A user study on people's perception to the credibility of online health information

Marcos Fernández-Pichel[1,*], Markus Bink[2], David E. Losada[1] and David Elsweiler[2]

[1]Centro de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain

[2]Chair of Information Science, Universität Regensburg, Regensburg, Germany

Abstract

Judging the credibility of information is a subjective process and prone to biases. This issue can be especially concerning in health information seeking. Some efforts have been made to define robust credibility assessment guidelines that support the development of reliable test collections. This is of the utmost importance since the applicability of retrieval algorithms to real use case scenarios relies on the quality of the labelled data. Yet, the question persists as to whether the labels created by these guidelines can effectively serve as a surrogate for the genuine judgements of credibility as perceived by end-users. Motivated by this, we conducted a user study with 1,000 participants. We demonstrate that there is a correlation between participants' judgements and the reference values produced following existing guidelines. Further analyses of the data reveal worrying insights into people's ability to judge the credibility of online medical content, leading to potential personal harm.

Keywords

Health-related content, Credibility, User study

1. Introduction

The Internet has become the dominant platform for accessing health information, offering convenient access to a wealth of medical knowledge [1, 2, 3]. Nonetheless, the abundance of information poses a challenge for users in discerning trustworthy sources from unreliable ones, potentially resulting in ill-informed choices regarding their health [4, 5, 6, 7]. In extreme cases, this situation can have severe consequences and even poses a risk to personal well-being [8]. Credibility has been defined as the extent to which information from a webpage or other online source can be believed [9]. It is a highly subjective concept that is susceptible to individual differences, such as user's reading skills [10, 11]. The subjective nature of credibility represents a barrier in creating reliable and robust test collections. In the context of shared-task evaluation campaigns, some researchers have critically analysed the quality of the credibility assessments and proposed a set of robust and traceable guidelines to improve the robustness of

ROMCIR 2024: The 4th Workshop on Reducing Online Misinformation through Credible Information Retrieval (held as part of ECIR 2024: the 46th European Conference on Information Retrieval), March 24, 2024, Glasgow, UK

[*]Corresponding author.

marcosfernandez.pichel@usc.es (M. Fernández-Pichel); markus.bink@ur.de (M. Bink); david.losada@usc.es (D. E. Losada); david.elsweiler@ur.de (D. Elsweiler)

0000-0002-6560-9832 (M. Fernández-Pichel); 0000-0002-3444-0990 (M. Bink); 0000-0001-8823-7501 (D. E. Losada); 0000-0002-5791-0641 (D. Elsweiler)

annotations [12]. This is important since the applicability of retrieval and machine learning algorithm relies on the quality of the annotation process.

Nevertheless, there is still a need for a rigorous examination of the relationship between this type of guidelines and credibility as real end-users perceive it. While we know that judgements vary across users, annotations need to be both consistent and reflective of average users' perceptions. Previous research has studied the main elements influencing individual credibility perceptions [13, 14]. However, no user-oriented study has attempted to understand annotation practices for shared-tasks. Such a study could also provide valuable cues on how people evaluate the credibility of websites posting medical information.

In this work, we perform a study to understand how end-users perceive the credibility of online health information[1]. The ultimate goal being to determine whether the labels created from guidelines can serve as surrogate of credibility perceived by end-users. We attempt to answer the following research questions:

- **RQ1.** Can current credibility annotation guidelines act as a proxy of the real perception of credibility of end-users?
- **RQ2.** To what extent users are able to recreate the judgements of experts?
- **RQ3.** How do user variables, such as familiarity with the search topic, educational background, and other human factors, affect the user's perception of credibility?

2. Related work

The credibility of online information and the spread of misinformation have been extensively studied [15, 16, 17, 18]. Viviani and Pasi reviewed the main automatic methods to estimate credibility in social media, focusing mainly on health content [19]. A further body of work has sought to understand how end-users assess the credibility of online content and why people make certain assessments. For instance, Fogg defined the *prominence-interpretation theory*, which helps to determine which website elements influence end-users' credibility [13]. This theory was later tested through a user study involving 2,500 participants, where authors found that 46% of the users mentioned design as a critical aspect influencing credibility [20].

Easting et al. [21] demonstrated that both the source and the prior knowledge about the content have influence on users' perception of online health information. Other studies have also demonstrated that, apart from the characteristics of the web elements, the receiver's characteristics also influence the perception of the information [22]. Other researchers analysed in-depth the factors that influence end-users' perceived credibility [15]. Previous studies have also evaluated the correlation between different users' judgements to test their feasibility as ground truth values [23, 24].

In this paper, we present a systematic user study that shows how well expert annotations reflect the subjective judgements of a broad population of users. We also evaluate a number of personal factors that may influence the credibility estimations.

[1] https://github.com/MarcosFP97/perceived-credibility-study

Topic id	Question	Number of docs.
T1	*Do antioxidants help female subfertility?*	41
T5	*Do sealants prevent dental decay in permanent teeth?*	45
T8	*Does melatonin help treat and prevent jetlag?*	39
T10	*Does traction help low back pain?*	37

Table 1
Health topics in the user study

3. Experimental setup

We hypothesise that reference values (created by expert annotators using formal guidelines) will correlate with users' judgements. To test this, we conducted a crowd-sourced user study whereby participants provided credibility judgements for webpages.

3.1. Dataset

We utilised a pre-existing dataset from the medical domain that had been originally compiled by Pogacar et al. [4] and later extended by Zimmerman et al. [25]. We extracted 162 screenshots of webpages from it, such that the selected webpages provide answers spanning four distinct health-related topics, as detailed in Table 1[2]. Each participant was presented with a full-scale screenshot of a randomly selected webpage and was asked to assess the webpage's credibility on a 7-point Likert-scale. Each webpage was evaluated at least 5 times by different participants to minimise personal bias [18, 10].

For these webpages, we also produced annotations generated by human assessors according to the guidelines from [12], detailed in Table 2. We recruited four different assessors[3]. For a given topic, the webpages were annotated by the same pair of assessors. Next, the two assessors responsible for each topic convened a meeting to discuss and consolidate their annotations and generate a final set of labels. These final annotations were used as reference values in our user study.

3.2. Variables

3.2.1. Independent variables

- **Reference values**: variable indicating the credibility-level perceived by the human annotators according to the guidelines. There are three possible levels: **0 (non-credible)**, **1 (credible)**, and **2 (highly credible)**.

3.2.2. Dependent variables

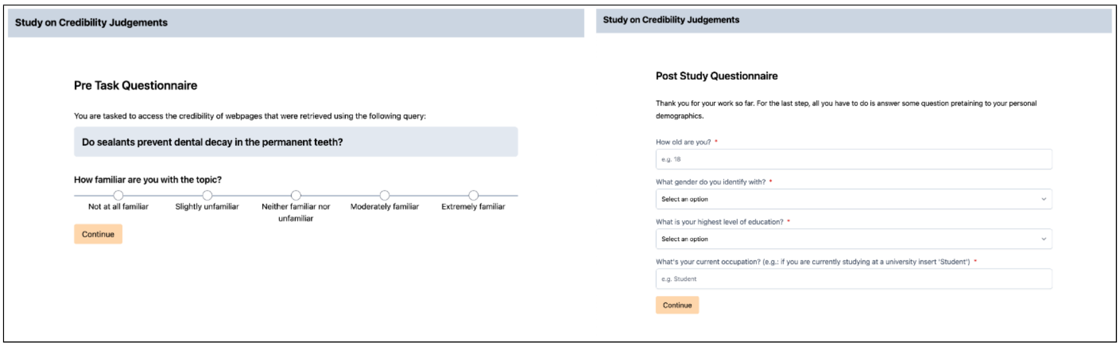
- **User credibility score**: the credibility score assigned by the crowdsourcers in a 7-point Likert scale.

[2] We did not use the raw HTML, since we consider visual elements as key to the perception of credibility.

[3] Three PhD students with background in British Studies, Computer Science and Information Science, respectively, and a Master's Degree student in Information Science

	Label	Guideline
G1	2	Source is a scientific paper, or a Medical publisher or hospital/clinic or government website or university.
G2	1	Document is citing the information they provide in their articles. They provide links or specific references to their sources. They cite sources with credibility 2 (i.e. medical publications and/or lab studies).
G3	1	Document is written by an expert in the field/someone qualified to write this document (irrespective of publishing venue).
G4	0	The document is actually for advertising or marketing purposes. If so, the website might be biased or a scam designed to trick people into fake treatments or into buying medical products that do not live up to their claim.
G5	0	The information posted by a non-expert person providing a medical product review or providing medical advice without proper citations (links/list of references).
G6	0	The website provides or states claims that go against well-known medical consensus (e.g. smoking cigarettes does not cause cancer).
NOTE: It is generally allowed to look up authors to check whether they have the required knowledge to be regarded as an expert and look up websites to find out if they are legitimate.		

Table 2
Guidelines proposed in Fernández-Pichel et al. [12].



Study on Credibility Judgements

Pre Task Questionnaire

You are tasked to access the credibility of webpages that were retrieved using the following query:

Do sealants prevent dental decay in the permanent teeth?

How familiar are you with the topic?

○ Not at all familiar ○ Slightly unfamiliar ○ Neither familiar nor unfamiliar ○ Moderately familiar ○ Extremely familiar

Continue

Study on Credibility Judgements

Post Study Questionnaire

Thank you for your work so far. For the last step, all you have to do is answer some question pretaining to your personal demographics.

How old are you? *
e.g. 18

What gender do you identify with? *
Select an option

What is your highest level of education? *
Select an option

What's your current occupation? (e.g.: if you are currently studying at a university insert 'Student') *
e.g. Student

Continue

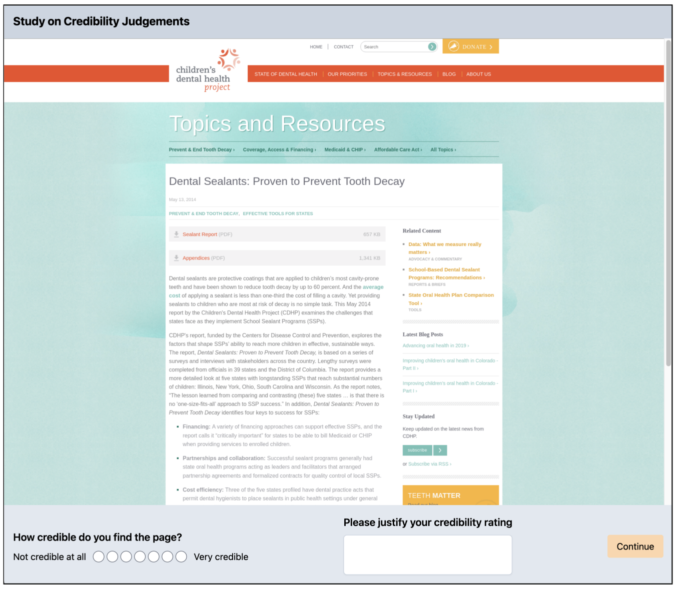
Figure 1: Pre-Task Questionnaire (left), which prompted the participant to give their self-assessed level of experience on the presented topic. Post-Study Questionnaire (right), which prompted the participant to enter demographic information such as age, gender, educational background and current occupation.

- **Time of completion (in minutes)**: variable representing the total amount of time it took for a crowdsourcer to complete the assessment (measured from the moment the screenshot of the website was shown until successful completion).

3.2.3. Descriptive and exploratory variables

We also studied some variables that can influence or have some connection with the credibility scores gathered in the study (this relation is further explored in Section 4):

- **Topic familiarity**: in the pre-task questionnaire, see Figure 1 (left side), participants were asked about their prior knowledge on the topic.


Study on Credibility Judgements
HOME | CONTACT
Search
DONATE ›
children's dental health project
STATE OF DENTAL HEALTH | OUR PRIORITIES | TOPICS & RESOURCES | BLOG | ABOUT US
Topics and Resources
Prevent & End Tooth Decay › Coverage, Access & Financing › Medicaid & CHIP › Affordable Care Act › All Topics ›
PREVENT & END TOOTH DECAY, EFFECTIVE TOOLS FOR STATES
Sealant Report (PDF) 657 KB
Appendices (PDF) 1,341 KB
Related Content
Data: What we measure really matters ›
ADVOCACY & COMMENTARY
School-Based Dental Sealant Programs: Recommendations ›
REPORTS & BRIEFS
State Oral Health Plan Comparison Tool ›
TOOLS
Latest Blog Posts
Advancing oral health in 2019 ›
Improving children's oral health in Colorado - Part II ›
Improving children's oral health in Colorado - Part I ›
Stay Updated
subscribe
or Subscribe via RSS ›
TEETH MATTER
Continue

Dental Sealants: Proven to Prevent Tooth Decay

May 13, 2014

Dental sealants are protective coatings that are applied to children's most cavity-prone teeth and have been shown to reduce tooth decay by up to 60 percent. And the average cost of applying a sealant is less than one-third the cost of filling a cavity. Yet providing sealants to children who are most at risk of decay is no simple task. This May 2014 report by the Children's Dental Health Project (CDHP) examines the challenges that states face as they implement School Sealant Programs (SSPs).

CDHP's report, funded by the Centers for Disease Control and Prevention, explores the factors that shape SSPs' ability to reach more children in effective, sustainable ways. The report, *Dental Sealants: Proven to Prevent Tooth Decay,* is based on a series of surveys and interviews with stakeholders across the country. Lengthy surveys were completed from officials in 39 states and the District of Columbia. The report provides a more detailed look at five states with longstanding SSPs that reach substantial numbers of children: Illinois, New York, Ohio, South Carolina and Wisconsin. As the report notes, "The lesson learned from comparing and contrasting (these) five states … is that there is no 'one-size-fits-all' approach to SSP success." In addition, *Dental Sealants: Proven to Prevent Tooth Decay* identifies four keys to success for SSPs:

- **Financing:** A variety of financing approaches can support effective SSPs, and the report calls it "critically important" for states to be able to bill Medicaid or CHIP when providing services to enrolled children.
- **Partnerships and collaboration:** Successful sealant programs generally had state oral health programs acting as leaders and facilitators that arranged partnership agreements and formalized contracts for quality control of local SSPs.
- **Cost efficiency:** Three of the five states profiled have dental practice acts that permit dental hygienists to place sealants in public health settings under general

Keep updated on the latest news from CDHP.

How credible do you find the page?

Not credible at all ○○○○○○○ Very credible

Please justify your credibility rating

Figure 2: Main-Study Screen, which presented participants with a screenshot of the assigned webpage. The bottom of the page presented a Likert scale to rate the perceived credibility of the website and a (non-mandatory) free-text input field to justify the judgement.

- **Personal data**: in a post-study questionnaire, see Figure 1 (right side), we gathered additional information about the participants in the study (educational background, gender, and age).
- **Justifications**: we also provided the participants with the possibility of justifying their rating in their own words (free text field).

3.3. Procedure

Once participants had been presented with the goals of the study, its methodology, and the implications of their involvement, they provided their permission by signing a consent form. Next, they proceeded with 3 steps to satisfactorily complete the study:

1. Each participant was randomly assigned one webpage from the collection. Before seeing the webpage's screenshot, they needed to fulfil a pre-task questionnaire about their expertise on the topic of the website, see Figure 1 (left side).
2. Subjects were shown a screenshot of the entire webpage and they needed to assess its credibility in a 7-point Likert scale (from *not credible at all* to *very credible*), see Figure 2. They had no time limit to provide this estimate, and they could scroll through the entire screenshot and provide a free-text justification about their judgement (this step was not mandatory, however a high number of participants provided this feedback, see Section 4.6).
3. Before ending the study, participants were shown a post-study questionnaire, see Figure 1 (right side). Our main goal was to gather additional data, such as educational background, age, and gender.

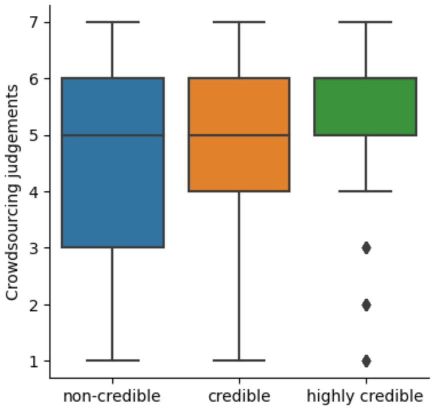


Figure 3: Boxplot of the crowdsourced assessments for the three types of webpages presented (non-credible, credible and highly credible).

3.4. Participants

We recruited a total of 1,000 users to guarantee at least 5 judgements per webpage. We used the *Prolific*[4] platform and each participant received £0.32 (equivalent to £9.60 per hour). Our participants were fluent English speakers, resident in the United States or the United Kingdom. The annotators belonged to an age range between 18 and 85 years old. 53% identified as female, 45% as male, and the remaining participants identified either as diverse or other. In terms of educational background, 40% had a bachelor's degree, 37% completed secondary education, and only 2.6% of the participants reported a level of education below high school.

4. Results

4.1. Reference credibility values vs users' judgements

RQ1 seeks to determine whether the current annotation guidelines, whose goal is to produce robust assessments of credibility for medical websites, serve as a proxy for the human's perception of credibility [12]. We analysed the distribution of the crowdsourced judgments according to the three levels of reference values. As can be seen in Figure 3, it seems that there is a relationship between both variables. The participants' judgements tend to be higher when presented with webpages of increasing credibility (according to experts). This is confirmed by a Spearman's rank correlation ($\rho = 0.26$ and a $p - value < 0.01$), indicating weak agreement according to [26].

Despite this correlation, there are some signs of concern regarding **RQ2**. Webpages annotated as non-credible according to the guidelines were often perceived as reliable by the participants. This can be observed by the fact that non-credible documents have very high perception scores and their median score is 5. This confirms previous research findings that people tend to overestimate credibility and have problems identifying low-quality sites [27, 28]. In general, webpages labelled as credible or highly credible by the reference judgements were

[4] https://www.prolific.co/

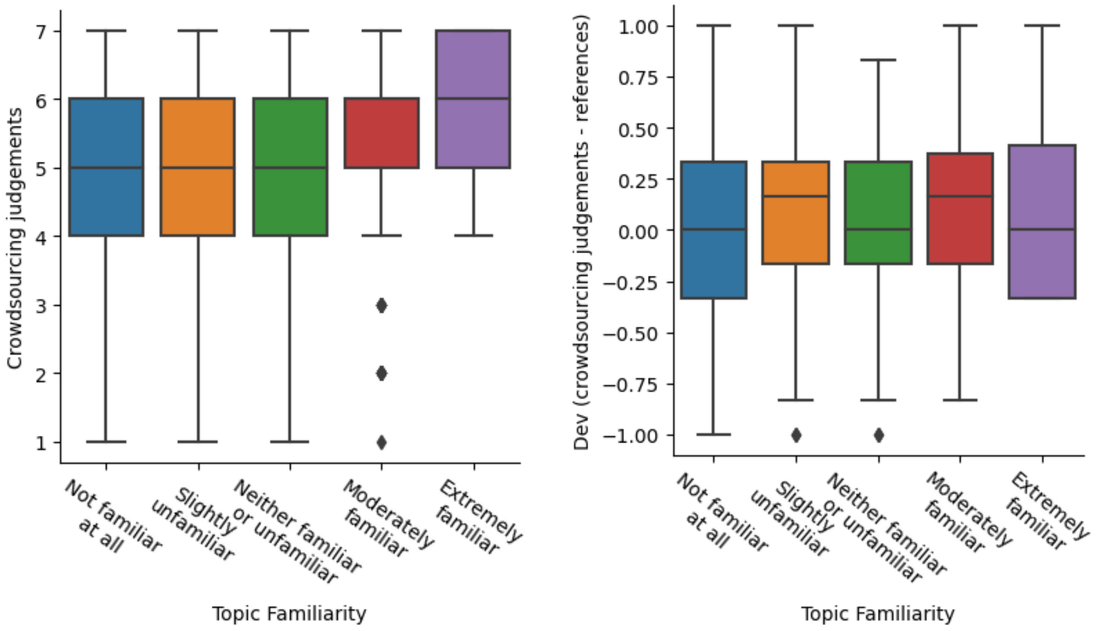


Figure 4: In the left plot, we can find crowdsourced credibility values grouped by familiarity with the topic. The right plot represents the deviations between the crowdsourced credibility values and the reference values (also grouped by familiarity). Positive (negative) values represent cases where users overestimated (underestimated) the credibility of the webpage.

also considered as of high quality by the crowdworkers. However, people struggled to detect contents that are regarded as low quality by the reference annotations.

Summing up, regarding RQ1, we can conclude that the participants' judgements and the reference values derived from the guidelines are correlated. As for RQ2, we found out that crowdworkers are less prone to errors when evaluating high quality pages, but they struggled for pages with lower levels of credibility.

4.2. Topic familiarity

Prior to completing the study and to partially answer **RQ3**, we asked participants about their level of knowledge or familiarity with their assigned topic in a 5-point Likert scale (ranging from *Not at all familiar* to *Extremely familiar*, see Figure 1 (left side)).

Figure 4 (left side) shows the relation between levels of familiarity and the participants' judgements. It seems that the higher the familiarity, the higher the credibility judgements provided by participants. Spearman's correlation yielded a $\rho = 0.10$ and $p - value < 0.01$, demonstrating that there is a very weak correlation between the two variables [26].

To further explore the user study data, we also computed the deviation per level of familiarity between the reference values and the user study's judgements. First, we applied a Min-Max normalisation to both sets of scores. Then, the difference between the crowdsourcing judgements and the reference values was computed −0 represents a perfect match, while a positive (negative) value means that people overestimated (underestimated) credibility– see Figure 4 (right side). From the figure, we might conclude that there is not a strong relation between familiarity and how effective are users at rating webpages. However, Spearman's correlation yielded a $\rho = 0.13$ and $p - value < 0.01$.

As a complementary analysis, we also computed the mean familiarity per topic (and its standard deviation): T1 has a mean familiarity of 1.55 (0.89), T5 of 1.79 (1.07), T8 of 2.44 (1.25), and T10 of 2.14 (1.15).

4.3. Topic analysis

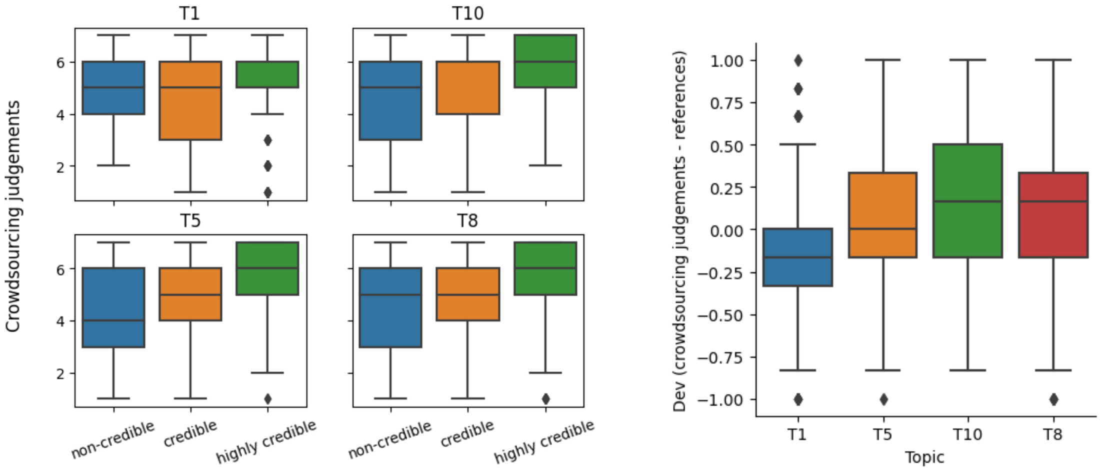


Figure 5: In the left plot, we show boxplots (per topic) of the crowdsourced assessments for the three types of webpages presented (non-credible, credible and highly credible). The right plot shows the deviations between the user credibility values and the reference values.

Figure 5 (left side) reports a topic-level analysis. As can be expected, there are individual differences among the topics. Spearman's test also revealed statistically significant correlations between reference credibility scores and crowdsourced credibility scores for all topics. However, the correlation for T1 was lower. These results fit with the familiarity scores described above, where T1 was shown to be the topic that users had less knowledge about.

Again, we also computed the deviation per topic between the reference values and the user study's judgements, see Figure 5 (right side). Spearman's test revealed an statistically significant correlation between both variables ($\rho = 0.20$ and $p - value < 0.01$). An interesting finding is that for the topics users are more familiar with, T8 and T10, they tend to overestimate their perceived credibility.

4.4. Other user variables

To fully answer **RQ3**, the relation between additional user variables (gender, age, and educational background) and their perception of credibility was explored. For the first two variables, no significant correlations or revealing trends were found. However, for the educational background

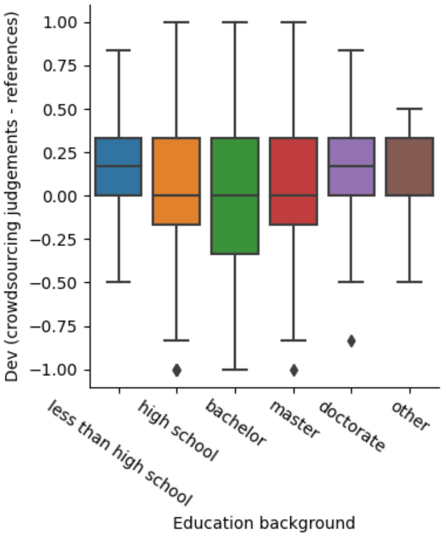


Figure 6: Box plots (per educational background) representing the deviations between the user credibility values and the reference values. Positive (negative) values represent cases where users overestimated (underestimated) the credibility of the webpage.

(Figure 6), we found an interesting conclusion: all groups were equally good at estimating credibility, except for the less educated (*less than high school* group) and the most educated (*doctorate* group). The Spearman's test reported a $\rho = 0.05$ and $p - value = 0.12$ rejecting the hypothesis that there is a correlation between the educational level and the quality of the assessments, measured by the deviation between the crowdsourced and reference values.

4.5. Time of completion

We also analysed the time (in minutes) users needed to complete the assessment. Figure 7 shows that users who spent less time analysing the web (between 0-6 minutes) tended to deviate less from the reference values (deviation close to 0). We speculate that "overthinking" might be counterproductive for this task. Alternatively, the lower quality of the estimates at the right end of the graph could be due to other factors such as distractions. Related to this, previous studies showed that people who take more time on this type of tasks tend to be more influenced by visual elements and their prior knowledge [29]. In any case, the Spearman's correlation test revealed no statistical significance (with a $\rho = 0.009$ and a $p - value = 0.78$) between completion time and deviation between crowdsourced and reference judgements.

4.6. Analysis of justifications

We offered users the possibility of justifying their judgements. This was actively used by participants, with 93% providing a textual explanation. This gave us valuable evidence to analyse the reasons behind credibility judgements. Yet, manual inspection was infeasible because we had thousands of datapoints. We therefore opted for exploiting the summarisation capabilities

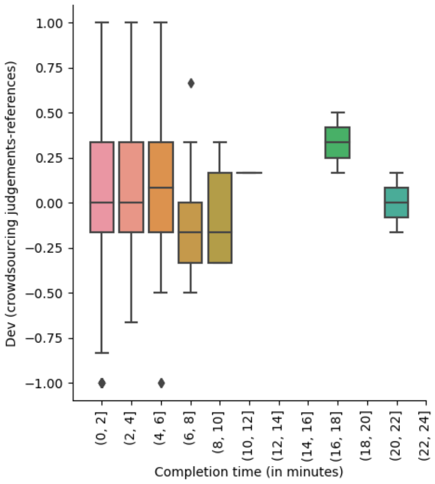


Figure 7: Box plots (per completion time, in minutes) between the user credibility values and the reference values. Positive (negative) values represent cases where users overestimated (underestimated) the credibility of the webpage.

of current Large Language Models (LLMs). To that end, the justifications were grouped by the different levels of perceived credibility and GPT-4 was provided with these textual extracts and asked to generate a summary for each level. The template used for prompting the LLM was as follows: "*We are a group of scientists that have conducted an online survey on webpage credibility. For each webpage, we asked a human assessor to provide a score credibility from 1 to 7 (very low credible to very credible). Assessors could also provide a justification on why they assigned a given credibility score. Given a series of justifications between <>, I want you to generate an understandable summary. <justifications>. The summary is:*". The resulting summaries are presented in Table 3.

Some interesting patterns were observed in the provided explanations. Low credibility was usually associated with a poor visual appearance of the webpage. Moreover, several users mentioned the lack of reputable references as a critical aspect to mistrust information. This suggests the existence of a direct relationship between the real perception of end-users and the existing guidelines [12], which incorporate criteria such as the presence of authoritative references. Regarding cases of high perception of credibility, users tend to trust professional appearance and well-structured content. They also mentioned the importance of the qualifications of the authors or the inclusion of credible sources. These two criteria were also considered in the original assessment guidelines (citing reputable sources or the expertise of the author). Crowdsourcers mentioned governmental sites, e.g. NHS or CDC, as the highest trustful sources. This also matches with the criteria stated in the guidelines. This qualitative analysis provides tangible evidence to support current guidelines as proxies of users' perceptions of credibility.

Perceived Credibility Level	Summary of the Justifications
1 (low credibility)	Many participants cited **unprofessional design**, **poor layout**, and lack of **reliable sources** as reasons for low credibility scores. Additionally, some expressed distrust towards **Wikipedia** and belief that the information could be biased or incorrect.
2	The justifications mainly indicate that the webpages appear **amateur**, **unprofessional**, and lacking **credible sources or evidence**. Additionally, many assessors express skepticism due to **outdated information**, **poor design**, and potential bias from **advertisements** or affiliate links.
3	The webpages were considered less credible due to factors such as **unprofessional design**, **outdated information**, **informal language**, **lack of evidence** or citations, and clickbait-like content. The credibility of some pages was questioned due to unfamiliar brands, **personal opinions**, and the absence of proper qualifications or accreditations.
4	Assessors found some webpages to be credible due to **professional appearance**, **links to studies** or association with reputable organizations, while others were seen as less credible due to informal language, lack of citations or references, and potential for errors. The credibility of some pages was difficult to judge without further investigation or knowledge of the subject matter.
5	The justifications highlight the presence of **credible sources**, **professional appearance**, and **author qualifications** as positive factors for credibility. However, some concerns are raised due to missing citations, outdated information, and potential biases.
6	Survey participants found the webpages credible due to their **professional appearance**, use of medical facts and **references**, reputable sources, **well-structured content**, and **qualified authors**. The credibility was also often influenced by **personal experiences** or previous knowledge about the subjects discussed.
7 (high credibility)	The majority of the justifications indicate that the webpages are credible due to their professional appearance, reputable sources, and being associated with trusted organizations such as the **NHS**, **CDC**, and various academic **journals**. Additionally, assessors mentioned the presence of **scientific research**, citations, **author credentials**, and detailed information as contributors to the credibility of the webpages.

Table 3
Summaries generated with LLMs of the justifications provided by the crowdsourcers. We have highlighted some key words that served as justification for user judgements

5. Discussion

In this study, we showed that there is correlation between the annotations produced from existing credibility guidelines and the end-user's perception of credibility. This highlights the value of existing guidelines [12] as proxies of credibility, thus endorsing these guidelines as a roadmap in the complex and subjective task of credibility tagging.

We also demonstrated that this relation is topic-dependent, as users tend to deviate more from the reference ground truth when they are less familiar with the topic. Results also confirmed previous research that suggests that people tend to overestimate credibility and, often, struggle to identify sites labelled as low-quality. We also studied the influence of other variables, such as the time of completion or the educational background, and some interesting conclusions arose. Regardless of the educational background, people have difficulties judging the credibility. This even happens with individuals that have a strong educational background (for example, graduated students who have often been trained in skills such as critical thinking). It also appears that "overthinking" and spending too much time to emit a judgement does not lead to better estimates of credibility. Text-free justifications were also inspected, confirming a direct relationship between certain elements from the credibility guidelines and user's perceptions.

6. Conclusions

In this paper, we have conducted a user study on people's perception to the credibility of online health information. First, we used a previous study in the field to produce reference values based on a series of guidelines. We found out a correlation between these values and the judgements collected in the study. However, some worrying facts were also found: people tend to overestimate the credibility of the sites (this can be specially damaging when health information seeking) and it seems that the educational background has not a direct effect in their perceptions. As future work, we want to differentiate between closely related concepts such as credibility (more subjective) or correctness (more factual) and study how they affect users judgements.

Acknowledgements

The authors thank: i) the financial support supplied by the Xunta de Galicia - Consellería de Cultura, Educación, Formación Profesional e Universidades (Centro de investigación de Galicia accreditation 2019-2022 ED431G-2019/04 and Reference Competitive Group accreditation 2021-2024, ED431C 2022/19) and the European Union (European Regional Development Fund - ERDF) and ii) the financial support supplied by project PID2022-137061OB-C22 (Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, Proyectos de Generación de Conocimiento; supported by the European Regional Development Fund).

The third author thanks the financial support obtained from project SUBV23/00002 (Ministerio de Consumo, Subdirección General de Regulación del Juego).

The authors also thank the funding of project PLEC2021-007662 (MCIN/AEI/10.13039/501100011033, Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, Plan de Recuperación, Transformación y Resiliencia, Unión Europea-Next Generation EU).

References

[1] S. Shepperd, D. Charnock, B. Gann, Helping patients access high quality health information, Bmj 319 (1999) 764–766.

[2] R. J. Cline, K. M. Haynes, Consumer health information seeking on the internet: the state of the art, Health education research 16 (2001) 671–692.

[3] S. Fox, Health topics: 80% of internet users look for health information online, Pew Internet & American Life Project, 2011.

[4] F. A. Pogacar, A. Ghenai, M. D. Smucker, C. L. Clarke, The positive and negative influence of search results on people's decisions about the efficacy of medical treatments, in: Proceedings of the ACM SIGIR Int. Conf. on Theory of Information Retrieval, 2017, pp. 209–216.

[5] G. Eysenbach, Infodemiology: The epidemiology of (mis) information, The American Journal of Medicine 113 (2002) 763–765.

[6] G. Eysenbach, J. Powell, O. Kuss, E.-R. Sa, Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review, Jama 287 (2002) 2691–2700.

[7] E. V. Bernstam, D. M. Shelton, M. Walji, F. Meric-Bernstam, Instruments to assess the quality of health information on the world wide web: what can our patients actually use?, International journal of medical informatics 74 (2005) 13–19.

[8] N. Vigdor, Man fatally poisons himself while self-medicating for coronavirus, doctor says, 2020. URL: https://www.nytimes.com/2020/03/24/us/chloroquine-poisoning-coronavirus.html, [accessed June 9, 2022].

[9] B. J. Fogg, Persuasive technologie301398, Communications of the ACM 42 (1999) 26–29.

[10] C. Hahnel, F. Goldhammer, U. Kröhne, J. Naumann, The role of reading skills in the evaluation of online information gathered from search engine environments, Computers in Human Behavior 78 (2018) 223–234.

[11] M. Kąkol, M. Jankowski-Lorek, K. Abramczuk, A. Wierzbicki, M. Catasta, On the subjectivity and bias of web content credibility evaluations, in: Proceedings of the 22nd international conference on world wide web, 2013, pp. 1131–1136.

[12] M. Fernández-Pichel, S. Meyer, M. Bink, A. Frummet, D. E. Losada, D. Elsweiler, Improving the reliability of health information credibility assessments, in: Proceedings of the 3rd Workshop on Reducing Online Misinformation through Credible Information Retrieval 2023 co-located with The 45th European Conference on Information Retrieval (ECIR 2023), 2023, pp. 43–50. URL: https://ceur-ws.org/Vol-3406/paper4_jot.pdf.

[13] B. J. Fogg, Prominence-interpretation theory: Explaining how people assess credibility online, in: CHI'03 extended abstracts on human factors in computing systems, 2003, pp. 722–723.

[14] J. Unkel, A. Haas, The effects of credibility cues on the selection of search engine results, Journal of the Association for Information Science and Technology 68 (2017) 1850–1862.

[15] S. M. Shariff, A review on credibility perception of online information, in: 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), IEEE, 2020, pp. 1–7.

[16] A. Bodaghi, K. A. Schmitt, P. Watine, B. C. Fung, A literature review on detecting, verifying,

and mitigating online misinformation, IEEE Transactions on Computational Social Systems (2023).

[17] A. L. Ginsca, A. Popescu, M. Lupu, et al., Credibility in information retrieval, Foundations and Trends in Information Retrieval 9 (2015) 355–475.

[18] D. H. McKnight, C. J. Kacmar, Factors and effects of information credibility, in: Proceedings of the ninth international conference on Electronic commerce, 2007, pp. 423–432.

[19] M. Viviani, G. Pasi, Credibility in social media: opinions, news, and health information—a survey, Wiley interdisciplinary reviews: Data mining and knowledge discovery 7 (2017) e1209.

[20] B. J. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, E. R. Tauber, How do users evaluate the credibility of web sites? a study with over 2,500 participants, in: Proceedings of the 2003 conference on Designing for user experiences, 2003, pp. 1–15.

[21] M. S. Eastin, Credibility assessments of online health information: The effects of source expertise and knowledge of content, Journal of Computer-Mediated Communication 6 (2001) JCMC643.

[22] C. N. Wathen, J. Burkell, Believe it or not: Factors influencing credibility on the web, Journal of the American society for information science and technology 53 (2002) 134–144.

[23] S. Sikdar, B. Kang, J. ODonovan, T. Höllerer, S. Adah, Understanding information credibility on twitter, in: 2013 International Conference on Social Computing, IEEE, 2013, pp. 19–24.

[24] S. K. Sikdar, B. Kang, J. O'Donovan, T. Hollerer, S. Adal, Cutting through the noise: Defining ground truth in information credibility on twitter, Human 2 (2013) 151–167.

[25] S. Zimmerman, A. Thorpe, C. Fox, U. Kruschwitz, Privacy nudging in search: Investigating potential impacts, in: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, 2019, pp. 283–287.

[26] N. S. Chok, Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data, Ph.D. thesis, University of Pittsburgh, 2010.

[27] J. Schwarz, M. Morris, Augmenting web pages and search results to support credibility assessment, in: Proceedings of the SIGCHI conference on human factors in computing systems, 2011, pp. 1245–1254.

[28] E. R. Carlson, Evaluating the credibility of sources: A missing link in the teaching of critical thinking, Teaching of Psychology 22 (1995) 39–41.

[29] M. Kattenbeck, D. Elsweiler, Understanding credibility judgements for web search snippets, Aslib Journal of Information Management (2019).