websites is assigned a numerical code depending on the type of source. The first entry is Wikipedia, which has the value 22. The next URL goes back to a university website, which is coded with 31. The same is true for the website of the PThU, which is also a university and thus received the value 31. The next two links are both established news websites and are coded with 52. Finally, youtube.com is a social media site and has the value 25 (for the coding manual, see **Supplementary Appendix 2**).

In a next step, we analyzed the quality of the content of the online sources used by students when solving a CORA task. To evaluate content quality, certain indicators were identified. In this part of the coding process, the raters followed the links of the log file and applied the coding scheme to the text on each webpage the participant visited. Since the *amount of information (text) on a webpage* and the *use of external sources* are particularly important

for learning and understanding concepts, these indicators were considered as quality dimensions in our study. With regard to the *amount of information*, the raters examined which sections of the individual webpages dealt with the topic covered in the task and counted the number of words in those sections. For coding the number of external sources (scientific and non-scientific), links in the texts as well as sources mentioned at the end of the texts were also counted. If sources were mentioned in the text, it was additionally coded if the text addressed the credibility of these sources (0 – no/1 – yes). Other quality indicators that are important for learning and understanding, and have thus been considered in various studies on media quality, are *balance* and *facticity* (McQuail, 1992; Gladney et al., 2007; Urban and Schweiger, 2014). They too were part of the coding manual. For balance, if the text has a clear stance that takes the side of, for example, a certain actor or on an issue, 1 was coded. If the text was rather balanced or did not take any side, 2 was coded. Facticity addressed the relation of opinion and facts. Three different codes indicated if the text contained (almost) exclusively opinions (1), was balanced in this regard (2) or was (exclusively) based on facts (3).

The coding process was conducted by two human raters. One of these raters was responsible for the first recording unit (log file), the other rater for the second (content of the websites). Before starting the coding process, the raters were intensively trained by the researchers of this study. To provide a sufficient level of reliability, the researches and the raters coded the same material and compared the results of the coding until the level of agreement between researchers and coders reached at least 80% for each category.
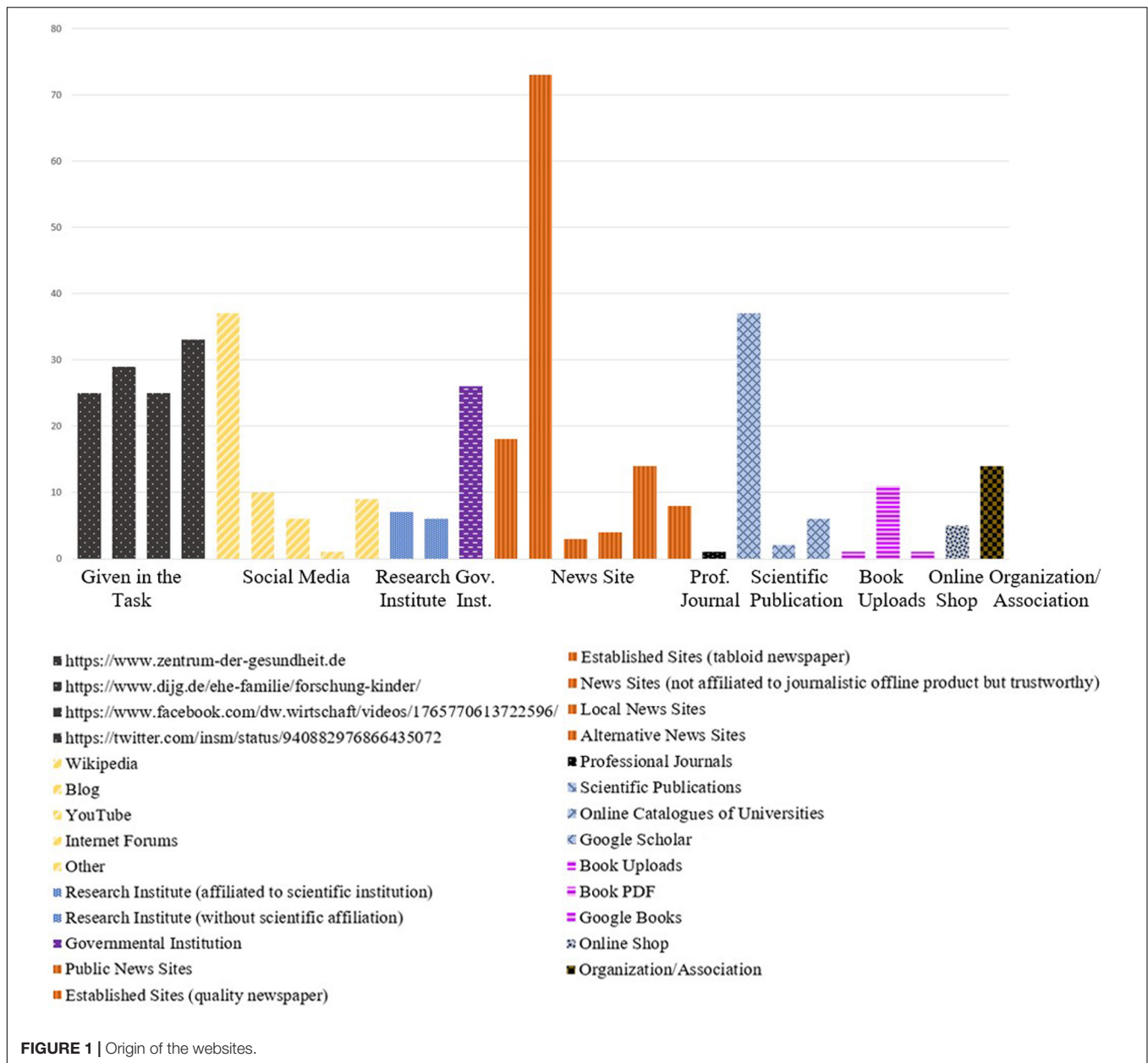
Sample

The sub-sample used in this study consisted of 45 economics students from one German university and is part of the overall sample (N = 123; see below) used in the overarching CORA study (see Molerov et al., 2019; Zlatkin-Troitschanskaia et al., 2020). Participation in the CORA study, which was voluntary, was requested in obligatory introductory lectures at the beginning of the winter semester 2018/2019, the summer semester 2019, and the winter semester 2019/2020. To ensure more intrinsic test motivation, for their participation in the study, the students received credits for a study module.

For this article, a selected sub-sample of the participant data was used, since the coding and analysis of all websites the CORA participants used to solve these tasks was hardly feasible for practical research limitations. When selecting this subsample, we included students from all study semesters represented in the overall sample. Another important criterion for the sampling was the students' central descriptive characteristics such as gender, age, migration background and prior education, which may influence students' web search behavior and COR task performance.

The majority of the 45 participants were at the beginning of their studies (m = 1.76, SD = 1.45, with an average age of 21.5 years (SD = 2.82) and an average school-leaving grade of 2.44 (SD = 0.62); 60% of the participants were women.

The subsample used in this study, which is relatively large with a view to the comprehensive analysis conducted in this

FIGURE 1 | Origin of the websites.

FIGURE 2 | Excerpt from a log file.

study, can be considered representative for both the total sample of the CORA study and for German economics students at the beginning of their studies in general: For instance, the overall CORA sample consists of 123 students with 61% women, an average age of 22 years (SD = 2.82), and an average study semester of 2.02 (SD = 1.82). There are also hardly any differences

TABLE 1 | Average number of visited websites and webpages per task.

Task	Sample	Websites (SD)	Webpages (SD)	Score (SD)
1	24	2.54 (SD = 1.77)	2.67 (SD = 2.35)	0.88 (SD = 0.82)
2	29	4.86 (SD = 1.87)	2.07 (SD = 1.79)	1.39 (SD = 0.62)
3	29	2.07 (SD = 1.13)	5 (SD = 4.18)	0.66 (SD = 0.64)
4	25	2.88 (SD = 1.67)	4.76 (SD = 4.91)	0.84 (SD = 0.53)
5	28	2.93 (SD = 1.80)	4.21 (SD = 4.28)	0.81 (SD = 0.40)

regarding the school-leaving grade, with an average of 2.41 (SD = 0.52). In comparison with a representative Germany-wide sample of 7111 students (Zlatkin-Troitschanskaia et al., 2019) in their first semester, there are also hardly any differences in terms of age (m = 20.41, SD = 2.69) and school-leaving grade (m = 2.37, SD = 0.57). Only the proportion of women in the German-wide representative sample is slightly lower (54%). However, since previous studies did not find any gender-specific effects on COR performance (Breakstone et al., 2019), this study assumes that there is no discrepancy of this kind that should be taken into consideration.

RESULTS

Students' Internet Search Behavior (H1)

To examine the online search behavior of students, both the number of websites and webpages as well as the type of websites students visited while solving the CORA tasks were examined. The findings are displayed in **Table 1**. On average, the number of visited websites was highest for task 2 and lowest for task 3. With regard to the visited webpages of a website, however, the pattern was exactly the opposite: Task 3, for example, had on average the most webpages visited per participant, whereas task 2, with m = 2.07 (SD = 1.79), had the fewest. This pattern shows that for task 2, participants rather relied on several websites, while for the other tasks they browsed the webpages of websites more intensively.

With regard to the type of source, a descriptive analysis showed substantial differences between the tasks. In tasks 1, 3, 4, and 5, for instance, one of the most frequently visited pages was the one given in the task prompt (n = 24 in task 1 – n = 33 in task 5), whereas no website was given as a starting point for task 2 (see **Table 2** and **Figure 1**).

All students complied with the task and accessed the pages from the task prompts at least once. By contrast, for task 2 which did not have a given start page, pages in the categories social media (n = 21), news (n = 65) and scientific publications (n = 34) were used more frequently than for the other tasks. In the other tasks, significantly fewer pages of these categories were visited: Social media pages were used four (task 3) to 17 times (task 5), news pages seven (task 5) to 31 times (task 4), and scientific publications or online catalogs were used not at all (task 4) to 5 five times (task 3).

Moreover, significantly more news pages were used for task 4 (n = 31) than for tasks 1, 3 and 5 (n = 7 – n = 9), and more governmental web pages for task 5 (n = 13) than for the other

TABLE 2 | Number of accessed websites by website type and per task[1].

Type of website	Task					Total
	1	**2**	**3**	**4**	**5**	
1 Given in the task	24	0	29	25	33	112
2 Social media	12	21	4	9	17	63
3 Research Institute	2	4	4	0	3	13
4 Gov. Institution	1	7	4	1	13	26
5 News site	9	65	8	31	7	120
6 Professional journal	0	1	0	0	0	1
7 Scientific publication	3	34	5	0	3	45
8 Book uploads	4	2	3	2	2	13
9 Online shop	3	2	0	0	0	5
10 Other	3	2	3	3	3	14
	61	139	60	71	81	**N = 412**

[1]Referring to the number of individual websites that all participants visited across all tasks.

tasks (n = 1 in tasks 1/4 – n = 7 in task 2). Overall, the participants most frequently used news pages (n = 120), pages from the tasks (n = 112), social media (n = 63), and scientific publications (n = 45). An examination using the chi^2-test confirmed that students' use of website categories significantly differed between the tasks (χ^2 = 180.81, df = 36, p = 0.00).

Since the differentiation observed so far was still rather rough, we conducted a more precise analysis based on the types of online sources most frequently used by the participants. In the case of news sites, the sites used by far most frequently (60.8%) were established news sites that can be assigned to a high-quality (national) newspaper or magazine[3,4] (see **Table 3**). Public broadcasting news sites[3] (e.g., 15%) and local news sites[4] were accessed occasionally (11.7%). In contrast, other sites that cannot be assigned to any journalistic offline product and/or that disseminate rather unreliable news were used little or not at all. When selecting news sites for research, students apparently mainly relied on well-known and established national and local news sites, while avoiding possibly less well-known sites without offline equivalents and well-known unreliable sites.

In the next most frequently used category, social media, the students' preference for Wikipedia was evident at 58.5% (see **Table 4**), followed by blogs (15.9%) and various sources that cannot be specifically assigned (14.3%). YouTube was also used a few times (9.5%), while Facebook, Twitter, Instagram, and forums were hardly or not at all used.

The third most frequently used type of website, scientific publications or online catalogs, was almost exclusively used in the form of pages of scientific journals (82.2%); online catalogs of universities (4.44%) or Google Scholar (13.33%) were hardly used at all. Overall, the analyses pertaining to H1 show clear differences in students' search behavior. Hypothesis H1 was therefore not rejected.

[3]faz.net
[4]zeit.de

TABLE 3 | Number of accessed news websites by subtype and per task[1].

	N = 120	
Type of Source	**Freq.**	**Percent**
Public News Sites	18	15.00
Established Sites of a Quality Newspaper	73	60.83
Established Sites of a Tabloid Newspaper	3	2.5
News Sites not Affiliated to any Journalistic Offline-Product (Trustworthy News)	4	3.33
Local News Sites	14	11.67
Alternative News Sites	8	6.67
	120	100.00

[1]Referring only to the sub-group of news websites that the participants visited.

TABLE 4 | Number of accessed social media websites by subtype and per task[1].

	N = 63	
Type of Source	**Freq.**	**Percent**
Wikipedia	37	58.73
Blog	10	15.87
YouTube	6	9.52
Online Forums	1	1.59
Other	9	14.29
	63	100.00

[1]Referring only to the sub-group of social media websites that the participants visited.

Quality of the Visited Websites (H2)

The second analyzed aspect of students' web search behavior concerns the quality of the webpage content that the respondents visited to solve the CORA tasks. In terms of length, participants visited websites that provided on average $m = 2,448.81$ (SD = 4,237.06) words that were relevant for the CORA tasks. An examination of the average number of words on the webpages used for each task showed that task 3 stands out. Here, the webpages provided an average number of 4,865.31 (SD = 6,631.01) words, followed by webpages visited to solve task 2 with $m = 2,797.67$ (SD = 5,230.25) words, task 1 with $m = 1,748.39$ (SD = 2,126.10) words, task 4 with $m = 1,559.41$ (SD = 1,565.11) words and task 5 with $m = 1,458.41$ (SD = 1,549.62) words. A one-way ANOVA of the extent of the webpages grouped by task revealed that there are significant differences between the tasks [$F(4) = 7.70$; $p < 0.001$]. A comparison between the groups with a *post hoc* test (Tamhane T2) showed that task 3 significantly differs from all other tasks except for task 2 ($p < 0.05$). All other group comparisons were not significant.

Another quality dimension concerned external sources (see **Table 5**). Here, we investigated whether scientific sources or non-scientific sources were mentioned on the accessed webpages and whether the response text discussed the quality of the used external sources. Concerning the first two categories, on average, the websites used to respond to the tasks provided $m = 6.92$

TABLE 5 | Average number of scientific and non-scientific sources per task[1].

Task	Average number of scientific sources (SD)	Average number of non-scientific sources (SD)
1	7.48 (SD = 21.14)	3.09 (SD = 5.55)
2	6.38 (SD = 12.64)	2.88 (SD = 6.52)
3	22.16 (SD = 42.80)	3.90 (SD = 11.32)
4	0.13 (SD = 0.34)	4.28 (SD = 12.86)
5	2.19 (SD = 5.34)	8.41 (SD = 13.51)

[1]Referring only to the sub-group of scientific/non-scientific websites that the participants visited.

(SD = 21.12) scientific sources and $m = 4.48$ (SD = 10.45) non-scientific sources.

The analysis of the average number of sources per task showed that for scientific sources, task 3 stands out. There, participants relied on websites that provided many scientific sources. For task 4, participants rather relied on sources with almost no scientific sources. According to a one-way ANOVA with the number of scientific sources as dependent and the respective tasks as grouping variable, the tasks had a significant effect for the number of scientific sources used [$F(4) = 11.57$; $p < 0.001$]. The *post hoc* tests (Tamhane T2) showed that except for task 1, all other tasks differed at least marginally from each other ($p < 0.10$). Concerning non-scientific sources, participants preferred websites with a higher number of this kind of sources for task 4 and 5. All other mean values were rather similar. Here, the effect of the task was also significant [$F(4) = 3.87$; $p < 0.01$]. However, according to the *post hoc* tests (Tamhane T2), only the differences between task 5 and task 2 as well as task 5 and task 1 were significant ($p < 0.05$).

Finally, we analyzed whether the visited webpages referenced additional sources that could allow for conclusions about the credibility of the websites to be drawn. On average, this was the case for 52 of the 379[5] analyzed webpages. If the tasks are also considered, for task 2, no webpage was used that addressed the reliability of the sources. For all other tasks, the share of webpages discussing the external sources was on a comparable level that varied between 21.2 and 28.8%. An examination using the chi^2-test confirmed that there are highly significant differences ($\chi^2 = 28.67$, df = 4, $p = 0.00$).

Further, we also investigated whether an article was balanced and based on facts. Concerning balance, 138 of the 379 (36.4%) webpages were rather unbalanced while the rest was classified as balanced. If the tasks were taken into account, for task 1, the share of webpages with rather unbalanced information was high with 70.7%, while for task 2 this share was rather low (10.6%). For task 3, about a third (32.8%) of the webpages was rather unbalanced while the share for task 4 was 46.5 and 41.8% for task 5. A chi^2-test confirmed that the differences were highly significant ($\chi^2 = 66.32$, df = 4, $p = 0.00$).

---

[5]This number differed from the number of 412 analyzed websites mentioned before because 33 of the links were expired when the content analysis was conducted. Even if, for example, the type of source could still be identified, the content of the websites could not be used for the content analysis.

TABLE 6 | Level of facticity of the visited websites per task[1].

Task	Mere opinion	Balanced	Mere facts	Total
1	33 (56.9%)	13 (22.4%)	12 (20.7%)	58
2	3 (2.7%)	58 (51.3%)	52 (46.0%)	113
3	14 (24.1%)	24 (41.4%)	20 (34.5%)	58
4	31 (43.7%)	32 (45.1%)	8 (11.3%)	71
5	31 (39.2%)	26 (32.9%)	22 (27.8%)	79
Total	112 (29.6%)	153 (40.4%)	114 (30.1%)	*N* = 379

[1]*Referring to the number of individual websites that all participants visited across all tasks.*

The last category was facticity (see **Table 6**). If the tasks were not considered, the participants chose roughly equal shares of websites that were more opinion-based, balanced, and fact-based. However, this depended again on the tasks, and every task had a different pattern. Participants chose almost no opinion-based websites for task 2 but rather preferred balanced or fact-based content. The highest values for opinion-based content were found for task 1, followed by tasks 4 and 5. In task 3, participants preferred especially balanced websites. A chi^2-test confirmed that the differences were highly significant (χ^2 = 79.70, df = 8, p = 0.00). Based on the analysis results, H2 was not rejected.

Correlation of Search Behavior With the COR Score (H3)

The descriptive examination of the average scores per task showed substantial differences between the tasks, with participants scoring best on average for task 2 with 1.39 points and worst for task 3 with 0.66 points (see **Table 2**). There were significant positive correlations between the number of websites visited and the task scores, for task 1 (n = 48, r = 0.59, p = 0.000), task 3 (n = 52, r = 0.33, p = 0.02) and task 5 (n = 53, r = 0.32, p = 0.02). Even if no significant effects were found for items 2 and 5, a correlation of the total number of websites used by the participants with their summed up overall scores confirmed the overarching tendency that the use of a larger number of websites in the processing of CORA tasks was associated with a higher CORA score (n = 87, r = 0.49, p = 0.000). Overall, the results indicated at least the tendency that visiting more websites during the search was associated with a better CORA test performance. Thus, H3 was not rejected.

Relationship Between the Quality Characteristics of Visited Websites and COR Score (H4)

When analyzing the relationship between the types of websites used by students and their CORA task score, a corresponding single-factor analysis of variance with the ten groups of websites as independent variable and the total score of the participants as dependent variable was just barely not significant (p = 0.06).

For a comprehensive analysis of the correlation between different aspects of search behavior, we considered whether the score was different between students who only visited the websites specified in the task and students who visited additional websites. To obtain the overall score, each participant's scores on all three completed tasks were added up, resulting in ranging between 0 and 6 points. A corresponding t-test for the overall score of both groups showed that the students who visited additional websites, on average, achieved a significantly higher total score of 3.20 points than the students who only stayed on websites linked in the CORA tasks (2.72 points, p = 0.002).

To analyze the effects of the quality characteristics of the website content on the score, the categories defined in the coding manual (length, use of scientific sources, use of non-scientific sources, discussion of the sources, balance and facticity) were examined using correlation analyses. Here, we only found a significant correlation for the discussion of external sources and the overall score r = 0.22, (p = 0.000). All in all, with regard to the characteristics of the visited websites, two characteristics in particular had a significant correlation with the COR results: While participants who only remained on the websites specified in the tasks performed worse, the use of websites that critically report sources had a positive effect. In sum, H4 had to be partly rejected.

DISCUSSION

Interpretation of the Results

For this study, students' critical online reasoning (COR) was assessed using open-ended performance tasks, and their web search behavior was analyzed using log files that recorded their actions while solving the tasks. Concerning the CORA task performance, the students show a low level of skill in judging the reliability of websites, which confirms previous findings (Wineburg et al., 2016; Breakstone et al., 2019). The students' web search behavior differs between tasks, and the type and wording of the task appears to have a noticeable correlation with the students' search behavior. The identified differences may be explained by the specific characteristics of the respective tasks, as some of the tasks, for instance, referred to everyday topics frequently addressed on the news (e.g., task 4). When solving tasks that included a link to a website, the participants tended to spend more time on these websites and look at a larger number of subpages, visiting fewer or no additional websites in the free web search. In contrast, the students visited significantly more websites while solving task 2, which had not included any links to websites.

Furthermore, the students were found to have preferences for certain types of websites, especially news sites, social media (Wikipedia), and scientific journals, both across all CORA tasks and with regard to the individual tasks, whereas other website types such as blogs and online shops were neglected. Taking into account the limited processing time, it can be assumed that after reading the task and visiting the corresponding website, the participants tried to gain an overview by visiting well-known news sites, scientific journals, and online encyclopedias. The deviations in task 2 also show that the content of the task prompt and, in particular, whether a link to a website was included therein appears to correlate with the resulting search behavior.

The findings confirm that students have a strong preference for Wikipedia as a source of information (Maurer et al., 2020). This indicates that even though the students could certainly pay more attention to scientific sources and should not rely on Wikipedia as much as they do, they at least refrain from using completely unreliable sources such as alternative news sites, online shops, or Facebook. Wikipedia has a special status in this regard. It has been repeatedly proven to be a reliable source of information, which studies have attributed to the collaboration between Wikipedia users. However, the fact that Wikipedia articles are often written as a collaborative effort between numerous users is also the reason why the credibility of the Wikipedia articles cannot be guaranteed (Lucassen et al., 2013).

Overall, students tend to rely on sources they also typically use to gain information in their everyday life (Beisch et al., 2019) and might know from a university context (Maurer et al., 2020). This is consistent with earlier findings that indicated that people prefer to search for information on websites they are familiar with through their own experience or that are generally well known, and that students in particular are more likely to use a limited range of media or sources, depending on the nature of the task and immediacy considerations (Oblinger and Oblinger, 2005; Walraven et al., 2009; List and Alexander, 2017). In our study, the students may have tried to avoid wasting time by visiting uninformative websites or unknown websites.

The fact that the students had a tendency to spend a large amount of time on the websites linked in the CORA tasks also confirms prior findings (e.g., Flanagin and Metzger, 2007; Kao et al., 2008; Hargittai et al., 2010; Wineburg et al., 2016), where the study participants (with the exception of professional fact checkers) showed a tendency to focus on individual websites and their features. This tendency may have been more pronounced in the context of the CORA due to the time limit of 10 min. This time limit may have also caused a tendency to neglect online catalogs of universities and Google Scholar, as the students did not have the time to read in-depth scientific articles. In this respect, our study is in line with prior findings that students also consult non-scientific sources when they need information in regular university life (e.g., when preparing for exams) and have more time (Maurer et al., 2020). With regard to scientific sources, the question arises how elaborate findings can be presented in a more comprehensible way and how it can be ensured that they are easier to understand for a wider audience by paying more attention to the needs of the readership.

Concerning the quality of the websites' content, the participants tended to rely on news based on both scientific and non-scientific sources. This especially holds true for task 3, where the participants used a larger number of scientific sources and wrote significantly more words than they did when solving the other tasks. A possible explanation for this finding might be that the topic of the task (Child development) has a more scientific background than the other tasks, which makes it necessary to rely on more comprehensive and more scientific websites. This is an important finding, as sources with these kinds of characteristics also provide a suitable basis for learning (Dalrymple and Scheufele, 2007; Dimitrova et al., 2011). For the other two quality dimensions, balance and facticity, we find a more mixed picture. In general, the students tended to also take into consideration unbalanced and opinion-based sources. However, this depends on the specific task. In this context, it has to be considered that some of the tasks contained links to websites that were categorized as unbalanced and opinion-based. Taking this into account, it can be concluded that students appear to know how to find websites with reliable and fact-based content, even though there is still some room for improvement.

Moreover, as assumed, we found a relationship between the number of visited websites (as an indicator of the students' web search behavior) and a higher CORA task performance, with the exception being task 2, where no significant correlation was found. This could be due to the fact that no links to websites were included in this task. In the other four CORA tasks that did include links to specific websites, these websites were usually either biased or only of limited reliability. Thus, spending a large amount of time on these websites alone may have a significantly negative effect on the final responses of the participants, whereas visiting other websites could cause them to detect the bias. In task 2, however, this situation did not apply, as no link was included in the task. Since the students with their heuristic approach of referring to well-known (news) sites and encyclopedias largely avoided particularly unreliable or biased websites when working on the CORA tasks, there may have been less of a correlation between the total number of websites visited and the quality of the students' task responses in task 2. Studies from the field of political education in particular indicate that established news sites that offer full-length articles can have a positive effect on knowledge acquisition (Dalrymple and Scheufele, 2007; Andersen et al., 2016).

Another finding is also indicative of a relationship between the task definition, search behavior, and score: Students who visited additional websites on average achieved a significantly higher COR score than those who only looked at the websites (and webpages) mentioned in the tasks. In addition, with regard to the quality dimension of facticity, the amount of "purely opinion-based argumentation" expressed in task 2 was particularly low and the amount of "purely facts-based argumentation" particularly high (see **Table 6**), while the average test score was significantly higher in task 2 than in the other tasks. These findings are consistent with prior research (e.g., Anmarkrud et al., 2014; Wineburg et al., 2016; List and Alexander, 2017), indicating that it is of great importance to at least check the reliability of a website and its contents and to cross-check the information stated on a website with that stated on other websites.

Contrary to our second hypothesis (H2), however, we hardly found any correlations between other website characteristics such as facticity or scientific/non-scientific sources and the students' test score. One reason for this finding could be that the participants used only a limited variety of different websites and had only a limited amount of time to perform their online searches, so that certain website characteristics such as the extent of task-related content included therein were not fully considered. In addition, it is not clearly evident from the log data how much time the participants spent on the individual webpages and which sections of these webpages they actually read, whereas the ratings based on the coding manual always refer to entire

webpages. Thus, higher correlations may have been determined for certain characteristics if only the sections of the webpages that the students actually read had been taken into account. The aspect of "number of scientific/non-scientific sources," however, may be of limited use as a quality feature of a website, as biased websites can also use external sources to convey the impression that they are credible sources of information.

In conclusion, the students showed a great heterogeneity both in terms of their Internet use and their performance on the CORA tasks. Taking into account task-specific characteristics (in particular the wording of the task), the most frequently used websites were the ones that had already been included in the tasks as well as news sites, social media sites, and scientific websites. In particular, preferences for established news sites, websites of scientific journals, and Wikipedia were found. The quality of the visited websites also varied and depended on the task that the test-takers were working on. On average, the websites provided a large amount of relevant information and used both scientific and non-scientific sources. However, only a rather small number of websites critically reflected on their sources, and the participants showed a preference for websites that were based on opinions. When it comes to the relationship between these content features and the scoring, we found a positive relationship between the score and the number of visited websites, and the use of additional websites beyond the ones already included in the tasks. Although no significant correlation between the type of website used and the students' CORA test performance could be determined, using websites that critically reflect on their sources also increased the students' test performance.

Limitations

There are also some limitations to our study that should be taken into account when interpreting the results. For instance, this paper used a subsample of the CORA study consisting only of students in the first phase of their economics studies at one German university. As Maurer et al. (2020) indicate, students' web search behavior may differ between different study domains and universities or change over the course of study. For example, Breakstone et al. (2019) found that students with more advanced education performed better on tasks on civic online reasoning. Therefore, both search behavior and CORA performance will be analyzed in a larger and more heterogeneous sample in follow-up studies to confirm the representativeness of the results found in this study.

The possible influences of personal characteristics on students' Internet search behavior and CORA performance were not considered in this paper. Previous studies found correlations between, for example, the influence of ethnicity and socio-economic background on civic online reasoning (Breakstone et al., 2019). The extent to which these correlations between the participants' personal characteristics and their CORA performance can be replicated and whether they also have an effect on search behavior is being clarified in further studies. With regard to the implications for university teaching, for example, it would be useful to learn more about the students' prior knowledge of strategies for searching information on the Internet

(acquired through, e.g., previously attended research courses at the university) and how they deal with misinformation.

To better understand the correlation between website characteristics and the students' performance on CORA tasks, their search processes should be analyzed in more detail, for instance, based on data from an eye-tracking study. In particular, the duration and frequency of the individual webpage visits should be included in further analyses and it should be examined in more detail which sections of the visited webpages the students focused on and what exactly they did there (e.g., reading certain sections). A useful empirical extension of our study might be to include a more qualitative approach to investigate the students' web search behavior. For example, using the think-aloud method (Leighton, 2017) would reveal in more detail which strategies the students apply when searching for information online, how they choose their sources, and how they judge the content of websites. For this purpose, it would be helpful to use experimental study designs that focus on examining specific aspects of the web search, for example, how trainings focused on using different web browsers und different search interfaces affect the search process. This might be considered in future research designs.

Implications

In addition to the implications for future studies resulting from the limitations described above, further implications can be derived from the findings of this study, especially for (university) teaching. In particular, the overall rather poor CORA performance of the students confirms that there is a clear need for support when it comes to dealing with online information in an appropriate way (e.g., Allen, 2008). This is of particular importance, as the Internet is the main source of information for students enrolled in higher education (Maurer et al., 2020). Thus, COR should be promoted, for example, by offering courses on web search strategies at the university library and by fostering COR skills in lectures or seminars in a targeted manner. Students should not only be taught suitable strategies for searching the Internet but also criteria and techniques for judging the credibility of (online) sources and information (e.g., Konieczny, 2014).

Seeing as web searches are firmly ingrained in teaching and task requirements in the university context nowadays, it is important to consider the influence the prompt of a task can have on students' search behavior and their COR. The identified effects of whether or not a link to a specific website was included in the task prompt on the students' search behavior should also be taken into account, both in future studies and with regard to designing new exercises in teaching. Although first efforts and successes have been reported when it comes to promoting these skills through targeted intervention measures aimed at students (McGrew et al., 2019), COR is still not an integral part of teaching at many universities (Persike and Friedrich, 2016). There is, therefore, an urgent need for instructional action in this context, especially as studies indicate that the specific teaching methods of individual universities have an influence on the students' use of online media and on which sources they tend to use (Persike and Friedrich, 2016).

CONCLUSION

This study provides an insight into the so far under researched relationship between students' web search behavior when evaluating online sources, the characteristics of the visited websites, and the information the students used. Our findings provide insights not only into students' preferences for certain types of websites in online searches and their quality, but also into the relationship between the characteristics of these websites and students' performance in the CORA tasks. In this respect, this study contributes to previous research, which had been mainly focused on students' website preferences for learning or for private purposes and, moreover, often collected this data through self-reports or in a simulated test environment. In particular, while previous studies on students' abilities and skills related to COR (e.g., searching and evaluation strategies) had been primarily focused on the test results (i.e., the score) and to what extent it is influenced by personal characteristics, this contribution analyzes the connection between students' search behavior during task processing and characteristics of the particular website they used in more detail.

Based on the unique analyses and results, this study highlights that there is a clear need for students to receive targeted support in higher education, which should be urgently addressed by implementing appropriate measures, as the ability to use online resources and critical online reasoning in a competent manner constitute not only an important basis for academic success but also for lifelong learning and for participation in society as an informed citizen.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2020.565062/full#supplementary-material

REFERENCES

Allen, M. (2008). Promoting critical thinking skills in online information literacy instruction using a constructivist approach. *College Undergr. Libr.* 15, 1–2. doi: 10.1080/10691310802176780

Andersen, K., Bjarnøe, C., Albæk, E., and de Vreese, C. H. (2016). How news type matters. Indirect effects of media use on political participation through knowledge and efficacy. *J. Media Psychol.* 28, 111–122. doi: 10.1027/1864-1105/a000201

Anmarkrud, Ø, Bråten, I., and Strømsø, H. I (2014). Multiple-documents literacy: strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learn. Individ. Diff.* 30, 64–76. doi: 10.1016/j.lindif.2013.01.007

Barczyk, B. C., and Duncan, D. G. (2013). Facebook in higher education courses: an analysis of students' attitudes,community of practice, and classroom community. *CSCanada* 6, 1–11. doi: 10.3968/j.ibm.1923842820130601.1165

Beaudoin, M. F. (2002). Learning or lurking? *Internet High. Educ.* 5, 147–155. doi: 10.1016/s1096-7516(02)00086-6

Beisch, N., Koch, W., and Schäfer, C. (2019). ARD/ZDF-Onlinestudie 2019. Mediale internetnutzung und video-on-demand gewinnen weiter an bedeutung. *Media Perspekt.* 9, 374–388.

Blossfeld, H. P., Bos, W., Daniel, H. D., Hannover, B., Köller, O., Lenzen, D., et al. (2018). *Digitale Souveränität und Bildung*. Gutachten: Waxmann.

Braasch, L. G., Bråten, I., and McCrudden, M. T. (2018). *Handbook of Multiple Source Use*. London: Routledge.

Brand-Gruwel, S., Kammerer, Y., van Meeuwen, L., and van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *J. Comput. Assist. Learn.* 33, 234–251. doi: 10.1111/jcal.12162

Brand-Gruwel, S., Wopereis, I., and Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Comput. Educ.* 53, 1207–1217. doi: 10.1016/j.compedu.2009.06.004

Bråten, I., Strømsø, H. I., and Salmeron, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learn. Instr.* 21, 180–192. doi: 10.1016/j.learninstruc.2010.02.002

Breakstone, J., Smith, M., and Wineburg, S. (2019). *Students' Civic Online Reasoning. A National Portrait*. Available online at: https://stacks.stanford.edu/

file/gf151tb4868/Civic%20Online%20Reasoning%20National%20Portrait.pdf (accessed May 16, 2020).

Brooks, C. (2016). *ECAR study of students And Information Technology*. Louisville: ECAR.

Brox, H. (2012). The elephant in the room. A place for wikipedia in higher education? *Nordlit* 16:143. doi: 10.7557/13.2377

Bullen, M., Morgan, T., and Qayyum, T. (2011). Digital learners in higher education. Generation is not the issue. *Can. J. Learn. Technol.* 37, 1–24.

Carbonell, X., Chamarro, A., Oberst, U., Rodrigo, B., and Prades, M. (2018). Problematic use of the internet and smartphones in University Students: 2006–2017. *Int. J. Environ. Res. Public Health* 15:475. doi: 10.3390/ijerph15030475

Ciampaglia, G. L. (2018). "The digital misinformation pipeline," in *Positive Learning in the Age of Information*, eds O. Zlatkin-Troitschanskaia, G. Wittum, and A. Dengel (Wiesbaden: Springer), 413–421. doi: 10.1007/978-3-658-19567-0_25

Dalrymple, K. E., and Scheufele, D. A. (2007). Finally informing the electorate? How the Internet got people thinking about presidential politics in 2004. *Harvard Int. J. Press Polit.* 12, 96–111. doi: 10.1177/1081180X07302881

Dimitrova, D., Shehata, A., Strömbäck, J., and Nord, L. W. (2011). The effects of digital media on political knowledge and participation in election campaigns. *Commun. Res.* 41:1. doi: 10.1177/0093650211426004

Eppler, M. J., and Mengis, J. (2004). The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. *Inform. Soc.* 20, 325–344. doi: 10.1080/01972240490507974

European Commission (2018). *Fake News and Disinformation Online*. Brussels: European Commission.

Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus on Educational Assessment and Instruction. Research Findings and Recommendations*. Newark, NJ: American Philosophical Association (ERIC).

Feezell, J., and Ortiz, B. (2019). 'I saw it on Facebook'. An experimental analysis of political learning through social media. *Inform. Commun. Soc.* 92:3. doi: 10.1080/1369118X.2019.1697340

Ferrari, A. (2013). *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*. Brussels: European Commission.

Flanagin, A. J., and 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 Soc.* 9, 319–342. doi: 10.1177/1461444807075015

Früh, W. (2017). *Inhaltsanalyse*. Stuttgart: UTB.

Gadiraju, U., Yu, R., Dietze, S., and Holtz, P. (2018). "Analyzing knowledge gain of users in informational search sessions on the web," in *CHIIR '18: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, New York, NY.

Gikas, J., and Grant, M. (2013). Mobile computing devices in higher education. Student perspectives on learning with cellphones, smartphones & social media. *Internet High. Educ.* 19, 18–26. doi: 10.1016/j.iheduc.2013.06.002

Gilster, P. (1997). *Digital Literacy*. New York: Wiley Computer Publications.

Gladney, G. A., Shapiro, I., and Castaldo, J. (2007). Online editors rate Web news quality criteria. *Newspaper Res. J.* 28, 1–55. doi: 10.1177/073953290702800105

Guess, A., Nagler, J., and Tucker, J. (2019). Less than you think. Prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* 5:eaau4586. doi: 10.1126/sciadv.aau4586

Hague, C., and Payton, S. (2010). *Digital Literacy Across the Curriculum*. Bristol: Futurelab.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., and Yates Thomas, K. (2010). Trust online: young adults' evaluation of web content. *Int. J. Commun.* 4, 468-494.

Heinström, J. (2002). *Fast Surfers, Broad Scanners and Deep Divers. Personality and Information-Seeking Behaviour*. Dissertation, Akademi University Press, Åbo.

Helms-Park, R., Radia, P., and Stapleton, P. (2007). A preliminary assessment of Google Scholar as a source of EAP students' research materials. *Internet High. Educ.* 10, 65–76. doi: 10.1016/j.iheduc.2006.10.002

Hölscher, C., and Strube, G. (2000). Web search behaviour of Internet experts and newbies. *Comput. Netw.* 33, 1–6. doi: 10.1016/S1389-1286(00)00031-1

IBM Corp (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.

JISC (2014). *Developing Digital Literacies*. Available online at: https://www.jisc.ac.uk/guides/developing-digital-literacies (accessed October 22, 2020).

Jones, C., and Healing, G. (2010). Net generation students. Agency and choice and the new technologies. *J. Comput. Assist. Learn.* 26, 344–356. doi: 10.1111/j.1365-2729.2010.00370.x

Judd, T., and Kennedy, G. (2011). Expediency-based practice? Medical students' reliance on Google and Wikipedia for biomedical inquiries. *Br. J. Educ. Technol.* 42, 351–360. doi: 10.1111/j.1467-8535.2009.01019.x

Kao, G. Y., Lei, P.-L., and Sun, C.-T. (2008). Thinking style impacts on Web search strategies. *Comput. Hum. Behav.* 24, 1330–1341. doi: 10.1016/j.chb.2007.07.009

Kelly, D. (2018). Evaluating the news: (Mis)perceptions of objectivity and credibility. *Polit. Behav.* 41, 445–471. doi: 10.1007/s11109-018-9458-4

Kennedy, G. E., Judd, T. S., Churchward, A., Gray, K., and Krause, K. (2008). First year students' experiences with technology. Are they really digital natives? *AJET* 24, 108–122. doi: 10.14742/ajet.1233

Kim, J. (2008). Task as a context of information seeking: an investigation of daily life tasks on the web. *Libri* 58:172. doi: 10.1515/libr.2008.018

Knight, C., and Pryke, S. (2012). Wikipedia and the University, a case study. *Teach. High. Educ.* 17, 649–659. doi: 10.1080/13562517.2012.666734

Konieczny, P. (2014). Rethinking Wikipedia for the classroom. *Contexts* 13, 80–83. doi: 10.1177/1536504214522017

Kopp, M., Gröblinger, O., and Adams, S. (2019). Five common assumptions that prevent digital transformation at higher education Institutions. *INTED2019 Proc.* 160, 1448–1457. doi: 10.21125/inted.2019.0445

Kruse, O. (2017). *Kritisches Denken und Argumentieren*. Konstanz: UVK.

Leighton, J. (2017). *Using Think-Aloud Interviews and Cognitive Labs in Educational Research*. Oxford: Oxford University Press.

List, A., and Alexander, P. A. (2017). Text navigation in multiple source use. *Comput. Hum. Behav.* 75, 364–375. doi: 10.1016/j.chb.2017.05.024

Lucassen, T., Muilwijk, R., Noordzij, M. L., and Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *J. Am. Soc. Inform. Sci. Technol.* 64, 254–264. doi: 10.1002/asi.22743

Mason, L., Boldrin, A., and Ariasi, N. (2010). Epistemic metacognition in context: evaluating and learning online information. *Metacogn. Learn.* 5, 67–90. doi: 10.1007/s11409-009-9048-2

Maurer, M., Quiring, O., and Schemer, C. (2018). "Media effects on positive and negative learning," in *Positive Learning in the Age of Information*, eds O. Zlatkin-Troitschanskaia, G. Wittum, and A. Dengel (Wiesbaden: Springer), 197–208. doi: 10.1007/978-3-658-19567-0_11

Maurer, M., Schemer, C., Zlatkin-Troitschanskaia, O., and Jitomirski, J. (2020). *Positive and Negative Media Effects on University Students' Learning: Preliminary Findings and a Research Program*. New York, NY: Springer.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., and Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory Res. Soc. Educ.* 46, 165–193. doi: 10.1080/00933104.2017.1416320

McGrew, S., Ortega, T., Breakstone, J., and Wineburg, S. (2017). The challenge that's bigger than fake news. Civic reasoning in a social media environment. *Am. Educ.* 41, 4–9.

McGrew, S., Smith, M., Breakstone, J., Ortega, T., and Wineburg, S. (2019). Improving university students' web savvy: an intervention study. *Br. J. Educ. Psychol.* 89, 485–500. doi: 10.1111/bjep.12279

McQuail, D. (1992). *Media Performance. Mass Communication and the Public Interest*. London: Sage.

Molerov, D., Zlatkin-Troitschanskaia, O., and Schmidt, S. (2019). Adapting the civic online reasoning assessment for cross-national use. *Paper presented at Annual Meeting of the American Educational Research Association*, (Toronto: APA).

Murray, M. C., and Pérez, J. (2014). Unraveling the digital literacy paradox: how higher education fails at the fourth literacy. *Issues Inform. Sci. Inform. Technol.* 11, 85–100. doi: 10.28945/1982

Nagler, W., and Ebner, M. (2009). "Is Your University ready for the ne(x)t-generation?," in *Proceedings of 21st ED-Media Conference*, New York, NY, 4344–4351.

Newman, T., and Beetham, H. (2017). *Student Digital Experience Tracker 2017: The Voice of 22,000 UK Learners*. Bristol: JISC.

Oblinger, D. G., and Oblinger, J. L. (2005). *Educating the Net Generation. n.p.* Washington, DC: Educause.

Persike, M., and Friedrich, J. D. (2016). *Lernen mit digitalen Medien aus Studierendenperspektive. Arbeitspapier.* Berlin: Hochschulforum Digitalisierung.

Prensky, M. (2001). Digital natives, digital immigrants. *On Horizon* 9, 1–6. doi: 10.1002/9781118784235.eelt0909

Riffe, D., Lacy, S., Fico, F., and Watson, B. (2019). *Analyzing Media Messages. Using Quantitative Content Analysis in Research.* New York: Routledge.

Roblyer, M. D., McDaniel, M., Webb, M., Herman, J., and Witty, J. V. (2010). Findings on Facebook in higher education. A comparison of college faculty and student uses and perceptions of social networking sites. *Internet High. Educ.* 13, 134–140. doi: 10.1016/j.iheduc.2010.03.002

Roscoe, R. D., Grebitus, C., O'Brian, J., Johnson, A. C., and Kula, I. (2016). Online information search and decision making. Effects of web search stance. *Comput. Hum. Behav.* 56, 103–118. doi: 10.1016/j.chb.2015.11.028

Scheufele, D., Hardy, B., and Brossard, D. (2006). Democracy based on difference: examining the links between structural heterogeneity, heterogeneity of discussion networks, and democratic citizenship. *J. Commun.* 56, 728–753. doi: 10.1111/j.1460-2466.2006.00317.x

Selwyn, N., and Gorard, S. (2016). Students' use of Wikipedia as an academic resource — Patterns of use and perceptions of usefulness. *Internet High. Educ.* 28, 28–34. doi: 10.1016/j.iheduc.2015.08.004

Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., and Marino, J. P. (2019). Assessment of University students' critical thinking: next generation performance assessment. *Int. J. Test.* 19, 337–362. doi: 10.1080/15305058.2018.1543309

Stanford, J., Tauber, E. R., Fogg, B. J., and Marable, L. (2002). *Experts vs. Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites.* Cham: Springer.

StataCorp (2017). *Stata Statistical Software: Release 15.* College Station, TX: StataCorp LLC.

Steiner, M., Magin, M., and Stark, B. (2018). Uneasy bedfellows. *Digital J.* 7, 100–123. doi: 10.1080/21670811.2017.1412800

Strømsø, H. I., Bråten, I., Britt, M. A., and Ferguson, L. (2013). Spontaneous sourcing among students reading multiple documents. *Cogn. Instr.* 31, 176–203. doi: 10.1080/07370008.2013.769994

Tess, P. A. (2013). The role of social media in higher education classes (real and virtual) – A literature review. *Comput. Hum. Behav.* 29, A60–A68. doi: 10.1016/j.chb.2012.12.032

Thompson, P. (2013). The digital natives as learners. Technology use patterns and approaches to learning. *Comput. Educ.* 65, 12–33. doi: 10.1016/j.compedu.2012.12.022

Tombros, A., Ruthven, I., and Jose, J. M. (2005). How users assess web pages for information seeking. *J. Am. Soc. Inform. Sci. Technol.* 56, 327–344. doi: 10.1002/asi.20106

Tribukait, M., Baier, K., Grzempa, H., Loukovitou, A., Sijakovic, R., Tettschlag, N., et al. (2017). Digital learning in European education policies and history curricula. *Eckert. Dossiers* 13:417.

Urban, J., and Schweiger, W. (2014). News quality from the recipients' perspective. *J. Stud.* 15, 821–840. doi: 10.1080/1461670X.2013.856670

Walraven, A., Brand-Gruwel, S., and Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Comput. Educ.* 52, 234–246. doi: 10.1016/j.compedu.2008.08.003

Wannemacher, K., and Schulenberg, F. (2010). "Wikipedia in academic studies: corrupting or improving the quality of teaching and learning?," in *Looking Toward the Future of Technology-Enhanced Education. Ubiquitous Learning and The Digital Native*, eds M. Ebner and M. Schiefner (Hershey, PA: IGI Global), 296–312.

Wellbrock, C. (2011). Die journalistische Qualität deutscher Tageszeitungen – Ein Ranking. *MedienWirtschaft.* 8:2. doi: 10.15358/1613-0669-2011-2-22

White, R. W., Dumais, S. T., and Teevan, J. (2009). "Characterizing the influence of domain expertise on web search behavior," in *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09*, New York, NY.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., and Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *Am. Educ. Res. J.* 46, 1060–1106. doi: 10.3102/0002831209333183

Wineburg, S., Breakstone, J., McGrew, S., and Ortega, T. (2016). *Evaluating Information: The Cornerstone of Civic Online Reasoning. Stanford History Education Group.* Available online at: https://stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf (accessed May 10, 2020).

Wineburg, S., and McGrew, S. (2016). *Why Students Can't Google Their Way to the Truth: Fact-checkers and students approach websites differently. Educational Week.* Available online at: https://www.edweek.org/ew/articles/2016/11/02/why-students-cant-google-their-way-to.html (accessed May 10, 2020).

Wolfsfeld, G., Yarchi, M., and Samuel-Azran, T. (2016). Political information repertoires and political participation. *New Med. Soc.* 18:9. doi: 10.1177/1461444815580413

Yadav, R., Tiruwa, A., and Suri, P. K. (2017). Internet based learning (IBL) in higher education: a literature review. *J. Int. Educ. Bus.* 10:2. doi: 10.1108/JIEB-10-2016-0035

Zaller, J. (2003). A new standard of news quality: burglar alarms for the monitorial citizen. *Polit. Commun.* 20, 109–130. doi: 10.1080/10584600390211136

Zimmermann, F., and Kohring, M. (2018). "Fake News" als aktuelle desinformation. Systematische bestimmung eines heterogenen Begriffs. *Med. Kommunikationswissenschaft* 66, 526–541. doi: 10.5771/1615-634x-2018-4-526

Zlatkin-Troitschanskaia, O., Beck, K., Fischer, J., Braunheim, D., Schmidt, S., and Shavelson, R. J. (2020). The role of students' beliefs when critically reasoning from multiple contradictory sources of information in performance assessments. *Front. Educ.* 11:2192. doi: 10.3389/fpsyg.2020.02192

Zlatkin-Troitschanskaia, O., Jitomirski, J., Happ, R., Molerov, D., Schlax, J., Kühling-Thees, C., et al. (2019). Validating a test for measuring knowledge and understanding of economics among university students. *Z. Pädagogische Psychol.* 33, 119–133. doi: 10.1024/1010-0652/a000239