

 Open access • Journal Article • DOI:10.1002/ASI.22743

## Topic familiarity and information skills in online credibility evaluation — [Source link](#)

Teun Lucassen, Rienco Muilwijk, Matthijs Leendert Noordzij, Jan Maarten Schraagen

**Institutions:** University of Twente

**Published on:** 01 Feb 2013 - Journal of the Association for Information Science and Technology (Wiley)

**Topics:** Credibility, Skills management and Think aloud protocol

Related papers:

- [Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research](#)
- [Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context](#)
- [Factual accuracy and trust in information: The role of expertise](#)
- [The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information](#)
- [Believe it or not: Factors influencing credibility on the Web](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/topic-familiarity-and-information-skills-in-online-2js9f8b6hr>



## Topic Familiarity and Information Skills in Online Credibility Evaluation

Journal:	<i>Journal of the American Society for Information Science and Technology</i>
Manuscript ID:	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Lucassen, Teun; University of Twente, Cognitive Psychology & Ergonomics Muilwijk, Rienco; University of Twente, Cognitive Psychology & Ergonomics Noordzij, Matthijs; University of Twente, Cognitive Psychology & Ergonomics Schraagen, Jan Maarten; University of Twente, Cognitive Psychology & Ergonomics
Keywords:	trust < psychological aspects < (sociocultural aspects), skills < (human attributes) < (attributes), domain knowledge < knowledge < information content < (knowledge and information)

SCHOLARONE™  
Manuscripts

Gepubliceerd als:

Lucassen, T. & Schraagen, J.M.C. (2013). Topic. familiarity and information skills in online credibility evaluation. *Journal of the American society for information science and technology*, 64(2), 254-264.

<http://onlinelibrary.wiley.com/doi/10.1002/asi.22743/>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12 Topic Familiarity and Information Skills in Online Credibility Evaluation  
13  
14  
15

16 Teun Lucassen\*

17 *t.lucassen@utwente.nl*  
18

19  
20 Rienco Muilwijk

21 *riencomuilwijk@gmail.com*  
22

23  
24 Matthijs L. Noordzij

25 *m.l.noordzij@utwente.nl*  
26

27  
28 Jan Maarten Schraagen

29 *j.m.c.schraagen@utwente.nl*  
30  
31  
32  
33  
34  
35

36 Department of Cognitive Psychology and Ergonomics, University of Twente

37  
38 P.O. Box 217

39  
40 7500 AE Enschede

41  
42 The Netherlands.  
43  
44  
45

46  
47 \*Corresponding author. Tel. +31 53 489 3604, Fax. +31 53 489 4241.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

1  
2  
3  
4  
5 With the rise of user generated content, evaluating the credibility of information has become  
6  
7 increasingly important. It is already known that various user characteristics influence the way  
8  
9 credibility evaluation is performed. Domain experts on the topic at hand primarily focus on  
10  
11 semantic features of information (e.g., factual accuracy), whereas novices focus more on  
12  
13 surface features (e.g., length of a text). In this study, we further explore two key influences on  
14  
15 credibility evaluation, namely topic familiarity and information skills. Participants with  
16  
17 varying levels of information skills (i.e., high school students, undergraduates, and post-  
18  
19 graduates) evaluated Wikipedia articles of varying quality on familiar and unfamiliar topics  
20  
21 while thinking aloud. When familiar with the topic, participants indeed focused primarily on  
22  
23 semantic features of the information, whereas participants unfamiliar with the topic paid  
24  
25 more attention to surface features. The utilization of surface features increased with  
26  
27 information skills. Moreover, participants with better information skills calibrated their trust  
28  
29 to the quality of the information, whereas trust of participants with poorer information skills  
30  
31 did not. This study confirms the enabling character of domain expertise and information skills  
32  
33 in credibility evaluation as predicted by the updated 3S-model of credibility evaluation.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Introduction

Nowadays, we live in a world in which anyone can go online to attain all information imaginable, and more. While this presents the opportunity to expand our knowledge very quickly, the freedom of the Internet also has its downsides. One particular issue is that of the credibility of online information. In the pre-Internet era, evaluating credibility was relatively easy, as usually one specific individual could be held accountable (i.e., the author). Moreover, this task was mostly performed by trained professionals, such as newspaper or book editors. Nowadays, credibility evaluation is increasingly a responsibility of the end user, who often lacks the required skills (and often motivation) for the job (Flanagin & Metzger, 2007). The second wave of Internet technology (Web 2.0) has amplified this problem, because nowadays anyone can make information available to everyone.

The topic of credibility evaluation in online environments has attracted numerous researchers trying to explain the behavior of Internet users. The influence of many aspects, such as user characteristics (Metzger, 2007; Hilligoss & Rieh, 2008), information features (Yaari, Baruchson-Arbib, & Bar-Ilan, 2010; Lucassen & Schraagen, 2010), or other situational factors (Fogg, 2003; Kelton, Fleischmann, & Wallace, 2008) have been shown. One particular study demonstrated the impact of three distinctive user characteristics (namely domain expertise, information skills, and source experience) on the information features used in credibility evaluation (Lucassen & Schraagen, 2011). Initial validation for the proposed relationship between user characteristics and information features (in the 3S-model, explained below) was provided by means of an online quasi-experiment, which mainly focused on the influence of domain expertise.

In the current study, we attempt to gain more insight into the influence of various user characteristics on credibility evaluation. Two key user characteristics for active credibility evaluation (domain expertise and information skills) are manipulated and controlled

1  
2  
3 systematically in a think-aloud experiment, in order to better understand their relationship  
4  
5 with credibility evaluation and, ultimately, trust. Moreover, the experiment conducted can  
6  
7 show which particular strategies to evaluate credibility are applied by various users.  
8

9  
10 The remainder of this paper is structured as follows. We start by discussing and  
11  
12 defining the concepts of trust and credibility in online environments. After this, the 3S-model  
13  
14 (Lucassen & Schraagen, 2011) is discussed and revised, and related research is reviewed. Our  
15  
16 method to explore the role of domain expertise and information skills in credibility evaluation  
17  
18 is explained, followed by the results. The paper ends with a discussion on the results, and  
19  
20 their implications for academic research and practice.  
21

### 22 23 Trust and Credibility Evaluation

24  
25 Trust is an important concept in a world where we rely on interactions with other people  
26  
27 (e.g., financial transactions, information exchange). Constant monitoring of the other person  
28  
29 is often impossible, so we need to have trust in this person such that his or her actions are  
30  
31 beneficial (or at least not detrimental) to us (Mayer, Davis, & Schoorman, 1995). This  
32  
33 implies that a certain risk is taken each time we trust someone (Kelton et al., 2008).  
34

35  
36 In the case of trust in other people during information exchanges, trust implies that we  
37  
38 believe the information to be correct. This aspect of information is often called credibility. In  
39  
40 psychology, two key elements of credibility are defined, namely trustworthiness and  
41  
42 expertise (Fogg & Tseng, 1999). The first refers to whether someone *wants* to give correct  
43  
44 information (well-intentioned), whereas the latter refers to whether he is *able* to do so  
45  
46 (knowledgeable). Information usually travels from one person to another, so one-way  
47  
48 relationships between the reader and author can be expected rather than mutual relationships  
49  
50 (Kelton et al., 2008).  
51

52  
53 In some situations, people may want to reduce the risk they take when they trust  
54  
55 information (or similarly: trust that the other person gives us correct information). This may,  
56  
57  
58  
59  
60

1  
2  
3 for instance, be the case when the consequences of incorrect information are high (i.e., taking  
4 important, but wrong decisions based on the information). The risk of trusting can be reduced  
5 by performing a credibility evaluation. In such an evaluation, users search for cues in the  
6 information which they apply as indicators of high or low credibility. Which cues these are, is  
7 largely dependent on the mental model of trust of each individual user (Hilligoss & Rieh,  
8 2008). Different users may have very different conceptions of what is important for  
9 credibility. It has, for instance, been shown that references are a very important indicator for  
10 credibility for college students (Lucassen, Noordzij, & Schraagen, 2011). This can be  
11 explained by an academic bias towards references. Users without academic training are  
12 expected to attribute less value to this particular cue. They may pay more attention to other  
13 aspects, such as understandability or images.

14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The extent to which a credibility evaluation is performed, is dependent on the motivation and ability of the user (Metzger, 2007). Following dual-process theory (Chaiken, 1980), users only perform a credibility evaluation when they have a motivation to do so. According to Metzger (2007), motivation stems from the “consequentiality of receiving low-quality, unreliable, or inaccurate information” (p. 2087). Moreover, the level of processing (i.e., heuristic vs. systematic) is dependent on the ability (skills) of the user; a systematic evaluation is thus only performed when a user is motivated and able to evaluate.

It should be noted that the apparent dichotomous choice between heuristic and systematic processing is somewhat simplistic in the domain of trust. Credibility evaluation as a strategy to reduce the risk of trusting is always heuristic to a certain extent. This claim can be illustrated by considering the extreme case of systematic processing. If a user would consider all aspects of credibility systematically, he or she would be certain of the credibility of the information. This means that the concept of trust is eliminated. Hence, absolute

1  
2  
3 systematic processing is not possible in credibility evaluation, and therefore always remains  
4  
5 heuristic to a certain extent.  
6

### 7 The 3S-model

8  
9  
10 In order to better understand how people form their judgments on the credibility of  
11  
12 information, the 3S-model was introduced by Lucassen and Schraagen (2011). In this model,  
13  
14 three strategies of credibility evaluation are proposed.  
15

16 The first strategy is to consider semantic features of the information, such as its  
17  
18 accuracy or neutrality. This requires a certain level of domain expertise from the user, as the  
19  
20 presented information is compared with his or her own knowledge on the topic. Following  
21  
22 this strategy, the most salient aspect of credibility is attended, namely factual accuracy.  
23  
24

25 When domain expertise is low or non-existent, it is nearly impossible to follow the  
26  
27 semantic strategy. Users can work around this deficit by considering surface features of the  
28  
29 information. These features pertain to the way the information is presented and include for  
30  
31 instance the length of the text or the number of references. However, this requires different  
32  
33 skills from the user, namely generic information skills. Such skills include knowledge of the  
34  
35 user on how particular features are related to the concept of credibility (e.g., the presence of  
36  
37 references suggests well-researched information).  
38  
39

40 A third strategy is to consider previous experiences with a particular source as an  
41  
42 indicator for credibility. As opposed to the first and second strategy, this is a passive strategy,  
43  
44 as the actual information itself is not considered, but only the source where it came from.  
45  
46

47 When evaluating credibility, users follow one or more of these strategies, leading to a  
48  
49 trust judgment. Following the outcome of the trust judgment, source experience can be  
50  
51 adjusted accordingly.  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 In this study, we propose an updated version of the 3S-model (see Figure 1). The key  
4 point of the model remains unchanged, but two adjustments have been made to enhance the  
5 clarity of the model.  
6  
7  
8

9  
10 <<Insert Figure 1 here>>

11 First, the originally introduced term “Trust Judgment” proved to be ambiguous. It can  
12 be interpreted as the *process* of judging trust (i.e., credibility evaluation, considering the  
13 various related features) or as the *outcome* of this process (i.e., trust in the information). Since  
14 the 3S-model is more of an information model than a process model, we decided to rename  
15 “Trust Judgment” to “Trust” in the revised version.  
16  
17  
18  
19  
20  
21  
22

23 Moreover, in the original model, the connecting arrows only indicated that a  
24 relationship existed between certain user characteristics and information features. By  
25 switching the position of information characteristics and user characteristics in the original  
26 model, we are better able to specify the nature of these relationships. The three user  
27 characteristics play an enabling role in the selection of information features; for instance,  
28 possessing domain expertise on a topic *enables* the utilization of semantic features. The same  
29 goes for information skills and source experience: possessing information skills *enables* the  
30 utilization of surface features and possessing source experience *enables* the utilization of  
31 source features in credibility evaluation. Considering the enabling character of the user  
32 characteristics it naturally follows that only those information characteristics which are  
33 enabled have an *influence* on trust. Consider, for example, a college student with no  
34 particular knowledge on the topic at hand. We can expect a reasonable level of information  
35 skills, which he or she can bring to bear when evaluating credibility. However, his or her  
36 domain expertise on the topic is low or even non-existent. This means that when the  
37 information looks credible on the surface level (e.g., lengthy, numerous references and  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 images), factual errors in the information are likely to go undetected by this student. This  
4  
5 may result in (unjustifiably) high trust in the information.  
6

7  
8 It is perhaps tempting to interpret the semantic strategy as a systematic approach and  
9  
10 the surface strategy as a heuristic approach. However, this is not necessarily the case. For  
11  
12 example, recognizing stated facts as you have learned them before can be considered  
13  
14 heuristic processing (Klein, Calderwood, & Clinton-Cirocco, 1986), but classifies as a  
15  
16 semantic strategy. Similarly, considering the quality of each of the references is clearly  
17  
18 systematic processing (Lucassen, Noordzij, & Schraagen, 2011), but classifies as a surface  
19  
20 strategy. Hence, both the semantic and surface strategy can be performed systematically and  
21  
22 heuristically. An exception has to be made for the source strategy: this is largely heuristic, as  
23  
24 the actual information is not considered at all, but only where it came from. Earlier  
25  
26 experiences with this source are in this case a predictor of the credibility of the current  
27  
28 information.  
29  
30

### 31 32 Domain expertise

33  
34 The important role of domain expertise in credibility evaluation has been shown on numerous  
35  
36 occasions. It has been demonstrated that having knowledge on the topic at hand leads to more  
37  
38 trust (Self, 1996; Chesney, 2006) in the information. However, we argue that this is not  
39  
40 necessarily the case. Following Lucassen and Schraagen (2011), trust will only be high if the  
41  
42 features incorporated in the evaluation indicate high credibility. Hence, domain experts will  
43  
44 only have high trust in information that is credible at the semantic level (e.g., factually  
45  
46 accurate). This claim is supported by Chesney (2006), who argued that Wikipedia is credible,  
47  
48 since domain experts trusted the information more than novices.  
49  
50

51  
52 Domain experts are expected to evaluate credibility better than novices. Kelton,  
53  
54 Fleischmann, & Wallace (2008) argued that their trust is better calibrated to the actual  
55  
56  
57  
58  
59  
60

1  
2  
3 credibility of information, as their general propensity to trust has less influence on their  
4  
5 judgments than novices' propensity to trust.  
6

#### 7 8 Information skills

9  
10 Information skills, or information literacy, can be defined as “the skills required to identify  
11  
12 information sources, access information, evaluate it, and use it effectively, efficiently, and  
13  
14 ethically” (Julien & Barker, 2008, p. 12). Brand-Gruwel, Wopereis, and Vermetten (2005)  
15  
16 defined five stages for information problem solving, namely defining, selecting, searching,  
17  
18 processing, and organizing. Information skills influence how well information users perform  
19  
20 each of these tasks. It has for instance been shown that experts (PhD students) spend more  
21  
22 time defining the problem than novices (undergraduates) before moving on to subsequent  
23  
24 stages (Brand-Gruwel, Wopereis, & Vermetten, 2005). Moreover, information experts more  
25  
26 often activate their prior knowledge, elaborate on the content, and regulate their process.  
27  
28 These differences result in a better task performance of experts.  
29  
30

31  
32 As follows from the definition given above, information skills do not only relate to the  
33  
34 ability to evaluate credibility. However, it is an important sub skill, which relates to multiple  
35  
36 stages of the information problem solving process. For instance, source credibility is of  
37  
38 importance when selecting appropriate information sources, but when processing information  
39  
40 that was found, credibility at the surface and semantic level can be evaluated.  
41  
42

43  
44 According to Alexander & Tate (1999), users who evaluate information should focus on  
45  
46 five criteria, namely accuracy, authority, objectivity, currency, and coverage. Walraven,  
47  
48 Brand-Gruwel, and Boshuizen (2009) showed that students often know more of such criteria  
49  
50 than they actually apply when searching for information, indicating that they lack a critical  
51  
52 disposition to information from online sources. Moreover, Julien & Barker (2009)  
53  
54 demonstrated large gaps in the information skills of students (e.g., lack of knowledge on how  
55  
56  
57  
58  
59  
60

1  
2  
3 search engines work), arguing that education in information skills should be improved,  
4  
5 especially at high-schools.  
6

7  
8 Following the various studies on information skills of students at various educational  
9  
10 levels (i.e., high-school, undergraduate, post-graduate) it can be concluded that information  
11  
12 skills improve with education. High-school students have very limited skills to evaluate  
13  
14 information, which means that they largely depend on the credibility of a source (e.g.,  
15  
16 university websites are credible) rather than evaluating the content itself (Julien & Barker,  
17  
18 2009). Undergraduate students are better able evaluate information, largely by applying  
19  
20 various heuristics (on the source of information *and* the content itself; Hilligoss & Rieh,  
21  
22 2008). Post-graduate (PhD) students can be considered experts in information problem  
23  
24 solving (at least in comparison with undergraduate students), as they focus much more on  
25  
26 various aspects (e.g., quality, relevance, reliability) of the actual content of information  
27  
28 (Brand-Gruwel, Wopereis, & Vermetten, 2005).  
29  
30

### 31 32 Wikipedia

33  
34 One particularly interesting source on the Web to study the evaluation behavior of lay  
35  
36 Internet users is Wikipedia. This vast online encyclopedia thrives on user contribution:  
37  
38 everyone can make changes or additions to the available articles, or create new ones.  
39  
40 Intuitively, this seems like a bad idea, as this open-editing model is bound to attract vandals  
41  
42 (Viégas, Wattenberg, & Kushal, 2004) and other individuals with bad intentions (consider the  
43  
44 “trustworthiness” aspect of credibility). Moreover, how can we know that contributors have  
45  
46 the appropriate credentials (consider the “expertise” aspect of credibility, Fogg & Tseng,  
47  
48 1999) to add information?  
49  
50

51  
52 Still, history has proven many of the early critics wrong. Wikipedia has been shown to  
53  
54 be a reliable source of information on numerous occasions (e.g., Giles, 2005; Chesney, 2006;  
55  
56 Rajagopalan et al., 2010). This has been attributed to the collaborative manner in which the  
57  
58  
59  
60

1  
2  
3 articles are written (Wilkinson & Huberman, 2007). However, because of this very same  
4  
5 principle, Wikipedia users can never be *entirely* certain of the credibility of the articles. This  
6  
7 imposes the need for trust and thus also the need to evaluate credibility before using the  
8  
9 information.  
10

11 The notion that traditional heuristics no longer apply on the Web is also true in the  
12  
13 domain of Wikipedia, perhaps to an even larger extent due to its open-editing model  
14  
15 (Magnus, 2009). This of course has implications for credibility evaluation on this source.  
16  
17 Lucassen and Schraagen (2011) showed that when factual errors are present in Wikipedia  
18  
19 articles, trust is only influenced when the user is a domain expert, and even then only to a  
20  
21 limited extent. Novices were not influenced at all by the factual errors. In an earlier study  
22  
23 (Lucassen & Schraagen, 2010), undergraduate students worked around their lack of domain  
24  
25 expertise by applying their information skills. By doing so, they were able to distinguish high  
26  
27 and low quality information on Wikipedia (Lucassen & Schraagen, 2010).  
28  
29  
30

### 31 Hypotheses

32  
33 In the original study on the 3S-model (Lucassen & Schraagen, 2011), initial validation  
34  
35 through an online quasi-experiment was provided. This was done by manipulating a key  
36  
37 semantic feature (factual accuracy) and showing that users with some domain expertise were  
38  
39 influenced by errors, whereas complete novices were not. However, this approach has certain  
40  
41 limitations, which we address in this study.  
42  
43  
44

45 The participants in the preceding study were recruited on the basis of their domain  
46  
47 expertise (high vs. low) in the field of automotive engineering. This means that their level of  
48  
49 information skills was not controlled for. The first goal in this study is to further explore the  
50  
51 influence of both domain expertise and information skills on the features used in credibility  
52  
53 evaluation. Domain expertise was manipulated in a more rigorous fashion by presenting the  
54  
55 participants information on topics on which they indicated themselves to have high or low  
56  
57  
58  
59  
60

1  
2  
3 prior knowledge. In line with the results of Lucassen and Schraagen (2011), we formulate the  
4  
5 following hypothesis:  
6

7 *Hypothesis 1: Users with more domain expertise utilize more semantic features in*  
8  
9 *credibility evaluation than users with less domain expertise.*  
10

11 Information skills were controlled by selecting three different groups that are known to  
12  
13 differ in their level of information skills, namely high school students, undergraduates, and  
14  
15 post-graduates (Brand-Gruwel, Wopereis, & Vermetten, 2005; Julien & Barker, 2009).  
16  
17 Naturally, these groups will also differ on other dimensions than information skills only (e.g.,  
18  
19 age), which may introduce confounding variables in our experiments. However, these three  
20  
21 groups are all regular information seekers in comparable contexts (i.e., education), which aids  
22  
23 to the external validity of this research. In contrast, we believe that a more controlled but  
24  
25 isolated approach (e.g., training one half of a coherent group with low information skills)  
26  
27 would harm the validity of this study.  
28  
29  
30

31  
32 Following the preceding discussion on information skills, we expect that users with  
33  
34 better information skills (e.g., post-graduate students) can bring to bear more strategies to  
35  
36 consider various features of the information, rather than only focusing on the factual accuracy  
37  
38 (i.e., semantics features) of information. By doing so, they can work around their lack of prior  
39  
40 knowledge by considering surface features. In contrast, users with poorer information skills  
41  
42 will incorporate fewer surface features, as they are unfamiliar with such indicators of  
43  
44 credibility. Instead, they will mainly consider the semantics of the information, also when  
45  
46 they have limited prior knowledge on the topic at hand. This leads to the second hypothesis:  
47  
48

49 *Hypothesis 2: Users with better information skills utilize more surface features in*  
50  
51 *credibility evaluation than users with poorer information skills.*  
52

53  
54 As noted, in the original experiment a semantic feature was manipulated. This means  
55  
56 that differences in trust between users were mostly caused by differences in their domain  
57  
58  
59  
60

1  
2  
3 expertise. The application of surface features in credibility evaluation was also shown in the  
4  
5 experiment. However, the articles were kept unchanged on the surface level, which means  
6  
7 that although information skills were applied, this had no influence on trust.  
8

9  
10 In this study, we manipulate the quality of the presented information following the  
11 classification of the Wikipedia Editorial Team (“Wikipedia:Version 1.0 Editorial Team”,  
12 2012). Their goal is to assess all Wikipedia articles on how close they are to a “distribution-  
13 quality article on a particular topic.” While this implies that the articles should be factually  
14 accurate, we expect that the difference between high-quality articles and low-quality articles  
15 is best visible on the surface level, for instance by the number of references, its length, and  
16 the presence of images. These characteristics are explicitly noted in the grading scheme<sup>2</sup> of  
17 the Wikipedia Editorial Team.  
18  
19  
20  
21  
22  
23  
24  
25  
26

27 When the quality levels of the Wikipedia Editorial Team are indeed best visible at the  
28 surface level, this means that a certain level of information skills is needed in order to be  
29 influenced by the quality. We expect that users with poorer information skills do not focus on  
30 the features which reflect the quality level, and are thus not influenced by them. This leads to  
31 the following hypotheses:  
32  
33  
34  
35  
36  
37

38 *Hypothesis 3: Trust of users with better information skills is influenced by the quality of*  
39 *the information.*  
40  
41

42 *Hypothesis 4: Trust of users with poorer information skills is not influenced by the*  
43 *quality of the information.*  
44  
45  
46

47 In contrast, we do not expect that domain expertise has much influence on trust in high-  
48 quality or low-quality information. Articles with lower quality are generally also not expected  
49 to feature major errors; they are mainly much shorter and unfinished compared to higher-  
50 quality articles.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Method

### *Participants*

A total of 40 participants took part in the experiment. Three participant groups were created, namely high school students, undergraduate students, and post-graduate (PhD) students.

Table 1 shows the key characteristics of each of the participant groups.

<<Insert Table 1 here>>

The high-school students were in their third year out of six years of pre-academic education (i.e., preparing them for a subsequent university or college education). They received monetary compensation for their participation. Their experience with Wikipedia ranged from 2-5 years with an average of 4. Only three of the high school students mentioned the open-editing model behind Wikipedia when asked to explain the basics of this website. One high school student had experience in editing articles on Wikipedia himself.

The undergraduates were all following education in the domain of behavioral sciences. They received course credits for participating. Their experience with Wikipedia ranged from 3-8 years with an average of 5. All undergraduates students were able to explain the basics of Wikipedia in their own words. None of them had contributed to Wikipedia before.

The post-graduates were from various disciplines, such as behavioral sciences, physics, and management sciences. Their experience with Wikipedia ranged from 4-10 years with an average of 7. All post-graduates described the online encyclopedia as an open source that anyone can edit. Three post-graduates had experience in editing articles on Wikipedia.

All participants in the three groups were proficient in the Dutch language and able to effortlessly express their thoughts in this language. Therefore Dutch was chosen for the think aloud method (Ericsson & Simon, 1984). The articles used in the experiment were obtained from the English Wikipedia for the undergraduate and post-graduate students. No major language barriers were reported after the experiment. The participating high school students



1  
2  
3 were not sufficiently proficient in the English language to be able to fully comprehend  
4  
5 information in this language, therefore the Dutch Wikipedia was used to select articles for  
6  
7 this participant group.  
8

### 9 10 *Task*

11 The participants performed the Wikipedia Screening Task (Lucassen & Schraagen, 2010). In  
12  
13 this task, a Wikipedia article is displayed in a web browser. The participants are asked to  
14  
15 evaluate its credibility, without imposing a particular method on them to do so. This means  
16  
17 that they are free (and encouraged) to employ their own approach for this task. While doing  
18  
19 this, the participants were asked to think aloud following standard think-aloud instructions  
20  
21 (Ericsson & Simon, 1984). The participants were not allowed to navigate away from the  
22  
23 article during the task. No time limit was set.  
24  
25  
26

### 27 28 *Design*

29 A 3 (student group) × 2 (familiarity) × 2 (article quality) mixed design was applied for the  
30  
31 experiment. Student group (high school, undergraduate, post-graduate) was a between  
32  
33 subjects factor, whereas familiarity (familiar/unfamiliar) and article quality (high/low) were  
34  
35 both within subject factors. Each participant evaluated ten articles in total.  
36  
37

38 Familiarity was manipulated by selecting articles to be used in the experiment for each  
39  
40 participant individually. This was done on the basis of a telephone interview, conducted a few  
41  
42 days before the actual experiment. In this interview, the participants were asked for their  
43  
44 personal interests and disinterests. Half of the articles were selected to be on familiar topics;  
45  
46 the other half were on unfamiliar topics. Each article was only used once throughout the  
47  
48 whole experiment. Familiarity alternated between trials, starting with a familiar topic.  
49  
50

51 Article quality was manipulated following the classification of the Wikipedia Editorial  
52  
53 Team (“Wikipedia:Version 1.0 Editorial Team”, 2012). Manual assessments of the quality  
54  
55 are available for most of the articles on Wikipedia, resulting in a categorization into seven  
56  
57  
58  
59  
60

1  
2  
3 classes (see Table 2). However, A-class articles are largely underrepresented on Wikipedia,  
4  
5 which makes it virtually impossible to find articles on specific topics in this class. Therefore,  
6  
7 it was excluded from the experiment, leaving a total of six classes. Articles in the highest  
8  
9 three classes were considered high quality (Featured articles, Good articles, and B-class  
10  
11 articles); articles in the three lowest classes were considered low quality (C-class articles,  
12  
13 Start articles, and Stub articles). Article quality was randomized between trials.  
14  
15

16 <<Insert Table 2 here>>  
17

18  
19 Unfortunately, no classification of quality is available on the Dutch Wikipedia (apart  
20  
21 from the Dutch equivalent of the “Featured articles”, but these are very few). Instead, we  
22  
23 applied the clear criteria of the Wikipedia Editorial Team to articles from the Dutch  
24  
25 Wikipedia ourselves to distinguish high and low quality. To ensure the validity of this  
26  
27 manipulation, inter-rater reliability was calculated after double-rating the selected articles.  
28  
29 The result was a Cohen’s Kappa of .89 (Landis & Koch, 1977), which indicates a near perfect  
30  
31 agreement.  
32  
33

34  
35 The articles used in the experiment were presented exactly as they appeared on  
36  
37 Wikipedia, with the exception of the removal of cues specific for Wikipedia, indicating  
38  
39 diminished credibility (e.g., *[citation needed]* indications) or high credibility (e.g., bronze  
40  
41 stars in Featured articles). The removal of such indicators ensured that the participants could  
42  
43 only utilize cues from the information itself in their credibility evaluations rather than cues  
44  
45 only valid in the domain of Wikipedia.  
46

#### 47 *Procedure*

48  
49 Upon arrival, participants were provided a brief explanation of the experiment and asked to  
50  
51 sign an informed consent. As all participating high school students were under 18 years of  
52  
53 age, we also asked their parents or legal guardians to sign an informed consent in advance.  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 After signing, the participants had to fill in a short questionnaire regarding standard  
4 demographic features, and their familiarity and experience with Wikipedia (on 7-point Likert  
5 scales) along with their quotidian usage. They were also asked to provide a short explanation  
6 of what Wikipedia is and how it works.  
7  
8

9  
10  
11 Following this questionnaire, the participants were instructed on the Wikipedia  
12 Screening Task and the course of the experiment. The participants practiced the Wikipedia  
13 Screening Task and the think-aloud task during two practice trials. The articles used in these  
14 trials were “Barcelona” and “Titanic” for the high school students, and “Flat earth” and  
15 “Ethnography” for the undergraduates and post-graduates. Task performance was considered  
16 sufficient for all participants after two practice trials.  
17  
18

19  
20  
21 When the participants finished a trial, they indicated this to the experimenter verbally,  
22 who then handed them a questionnaire, on which perceived credibility and familiarity were  
23 measured on 7-point Likert scales. This was repeated 10 times for each participant, resulting  
24 in a total duration of approximately 90 minutes.  
25  
26

### 27 *Data analyses*

28 All sessions were audio-recorded and transcribed afterwards. In a protocol analysis, all  
29 utterances regarding credibility were marked and categorized. Each utterance was coded on  
30 the following aspects:  
31  
32

- 33 - The *component* of the article to which the utterance referred (Introduction, Text,  
34 Table of contents, Images, References, and Other).  
35
- 36 - The *strategy* applied by the participant (Semantic, Surface). Note that the Source  
37 strategy of the 3S-model (Lucassen & Schraagen, 2011) was not used here, as the  
38 source remained constant throughout the experiment.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 - Which *feature* of the component was mentioned by the participant (e.g., *number* of  
4 references, *quality* of the pictures). Ad hoc categories were created for the features  
5 mentioned.  
6  
7  
8

9  
10 The protocol of one student in each group was double-coded by two experimenters.  
11 Based on this overlap, the inter-rater reliability was calculated. A Cohen's Kappa of .87  
12 (Landis & Koch, 1977) indicated a near-perfect agreement.  
13  
14

15  
16 During the protocol analysis, it became apparent that the participants greatly differed in  
17 number of utterances. In order to ensure that expressive participants did not have a larger  
18 influence on the outcome of each group than the others, the number of utterances of each  
19 participant (*i*) in each category (*n*) was corrected. This was done by multiplying each number  
20 by the correction factor derived in the following formula:  
21  
22  
23  
24  
25

$$26 \text{ correction factor} = \frac{\text{number of remarks}_n/n}{\text{number of remarks}_i}$$

27  
28  
29  
30  
31 After this correction, the number of utterances was averaged over each group to create a  
32 coding scheme for each group.  
33  
34

35 Only non-parametric tests were performed on all data gathered on Likert-scales, as they  
36 are assumed to be measuring at the ordinal rather than the nominal level (Jamieson, 2004).  
37  
38

## 39 Results

### 40 *Familiarity manipulation check*

41  
42 The questionnaires after each article indicated that the manipulation of familiarity was  
43 successful. On a 1-7 familiarity scale, familiar topics were rated higher ( $M = 5.20$ ,  $SD = .92$ )  
44 than unfamiliar topics ( $M = 1.94$ ,  $SD = .87$ ),  $Z = 5.48$ ,  $p < .001$ . A more detailed analysis  
45 showed that this was the case for all participating groups (high school students,  
46 undergraduates, and post-graduates).  
47  
48  
49  
50  
51  
52

### 53 *Credibility evaluation*

1  
2  
3 Table 3 shows the number of remarks indicating the application of a semantic or surface  
4 strategy in the credibility evaluations of our participants. Typical examples of remarks  
5 categorized as a semantic strategy were “Yes, I know this is true, because the things I know  
6 about it are in line with the text.” and “I know this already, because I traveled by airplane last  
7 year.” Remarks such as “There are images everywhere, which seems trustworthy to me.” and  
8 “Every claim is referenced, that’s a good thing.” were typical for the surface strategy.  
9  
10  
11  
12  
13  
14  
15

16 <<Insert Table 3 here>>  
17

18 Participants evaluating articles on familiar topics used more semantic cues than when  
19 evaluating unfamiliar topics,  $\chi^2(1, N = 931) = 24.40, p < .001$ . This was the case for all  
20 participant groups (high-school students:  $\chi^2(1, N = 661) = 11.05, p < .01$ ; undergraduates:  
21  $\chi^2(1, N = 1122) = 41.74, p < .001$ ; post-graduates:  $\chi^2(1, N = 1010) = 29.43, p < .001$ ).  
22  
23  
24  
25  
26  
27

28 Moreover, the participant groups differed in their application of the semantic and  
29 surface strategy regardless of familiarity,  $\chi^2(2, N = 931) = 111.35, p < .001$ . This effect was  
30 caused by the high-school students using less surface features than the other groups. No  
31 difference was found between undergraduates and post-graduates.  
32  
33  
34  
35

36 Table 4 shows the number of remarks concerning the various components of the  
37 articles.  
38  
39

40 <<Insert Table 4 here>>  
41  
42

43 Participants in the various groups considered the different components of the article to  
44 varying degrees,  $\chi^2(12, N = 2793) = 435.85, p < .001$ . Post-hoc analyses showed that this was  
45 caused by post-graduates having fewer remarks on images and more on the introduction than  
46 the other groups. The number of remarks on references differed between all groups,  
47 increasing with education level. Finally, high school students mentioned the component ‘text’  
48 more and ‘table of contents’ less than the other groups.  
49  
50  
51  
52  
53  
54  
55

56 Table 5 shows the key features used by each group.  
57  
58  
59  
60

1  
2  
3 <<Insert Table 5 here>>  
4

5 As can be seen in Table 4, high school students had a smaller arsenal of strategies to  
6 evaluate credibility than the other groups. Some evident strategies are mentioned by all users  
7 (e.g., factual accuracy), but other strategies such as considering the references and the  
8 objectivity of the information were only mentioned by undergraduates and post-graduates.  
9

### 10 *Trust*

11 Table 6 shows trust in the information of all participants in all conditions.  
12

13 <<Insert Table 6 here>>  
14

15 No effect of student group on trust was found,  $\chi^2(2, N = 40) = .21, p = .90$ . This  
16 indicates that high school students, undergraduates, and post-graduates all have similar trust  
17 in Wikipedia.  
18

19 Moreover, no effect of familiarity on trust was found,  $Z = 1.68, p = .09$ . This was also  
20 the case for each individual student group (high school:  $Z = 1.09, p = .28$ ; undergraduates:  $Z$   
21  $= 1.30, p = .19$ ; post-graduates:  $Z = .66, p = .51$ ).  
22

23 Quality had a significant effect on trust: high-quality articles were trusted more than  
24 low-quality articles ( $Z = 3.62, p < .01$ ). However, a more detailed analysis showed that this  
25 was only the case for undergraduates ( $Z = 2.67, p < .01$ ) and post-graduates ( $Z = 2.84, p <$   
26  $.01$ ), but not for high school students ( $Z = 1.37, p = .17$ ).  
27

### 28 Discussion

29 In this study the influence of domain expertise and information skills on credibility  
30 evaluation and trust was examined. The results supported the updated 3S-model. It was found  
31 that users with domain expertise tended to focus more on semantic features than users  
32 without domain expertise. Moreover, surface features were used more by users with better  
33 information skills. Information quality was manipulated following the classification of the  
34 Wikipedia Editorial Team. We hypothesized that this would be mainly visible at the surface  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 level of the articles. Our experiment confirmed that indeed only trust of users with better  
4 information skills was influenced by the quality manipulation (i.e., only undergraduates and  
5 post-graduates, not high-school students). As expected, domain expertise had no influence on  
6 trust in high or low quality articles, as low-quality articles are also expected to be free of  
7 (large) factual errors.  
8  
9

10  
11  
12  
13  
14 The main contribution of this study is that the enabling character of domain expertise  
15 and information skills has been demonstrated, together with the influence of the  
16 corresponding information features on trust. In the original study on the 3S-model (Lucassen  
17 & Schraagen, 2011), it was already shown that trust of domain experts was influenced when a  
18 semantic feature (enabled for domain experts) indicated low credibility. Now, we also  
19 demonstrated that when surface features indicate lower credibility, this only has an influence  
20 on users with sufficient information skills. Hence, only undergraduates and post-graduates  
21 were influenced, whereas high school students were not.  
22  
23  
24  
25  
26  
27  
28  
29  
30

31  
32 This observation is very much in line with prominence-interpretation theory (Fogg,  
33 2003), which states that each cue in a piece of information has a certain prominence to a  
34 certain user. Only when a cue is prominent, the user can give an interpretation to this cue  
35 (i.e., consequences for credibility), and have an influence on trust. The key addition of the  
36 3S-model in comparison to prominence-interpretation theory is that we attribute specific user  
37 characteristics to specific information features.  
38  
39  
40  
41  
42  
43  
44

45 In this experiment, we showed that people with knowledge on the topic evaluate the  
46 credibility of information differently than people without such knowledge. The key difference  
47 is the utilization of semantic features, such as the accuracy of information. Novices on the  
48 topic at hand are not able to compare presented information with their pre-existing  
49 knowledge, which leads them to the consideration of other, surface features. Interestingly,  
50 one does not have to be an absolute domain expert to apply the “semantic strategy” of  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 credibility evaluation. Whereas in Lucassen and Schraagen (2011), domain experts were self-  
4  
5 selected from various Internet forums on automotive technology, in this experiment  
6  
7 familiarity was merely manipulated by asking the participants for their topics of interest. This  
8  
9 does not ensure a substantial level of expertise at all. Still, the influence of familiarity at this  
10  
11 level on credibility evaluation was made quite clear. Participants familiar with the topic at  
12  
13 hand used nearly twice as many semantic features in their credibility evaluations than  
14  
15 participants unfamiliar with the topic (except for high-school students, to be discussed later).  
16  
17

18  
19 Interestingly, when users encounter information on a familiar topic, they do not shift to  
20  
21 semantic features completely. Instead, they apply a combination of surface and semantic  
22  
23 strategies to evaluate credibility. This means that familiar users (with sufficient information  
24  
25 skills) are best equipped to evaluate credibility in a meaningful manner. However, this  
26  
27 experiment merely indicated the capabilities of various users to evaluate, which may differ  
28  
29 from their actual behavior in real-life. As predicted by Metzger (2007), the motivation of  
30  
31 users primarily determines to which extent credibility is evaluated. This experiment showed  
32  
33 what they are capable of when they are motivated.  
34  
35

36  
37 As stated earlier, the shift towards semantic features when evaluating familiar  
38  
39 information was much less distinctive for high-school students than for the other two groups.  
40  
41 We do not attribute this to an unexpected high level of expertise in unfamiliar topics (the  
42  
43 familiarity manipulation proved successful), but to a low level of information skills. We  
44  
45 argued before that the most salient strategy for credibility evaluation is to consider the factual  
46  
47 accuracy. This is also what the high-school students did. However, when evaluating  
48  
49 information on unfamiliar topics, this strategy is quite unsuccessful. We also observed this in  
50  
51 many participants remarking that they felt unable to evaluate the article at hand, as they did  
52  
53 not know anything about the topic (e.g., “If you don’t know anything about it, it is tempting  
54  
55 to believe the information is correct.” or “It doesn’t ring a bell, it could be true.”).  
56  
57  
58  
59  
60



1  
2  
3 Participants with better information skills worked around this deficit by considering various  
4  
5 alternative (surface) features. However, high-school students were not able to do so, due to  
6  
7 their limited information skills, which meant that a large portion (about half) of the remarks  
8  
9 still included semantic features (albeit unsuccessful). This was reflected in the trust of high-  
10  
11 school students in the information; as opposed to the other groups, no difference in trust was  
12  
13 observed between high-quality and low-quality information. The key surface feature that  
14  
15 high-school students did not consider at all as opposed to the other groups, was references.  
16  
17 High-school students were not at all aware of the importance of references, whereas a large  
18  
19 part of the remarks of undergraduates and post-graduates considered this feature (about 30%).  
20  
21 This replicates the finding of Lucassen, Noordzij, and Schraagen (2011), who found that  
22  
23 undergraduates consider the references of information on various levels.  
24  
25  
26

27  
28 The limited information skills of high-school students could lead one to believe that  
29  
30 they could still perform a meaningful credibility evaluation on familiar topics, as they can  
31  
32 bring their knowledge on the topic at hand to bear. However, no influence of information  
33  
34 quality on trust was found for high-school students, regardless of their familiarity. It could be  
35  
36 argued that this is due to their limited domain expertise, also on familiar topics. However, a  
37  
38 more plausible explanation can be found in the nature of the quality manipulation at hand.  
39  
40 We decided to replicate normal quality fluctuations as can be observed on Wikipedia.  
41  
42 However, these fluctuations can primarily be found at the surface level, as the information is  
43  
44 expected to be generally factually accurate (Giles, 2005), also in low-quality articles. High-  
45  
46 school students utilize a lot less surface features in their evaluations, which means they did  
47  
48 not notice the differences in quality.  
49  
50

51  
52 The manner of manipulating information quality also explains why despite earlier  
53  
54 findings (Self, 1996; Eastin, 2001; Chesney, 2006), familiarity had no influence on trust. It  
55  
56 can be expected that most of the articles used in this experiment were factually accurate.  
57  
58  
59  
60

1  
2  
3 Thus, no negative influence of knowledge on the information is expected. However, given the  
4 overall high trust in the presented information (> 5 on a 1-7 scale), it is questionable whether  
5 familiarity would increase trust even more. A study in which the role of familiarity is  
6 examined in trust in information of more questionable credibility would be of interest to  
7 further explore this topic.  
8  
9  
10  
11  
12

13  
14 No effect of participant group on trust was found. This means that trust in Wikipedia is  
15 the same for high-school students, undergraduates, and post-graduates. This is remarkable,  
16 since knowledge on the open-editing system behind Wikipedia (largely absent in high-school  
17 students) could lead to less trust. On the other hand, accumulating positive experiences with  
18 Wikipedia may increase trust in this source (Lucassen & Schraagen, 2011). This would  
19 indicate that the strategy of considering the source of information was also applied, but  
20 implicitly, as no participants mentioned this in the think aloud protocols (Taraborelli, 2008).  
21  
22  
23  
24  
25  
26  
27  
28

### 29 *Limitations*

30  
31 A few limitations should be kept in mind regarding the interpretation of the results of this  
32 study.  
33  
34  
35

36 The three participating groups of students were selected on the basis of their expected  
37 level of information skills. We have shown that this had a direct influence on credibility  
38 evaluation. However, other factors will also inevitably vary among these groups (e.g., age).  
39 These factors may act as confounding variables. However, we would argue that an isolated  
40 approach of varying information skills (e.g., training half of a coherent group of participants  
41 with low information skills) does not add to the external validity of the study.  
42  
43  
44  
45  
46  
47  
48

49 In this study, Wikipedia served as an information source for our stimuli. This online  
50 encyclopedia is always a great case study, as information quality is generally very high  
51 (Giles, 2005), but changeable (e.g., Cross 2006; Dooley, 2010). However, certain  
52 characteristics of this source may limit the potential for generalization to other sources  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 (online and offline). An example of such a characteristic is the open editing model behind  
4  
5 Wikipedia; this mechanism may cause users to approach the information differently (e.g., in a  
6  
7 more skeptical manner). For future research, it is important to verify the validity of the  
8  
9 proposed 3S-model in different contexts, such as other websites, or offline sources (e.g.,  
10  
11 books, newspapers).  
12

13  
14 The think-aloud method is a great tool to gain insights in the task performance of  
15  
16 participants. It should however be noted that in this study, the participants were explicitly  
17  
18 asked to evaluate credibility, whereas normally this is a subset of a larger task set (i.e.,  
19  
20 finding and evaluating information). Therefore, the observed behavior in this experiment  
21  
22 should not be interpreted as the way users always perform credibility evaluation. The degree  
23  
24 to which credibility is actually being evaluated may vary heavily (Metzger, 2007). The  
25  
26 behavior we observed in this study can rather be seen as credibility evaluation under optimal  
27  
28 circumstances (in terms of motivation and ability). In real life, users may pick a few  
29  
30 strategies from the set we found, depending on the context of the information.  
31  
32

### 33 34 *Further Research*

35  
36 This study has shed more light on the role of user characteristics in online credibility  
37  
38 evaluation. Additional validation was found for the 3S-model (Lucassen & Schraagen, 2011).  
39  
40 However, this study primarily aimed at validating the semantic and surface components of  
41  
42 the model. Future studies should also focus on the third strategy, considering the source of  
43  
44 information.  
45  
46

### 47 48 Acknowledgments

49  
50 The authors would like to thank Andreas Bremer, Knut Jägersberg, and Koen Remmerswaal  
51  
52 for their efforts in gathering the data. Also, we are grateful to C.S.G. Dingstede, for allowing us  
53  
54 to perform the experiments with high-school students at their school.  
55

### 56 57 References

1  
2  
3 Alexander, J. E. & Tate, M. A. (1999). *Web Wisdom; How to Evaluate and Create*  
4  
5 *Information Quality on the Web*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1st edition.  
6

7 Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving  
8  
9 by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior*,  
10  
11 *21*, 487-508.  
12

13  
14 Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source  
15  
16 versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*,  
17  
18 752-766.  
19

20 Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, *11*.  
21

22  
23 Cross, T. (2006). Puppy smoothies: Improving the reliability of open, collaborative  
24  
25 wikis. *First Monday*, *11*.  
26

27  
28 Dooley, P. L. (2010). Wikipedia and the two-faced professoriate. In *Proceedings of the 6th*  
29  
30 *International Symposium on Wikis and Open Collaboration*, WikiSym '10, pages 1-2,  
31  
32 New York, NY, USA. ACM.  
33

34 Eastin, M. S. (2001). Credibility assessments of online health information: The effects of  
35  
36 source expertise and knowledge of content. *Journal of Computer-Mediated*  
37  
38 *Communication*, *6*.  
39

40 Ericsson, K. A. & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*.  
41  
42 Cambridge, Massachusetts, USA: The MIT Press.  
43

44  
45 Flanagin, A. J. & Metzger, M. J. (2007). The role of site features, user attributes, and  
46  
47 information verification behaviors on the perceived credibility of web-based  
48  
49 information. *New Media & Society*, *9*, 319-342.  
50

51  
52 Fogg, B. J. (2003). Prominence-interpretation theory: Explaining how people assess  
53  
54 credibility online. *CHI '03 extended abstracts on Human factors in computing systems*,  
55  
56 *CHI EA '03*, 722-723, New York, NY, USA. ACM.  
57  
58  
59  
60

- 1  
2  
3 Fogg, B. J. & Tseng, H. (1999). The elements of computer credibility. *Proceedings of the*  
4  
5 *SIGCHI conference on Human factors in computing systems: the CHI is the limit, CHI*  
6  
7 *'99*, 80-87, New York, NY, USA. ACM.
- 8  
9  
10 Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438, 900-901.
- 11  
12 Hilligoss, B. & Rieh, S. (2008). Developing a unifying framework of credibility assessment:  
13  
14 Construct, heuristics, and interaction in context. *Information Processing &*  
15  
16 *Management*, 44, 1467-1484.
- 17  
18 Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- 19  
20 Julien, H. & Barker, S. (2009). How high-school students find and evaluate scientific  
21  
22 information: A basis for information literacy skills development. *Library & Information*  
23  
24 *Science Research*, 31, 12-17.
- 25  
26  
27 Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital  
28  
29 information. *Journal of the American Society for Information Science and Technology*,  
30  
31 59, 363-374.
- 32  
33  
34 Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the  
35  
36 fire ground. *Proceedings of the Human Factors Society*, 576-580, Human Factors and  
37  
38 Ergonomics Society.
- 39  
40 Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical  
41  
42 data. *Biometrics*, 33, 159-174.
- 43  
44  
45 Lucassen, T., Noordzij, M. L., & Schraagen, J. M. (2011). Reference blindness: The  
46  
47 influence of references on trust in Wikipedia. In *ACM WebSci '11*, New York, NY,  
48  
49 USA. ACM.
- 50  
51  
52 Lucassen, T. & Schraagen, J. M. (2010). Trust in Wikipedia: How users trust information  
53  
54 from an unknown source. *Proceedings of the 4th workshop on Information credibility*,  
55  
56 *WICOW '10*, 19-26, New York, NY, USA. ACM.
- 57  
58  
59  
60

- 1  
2  
3 Lucassen, T. & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role  
4  
5 of expertise. *Journal of the American Society for Information Science and Technology*,  
6  
7 62, 1232-1242.  
8  
9  
10 Magnus, P. D. (2009). On trusting Wikipedia. *Episteme*, 6, 74-90.  
11  
12 Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of  
13  
14 organizational trust. *The Academy of Management Review*, 20(3):709-734.  
15  
16 Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online  
17  
18 information and recommendations for future research. *Journal of the American Society*  
19  
20 *for Information Science and Technology*, 58, 2078-2091.  
21  
22  
23 Rajagopalan, M. S., Khanna, V., Stott, M., Leiter, Y., Showalter, T. N., Dicker, A., &  
24  
25 Lawrence, Y. R. (2010). Accuracy of cancer information on the internet: A comparison  
26  
27 of a wiki with a professionally maintained database. *Journal of Clinical Oncology*  
28  
29 *(Meeting Abstracts)*, 28, 6058.  
30  
31  
32 Self, C.S. (1996). Credibility. In M. Salwen & D. Stacks (Eds.), *An integrated approach to*  
33  
34 *communication theory and research*. Mahwah, NJ: Erlbaum.  
35  
36  
37 Taraborelli, D. (2008). How the web is changing the way we trust. In K. Waelbers, A.  
38  
39 Briggie, & P. Breg (Eds.), *Current Issues in Computing and Philosophy* (pp. 194-  
40  
41 204). Amsterdam, The Netherlands: IOS Press.  
42  
43 Viégas, F., Wattenberg, M., & Kushal, D. (2004). Studying cooperation and conflict between  
44  
45 authors with history flow visualization. *Proceedings of the 2004 conference on Human*  
46  
47 *factors in computing systems*, New York. ACM.  
48  
49  
50 Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2009). How students evaluate information  
51  
52 and sources when searching the world wide web for information. *Computers &*  
53  
54 *Education*, 52, 234-246.  
55  
56  
57  
58  
59  
60

1  
2  
3 Wikipedia:Version 1.0 Editorial Team. (n.d.). In *Wikipedia*. Retrieved April 16, 2012, from

4  
5 [http://en.wikipedia.org/wiki/Wikipedia:Version\\_1.0\\_Editorial\\_Team](http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team)

6  
7  
8 Wilkinson, D. M. & Huberman, B. A. (2007). Cooperation and quality in Wikipedia.

9  
10 *Proceedings of the 2007 international symposium on Wikis, WikiSym '07*, 157-164, New  
11 York, NY, USA. ACM.

12  
13  
14 Yaari, E., Baruchson-Arbib, S., & Bar-Ilan, J. (2011). Information quality assessment of

15  
16 community generated content: A user study of Wikipedia. *Journal of Information*  
17  
18 *Science*, 37, 487-498.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure Captions

**Figure 1: Revised 3S-model of credibility evaluation.**

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



**Table 1: Key characteristics of all participant groups. Standard deviation for age is given in parentheses.**  
\* For the undergraduate group, the uncoded transcriptions (raw utterances of the participants of Lucassen and Schraagen (2010) were used in the data analyses.

	N	Age	Gender		Nationality	
			Male	Female	Dutch	German
High school	13	14.3 (0.6)	5	8	13	0
Undergraduate*	12	23.4 (6.3)	5	7	7	5
Post-graduate	15	27.0 (1.9)	7	8	13	2

For Peer Review

**Table 2: Quality classes according to the Wikipedia Editorial Team Assessment. Detailed descriptions are available on [http://en.wikipedia.org/wiki/Wikipedia:Version\\_1.0\\_Editorial\\_Team/Assessment](http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Assessment).**

Status	Description
FA	The article has attained Featured article status.
A	The article is well organized and essentially complete, having been reviewed by impartial reviewers from a WikiProject or elsewhere. Good article status is not a requirement for A-Class.
GA	The article has attained Good article status.
B	The article is mostly complete and without major issues, but requires some further work to reach Good Article standards. B-Class articles should meet the six B-Class criteria.
C	The article is substantial, but is still missing important content or contains a lot of irrelevant material. The article should have some references to reliable sources, but may still have significant issues or require substantial cleanup.
Start	An article that is developing, but which is quite incomplete and, most notably, lacks adequate reliable sources.
Stub	A very basic description of the topic.

**Table 3: (Corrected) Number of remarks indicating semantic or surface strategy application by the participants of all three groups. Percentages are given in parentheses.**

	Familiar		Unfamiliar		All	
	Semantic	Surface	Semantic	Surface	Semantic	Surface
High-school	215 (63.4%)	124 (36.6%)	163 (50.7%)	159 (49.3%)	378 (57.2%)	283 (42.8%)
Undergraduates	241 (41.6%)	339 (58.4%)	127 (23.4%)	415 (76.6%)	368 (32.8%)	754 (67.2%)
Post-graduates	257 (43.1%)	338 (56.9%)	110 (26.5%)	305 (73.5%)	367 (36.3%)	643 (63.7%)
Average	238 (47.1%)	267 (52.9%)	133 (31.2%)	293 (68.8%)	371 (39.8%)	560 (60.2%)

**Table 4: (Corrected) Number of remarks indicating the utilization of several components of the information by the participants of all three groups. Percentages are given in parentheses.**

	Introduction	Text	Table of contents	Images	Internal links	References	Other
High-school	18 (2.7%)	532 (80.5%)	3 (.4%)	72 (10.9%)	11 (1.7%)	4 (.5%)	21 (3.2%)
Undergraduates	49 (4.3%)	485 (43.2%)	38 (3.4%)	140 (12.5%)	33 (2.9%)	319 (28.4%)	59 (5.2%)
Post-graduates	99 (9.8%)	408 (40.4%)	38 (3.7%)	66 (6.5%)	27 (2.7%)	337 (33.4%)	34 (3.4%)
Average	166 (5.9%)	1425 (51.0%)	79 (2.8%)	278 (10.0%)	71 (2.5%)	660 (23.6%)	114 (4.1%)

**Table 5: Key features used by participants of each group. Strategies were included if at least 50% of the participants in that group applied the strategy at least once.**

	High school	Undergraduates	Post-graduates
Factual accuracy	X	X	X
Completeness	X	X	X
Images	X	X	X
Length of text	X	X	X
Writing style	X	X	X
Quality of text	X	X	X
Scope of text	X	X	X
Understandability	X	X	X
References		X	X
Objectivity		X	X
Structure		X	X
Statistics		X	

**Table 6: Trust in the information on 7-point Likert scales in all conditions (HQ = High quality, LQ = Low quality). Standard deviations are given in parentheses.**

	Familiar			Unfamiliar			All		
	HQ	LQ	All	HQ	LQ	All	HQ	LQ	All
High-school	5.77 (1.03)	5.00 (1.78)	5.29 (1.09)	5.62 (.97)	4.44 (1.96)	5.03 (.96)	5.69 (.71)	4.74 (1.82)	5.23 (.94)
Undergraduates	5.86 (.72)	4.63 (1.31)	5.28 (.75)	5.71 (.56)	4.38 (1.05)	5.00 (.72)	5.76 (.52)	4.52 (1.05)	5.14 (.65)
Post-graduates	5.46 (.49)	4.78 (1.05)	5.16 (.63)	5.33 (.61)	4.82 (.89)	5.05 (.69)	5.37 (.44)	4.84 (.76)	5.11 (.54)
Average	5.69 (.75)	4.80 (1.38)	5.24 (.82)	5.55 (.71)	4.55 (1.30)	5.03 (.79)	5.61 (.56)	4.70 (1.21)	5.16 (.71)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

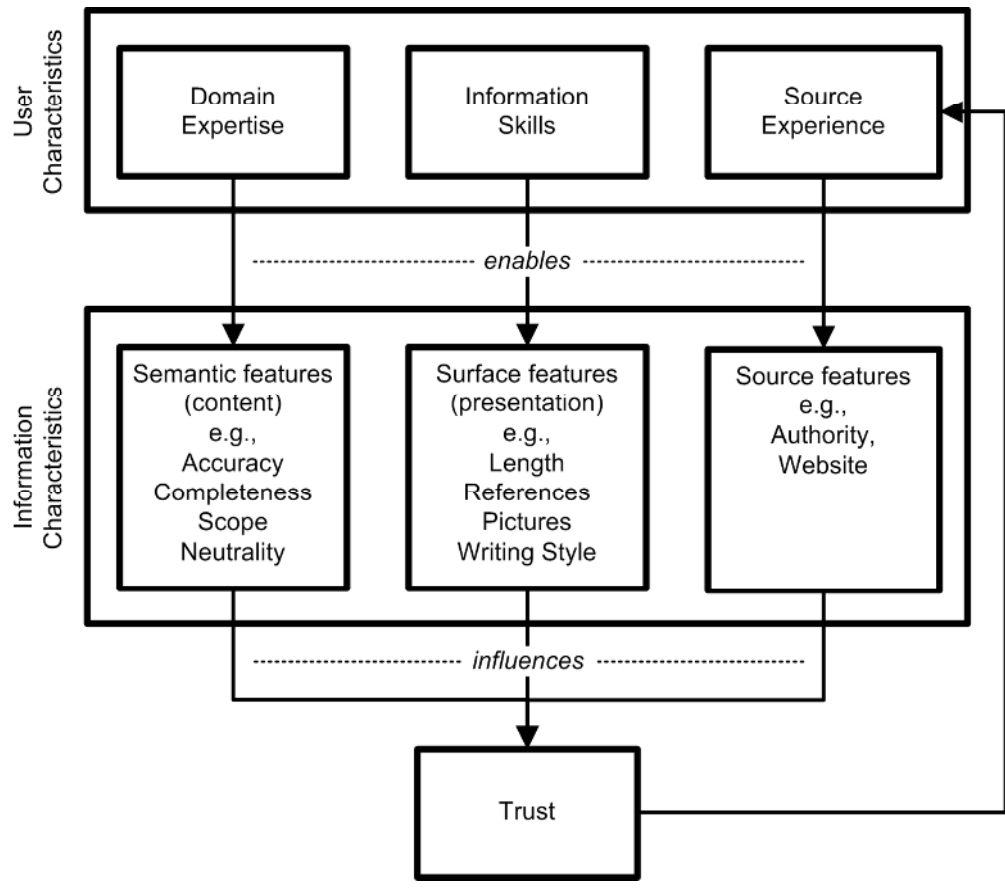


Figure 1: Revised 3S-model of credibility evaluation.