Journal of Information Literacy

ISSN 1750-5968

Volume 12 Issue 2

December 2018

Article

Silva, E., Green, J. and Walker, C. 2018. Source evaluation behaviours of first-year university students. *Journal of Information Literacy*, 12(2), pp.24-43.

http://dx.doi.org/10.11645/12.2.2512

Source evaluation behaviours of first-year university students

Elise Silva, Brigham Young University. Email: elise_silva@byu.edu

Jessica Green, Brigham Young University. Email: jessica_green@byu.edu

Cole Walker, Utah State University. Email: coleericwalker@gmail.com

Abstract

Researchers at Brigham Young University studied first-year students' information evaluation behaviours of open-access, popular news-based, non-academic source material on a variety of subjects. Using think-aloud protocols and screen-recording, researchers coded most and least used evaluation behaviours. Students most used an article's sources, previous experience with the source or subject matter, or a bias judgement to decide whether the source was reliable. Researchers also compared what students said was important when evaluating information vs. what behaviours students actually exhibited and found significant differences between the two. Namely, students did not think their previous experience or bias judgement affected the way they assessed sources; however, both behaviours played prominently in their observed source evaluation techniques across the study.

Keywords

first-year students; information literacy; popular sources; source evaluation; US

---

1. Introduction

In an era of fake news and disinformation, as well as politicised social media content, the question of how to teach students basic information evaluation behaviours is critical. Given rapidly changing information-consuming behaviours and the fact that the internet has no traditional gatekeepers, the onus is on individuals to establish the credibility and reliability of the information they are consuming (Metzger, 2007). As the ubiquity of internet access in developed countries increases, the way people access news has the power to affect world views. In an age where authority is being questioned and traditional markers of authority dismantled, how do librarians keep their fingers on the pulse of an information society their discipline tasks them with educating?

Though librarians are eager to respond and help instruct students in productive source evaluation techniques, little research exists investigating how students actually evaluate online, open-access, popular content. Though studies exist on how students evaluate source materials, methodologies vastly vary from study to study. Furthermore, given how quickly social media and information usage change, more up-to-date, data-driven studies are needed to study the information evaluation behaviours of university students, who are inundated with vast amounts of material on a daily basis.

Studies have established that students struggle with information evaluation. As the Stanford History Education Group recently reported, students of all ages struggle to critically engage with

online information sources, and, even though it is counterintuitive (especially given the amount of time university students spend online each day) college students struggle with age-appropriate critical thinking tasks (2016'). What the studies could expand upon, however, is *how* students are evaluating information sources organically, as they come across them online.

The researchers who designed the current study wished to understand the most and least used information evaluation behaviours of first-year university students as well as compare what these students thought was important when evaluating information vs. what evaluative behaviours they actually performed when accessing a source. In many ways, this study responds to recent information science research which asserts that 'to understand how students evaluate sources, a clearer understanding of the criteria they use to evaluate those sources is needed' (Taylor & Dalal, 2014). In contrast to much research done in this area that includes students self-reporting evaluation methods, the current study's methodology includes think-aloud protocols (TA) and screen recording. These methods allow researchers to gain a clearer picture of how students are actually evaluating information sources. As the study furthers knowledge about how students evaluate online source materials, it forwards strong implications for librarians and others who instruct students in digital and information literacies.

2. Literature review

This literature review will focus on three main discussions relating to source evaluation scholarship: research related to credibility and how it is defined, studies on how undergraduate students evaluate and use online source material, and trends in information literacy (IL) instruction related to source evaluation.

2.1 Defining credibility

The definition of information credibility has been historically hard to pin down, yet according to Metzger and Flanagin (2013) the concept goes back to the teachings of Aristotle and the birth of rhetorical principles like ethos (appeals to authority), pathos (appeals to emotion), and logos (appeals to logic). However, "more modern accounts of credibility define it as the *believability* of a source' and believability is in the eye of the beholder – or in this case, the information consumer (p.211, emphasis in original). Credibility is so important because it precedes trust: 'credibility [is the most] important antecedent to trust' (Rowley, Johnson, & Sbaffi, 2015, p.324). Librarians have described the concept of 'trust' as something highly contextual and personal. For Johnson, Rowley, and Sbaffi (2015a) 'trust' is a complex contextual attitude with several contributing factors including perceived competency and honesty (p.416).

As the current information environment online does not often include professional gatekeepers who filter information sources, traditional indicators of authority, competency, and honesty can be hard to spot by consumers – and therefore what to trust online is hazy. Given the way information is disseminated on the web, 'the need for effective critical evaluation [is] pressing' (Metzger & Flanagin, 2013, p.211) and defining credibility remains important. Taylor and Dalal (2014) summarise that there is often a disconnect between how instructors define credible and reliable material and how students understand those instructions. Students have trouble understanding what instructors mean when they ask them to find reliable source material, even though instructors believe they have fully fleshed out this concept for their students. With this in mind, even though credibility is difficult to define, its definition still matters to students who wish to understand what makes source material reliable, and to instructors who wish to be more transparent in their instructions to students.

As Flanagin and Metzger (2008) assert, credibility studies are interdisciplinary, and definitions across fields have both commonalities and differences. Most importantly, they break down credibility definitions into sub-areas like source credibility, message credibility, and media credibility. Although scholars distinguish between these different markers of credibility, information consumers, such as students, may not (p.9). Those who study trust in source evaluation likewise

look at different meanings of the term. Johnson, Sbaffi, and Rowley (2014) suggest that trust is a 'dynamic concept' (p.68). For instruction librarians this begs the following question: how are information consumers, particularly college students, defining 'credible' information for themselves? What do they trust, and why?

2.2 Source evaluation behaviours of university students

Studies repeatedly show that students struggle to evaluate information sources. According to Faix (2014), novices (like those new to college) are most likely to use superficial criteria and surface features such as ease of access and web design to evaluate how credible source material is. While they show some hesitancy with design that is too flashy or overbearing, these novices are generally unable to point out what characteristics authoritative sources actually display. Ease of use factors into how students go about finding information – explaining student affinity for Google searches (Johnson, Rowley, Sbaffi, 2015b, p.835). Other studies have backed up the findings that those new to research are more interested in the ease and accessibility of a source; its surface features (like pictures or graphics); and its perceived 'coolness' than other, more objective, markers of reliability, including content (Sundar, 2008; Brand-Gruwel, Kammerer, Meeuwen, & Gog, 2017; Harris, 2008; Sillence, Briggs, Fishwick, & Harris, 2004). Connaway, Dickey, and Radford (2011) explain that students are self-interested and have serious time constraints. They choose the easiest sources that are 'good enough' for them to finish their research task (p. 180). In brief, students settle.

Despite growing evidence that students struggle to evaluate information, 'students tend to trust their own judgement when determining the credibility of sources they [find] on the internet.' (Dubicki, 2015, p.676). Interestingly, there is a discrepancy between what students say is important vs. how they actually perform in source evaluation situations. In other words, 'the criteria students say they will use are not the same criteria they actually use' when evaluating information (Taylor & Dalal, 2014). Students especially struggle with identifying sourced material within information sources, author information, and document type (Strømsø, Bråten, Britt, & Ferguson, 2013, p. 198). Not surprisingly, Google search is widely used by university students who are seeking information, and the majority of those students believe 'that pages returned by search engines usually contain accurate information' (Taylor & Dalal, 2014). Perhaps this has to do with the motivation of the user.

Information evaluation behaviour depends on how motivated a student is, and the context of his/her information-finding situation (Connaway et al., 2011). Most students new to research have difficulty defining their information need and do not naturally evaluate texts. Instead, source evaluation is found to be a learned behaviour that develops in conjunction with domain expertise (von der Mühlen, Richter, Schmid, Schmidt, & Berthold, 2016). Finally, students rely on easier behaviours to evaluate – like ones that include opinions or quick judgements – than behaviours that require more work – like research or corroboration – to assess information value (Metzger, 2007).

The previous point is related to confirmation bias – another issue that affects the ways students interact with information. While this phenomenon is widely studied, it is concerning (especially given student tendencies to look at surface features, and trust gut reactions to information found online). As van Strien, Kammerer, Brand-Gruwel, and Boshuizen (2016) assert, 'Prior attitudes can affect information processing in profound ways. For instance, people are almost two times more likely to select information that supports their prior attitudes' (p. 246). Other researchers who have studied issues related to fake news find that falsehoods spread faster than truth (Vosoughi, Roy, & Aral, 2018). Because of highly politicised and polarising media content, the trend of disinformation on the web continues to be concerning for information professionals seeking to help create a well-informed, decision-making public.

Finally, there have been conversations about the way students juggle evaluating multiple texts at once. Strømsø et al. found that 'students fail to pay sufficient attention to details of the source

information' to make informed decisions about conflicting information sources (2013, p.195). Other sources corroborate these students have difficulty assessing and corroborating information from various sources at once (von der Mühlen et al., 2016).

Given the difficulty students generally have in assessing source material in contextual reading situations, how are librarians and curriculum experts responding to, or teaching about, information literacy in a digital age?

2.3 Trends in information literacy source evaluation instruction

In response to recent studies which show that students are woefully undereducated on source evaluation techniques, the Stanford History Education Group calls for more robust curriculum 'focused on developing students' online reasoning' (2016, p. 7). Yet, what is currently taught about this topic?

Librarians approach teaching source evaluation in a variety of ways. Most common, perhaps, is the checklist or acronym approach. The acronyms like CRAP (which stands for currency, reliability, authority, purpose/point of view) walk students through evaluation criteria to consider when evaluating a source. Checklists have students look at elements of a source to rate it on a credibility scale. This way of teaching source evaluation has been questioned, however, in its ability to teach critical thinking (Ostenson, 2014). In lieu of a checklist or acronym to teach source evaluation concepts, Ostenson (2014) and Chinn and Rinehart (2016) have suggested decision models to help students walk through problem-based source evaluating practice (Ostenson, p. 39). Chinn and Rinehart (2016) suggest that although surface features (like those considered in a checklist approach) can be reliable in source evaluation, augmenting such behaviours with 'an analysis of the underlying processes that are actually used by the source' is a better way for students to understand how information comes to be and why sources function the way they do (p.1714).

Other evaluative teaching methods include having students elaborate on sets of questions that ask about the genre, bias, purpose, or other aspects in a source (Pickard, Shenton, & Johnson, 2014, p. 6). This is a close cousin to the checklist, though more open-ended. Furthermore, this kind of approach encourages more close reading of a source, and arguably, more critical thinking,

Finally, lesser-known models include assessing a source's cognitive authority – how the source fits within the wider body of information – and iterative models of source evaluation – which walk students through levels of credibility assessment (Metzger, 2007). With this in mind, it is clear that there are several different ways librarians and IL instructors have used to teach students about source evaluation. Yet, without a clear understanding of how students are actually approaching sources they consume online, how are library professionals to create solid curricular standards? The current study hopes to shed light on this quandary.

3. Methods

3.1 Research questions and recruitment

Brigham Young University (BYU) in Provo, Utah, USA is a large, private, highly-selective religious university. Of the 31,000 students who attend the university, approximately 65% take a first-year writing (FYW) course at some point during their studies. Source evaluation behaviours are taught during this course along with other IL skills. Librarians wondered how these students, new to the university, navigated information they found on the unpaid, open web, and specifically, what source evaluation behaviours were most and least used by them before they received any information literacy instruction by librarians. Furthermore, researchers wondered what students thought were the most important things to consider when evaluating source materials, and if that would match up to observable source evaluation habits.

In the summer of 2017 there were 23 FYW classes taught. To recruit participants, researchers advertised using in-person class visits, flyers, and emails. To encourage participation, FYW students were given $10 campus gift cards to participate in the study. In total, 89 FYW students participated, which was 20% of the total enrolment over the summer of 2017.

3.2 Test design

The test was designed as a survey where students had to evaluate five different source types using think-aloud (TA) protocol and screen recording. In order to assess the students' source evaluation skills, researchers selected five articles of varying degrees of quality. The articles selected, as well as justification for why they were chosen, are listed below:

Table 1: Articles selected for study

Article title	Source	Selection criteria
'Over-The-Counter Birth Control Pills Would Be Safe For Teens, Researchers Say' (Hersher, 2017)	*National Public Radio (NPR)*	Chosen as a reliable news article that cited academic research and other reliable sources.
'A Third Way for Universities' (Wallace, 2017)	*Huffington Post*	Chosen as an opinion editorial from a well-known website that appeals to millennials.
'Global Warming Fail: Study Finds Melting Sea Ice is Actually Helping Arctic Animals' (Haskins, 2017)	*The Blaze*	Chosen for a far right bias, inaccurate use of data and information, a hot-button political issue, and use of visuals. (The online version of this article no longer includes a graph from NASA delineating melting sea ice levels that was in the original version)
'Elon Musk's SpaceX Makes History by Launching a 'Flight-Proven' Rocket' (Davenport, 2017)	*The Washington Post*	Chosen to represent a well-known newspaper reporting on an event, for its apolitical subject, and inclusion of a video.
'There's a Growing Crisis in Care for Disabled and Elderly People. Oh, and It's a Jobs Crisis, Too' (Clawson, 2017)	*Daily Kos*	Chosen for a far left bias and use of a combative, casual tone.

Researchers felt that these five articles represented a fair mix of biases and source types.

The test was designed using Qualtrics, a web-based survey design tool. The first page of the test asked basic demographic questions of the participant. The next five pages (one for each article) asked the students to evaluate the sources. The test asked the students to take two minutes to evaluate the screen shots of the article, and rate it on a scale from 1–10 based on how reliable they thought the article was (1 being very unreliable and 10 being very reliable). After the article screenshots, the test asked the students to take another two minutes to use the internet browser to do any online research required to evaluate the credibility of the overall source, and rate its reliability again on a scale of 1–10. After the second rating, the test presented two free response boxes that gave students space to write what qualities/attributes made the source more or less reliable to them (See Appendix I, Figure 5).

After the five sources were evaluated, the last page of the test asked the students to select from a list the attributes that help them determine source reliability. These included the following: publishing date; currency of facts; hyperlinks to outside sources; sources cited at the end of article/website; writer's authority or background; design of website; advertising on/in website or

article; domain type; the way the article is written (tone, grammar, etc.); the genre of the source; how they found the source; and their emotional response to the source. The final question asked how often they checked the previously selected attributes when considering a source (See Appendix I, Figure 6).

To understand what participants were thinking and why they were making certain decisions, researchers implemented think-aloud (TA) protocols into the study. Each page informed the participant that their voices would be recorded and encouraged them to 'think out loud'. To gain further information, researchers also recorded the computer screen of each participant. Between the concurrent TA reporting, screen capture, and written test results, researchers hoped to understand accurately how students were evaluating sources and why they made the choices they did.

3.3 Pre-testing

Before the test instrument was used with FYW students, researchers tested the instrument with ten library student employees to make sure the directions were clear and to make any last minute changes. From the pre-testing, researchers added additional clarifications to the test to minimise user errors. Furthermore, researchers realised the need for students to use an incognito browser to eliminate search engine auto-suggestions while participants conducted research on the different articles.

3.4 Test administration

Students that indicated their interest, either through email or sign-up, were contacted by a member of the research team to schedule a time to come to the library and participate in the study. The study was conducted in the library's usability lab, where researchers were able to easily record voice and capture the screen of each participant. Before taking the test, a test administrator showed students into the lab, briefly explained the test, and then student participants read and signed a consent form before beginning. Students remained in the room alone until they finished, which took anywhere between 20 and 50 minutes. Participants could leave the room at any time during the test if they wished, or could leave the test unfinished if they wished. All participants were given a $10 campus gift card as they exited the room whether they finished the test or not.

Out of the 89 students who participated in the test, 5 of the students were unable to complete the test, and their results were consequently removed from analysis.

3.5 Analysis

After completion of the test, each participant's session was transcribed. During this process, the transcriber also made note of what was occurring on the screen and included any unspoken research behaviours, such as following hyperlinks or fact checking. After this initial transcription, a second member of the research team went through the transcription, checking it against the recording and ensuring that it was transcribed accurately.

Next, each transcription was coded for observed evaluation behaviours. While the researchers did survey much of the previous research completed on specific evaluative behaviours (Metzger, 2007; Pickard et al., 2014), ultimately they decided to allow the evaluative behaviours that they focused on to emerge from the transcriptions rather than from a pre-determined list. One area where the coding became more complex was in the area of 'emotional response'. Researchers found that there were three distinct (but related) behaviours they witnessed in this category, emotional response, bias judgement, and previous experience.

The researchers ultimately selected 15 behaviours to code for:

Table 2: Evaluation behaviours

Behaviour	Explanation	Example quote from study
Sources	the student observed that the article cited/did not cite other sources	'And NASA is automatically reliable, if they correctly sourced them.'
Previous experience	the student referred to his or her own previous experiences with the website or content	'I've never heard of this.' Or 'I am pretty familiar with other things, op eds, that *The Washington Post* has published.'
Bias judgement	the student was influenced by his or her own biases (political, religious, etc.)	'This just seems like a really conservative post or media company.' Or 'I already know I don't believe half the things they say.'
Venue	the student commented on the reliability of the venue of the article	'The Washington Post is a very recognised, very famous newspaper.'
Author	the student commented on the reliability or authority of the author	'Contributing editor. Linked in. She has a PhD in sociology…so she has good credentials.'
Visuals	the student commented on the use/lack of visuals (pictures, graphs, videos, etc.) in the article	'So this is actually pictures, a video. So I'm going to go ahead and watch this video.'
Genre	the student commented on the genre (opinion editorial, news report, etc.) of the article	'It comes from a radio news broadcast, which makes me feel like the source may be more general.'
Fact checking	the student exhibited clear fact checking behaviours to verify information in the article	'Let's see if she is actually the president.' Searches contributor name.
Ads	the student commented on ads on the page	'This [ad] right here just seems…a little questionable.'
Website design	the student commented on the design (layout, professionalism, etc.)	'The news are not divided by section as in the previous websites.'
Grammar/style	the student commented on the grammar or style of the article (tone, typos, misspellings, etc.)	'It's one long run-on sentence pretty much.'
Hyperlinks	the student commented on or clicked on hyperlinks within the article	'The fact that it actually has links to the study and information cited.'
Emotional response	the student showed clear signs of an emotional reaction to the content of the article or website	Reads title. 'Oh, that's awesome!'
Date	the student commented on the publishing date of the article	'April 2017, so that's just, you know, a few months ago.'
Domain	the student commented on the domain type (.org, .com, .edu, etc.) of the article	'So this Huff Post, yeah, it's a .com.'

Where a specific instance in the transcription could be construed as more than one criterion, three team members met to decide how to code that item. After coding on all transcriptions was completed, each behaviour was tallied for later statistical analysis.

3.6. Ethics

BYU's Institutional Review Board (IRB) approved study design, advertisement, and use of incentives. The IRB protocol number for this study is as follows: E17139.

4. Findings

4.1 Demographics

To understand who the participants of the test were, researchers asked a few questions about age, sex, and how much college the participant had completed. Out of the 84 students whose results were analysed, 55% were male and 45% were female. About half of the test participants were aged 18, with 38% aged 21 or over. The majority of the students were starting their first semester of university, with 29% having completed one to two semesters and 14% with three or more semesters completed.

4.2 Results of student ratings

The first task researchers assigned each participant was to evaluate screenshots of five different articles based on a brief, two-minute assessment (see Table 1 for article hyperlinks). They could rate the articles' reliability anywhere from a one to a ten on a slider scale (one being low reliability and ten being high reliability). After this initial score, students were asked to perform two minutes of online research and score the articles again. Students had to self-regulate their time so some took significantly longer or shorter to complete these tasks.

Researchers analysed the students' mean initial ratings (light blue below) with their mean ratings of the sources after additional research (dark blue below).

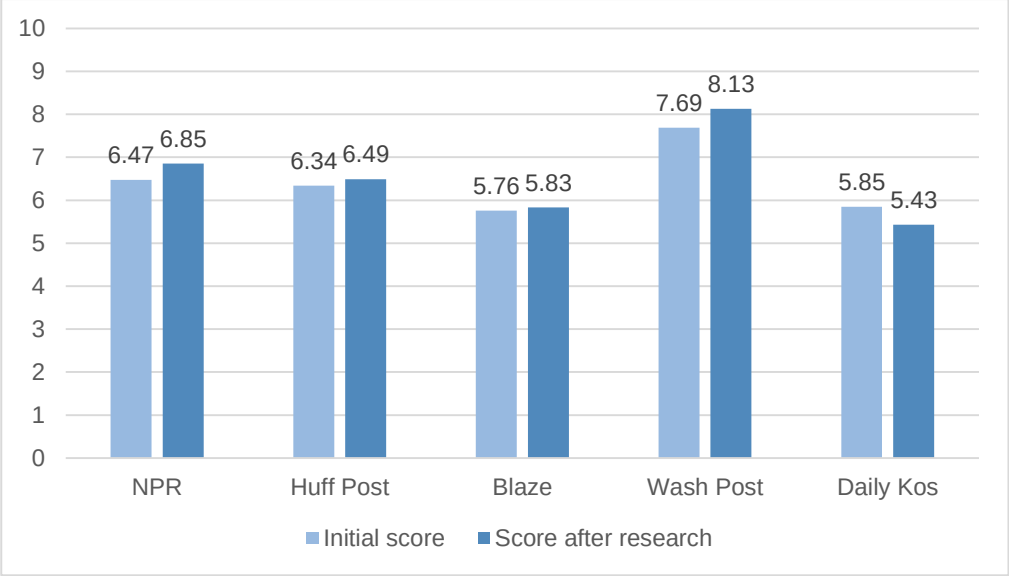


Figure 1: Mean student scores for sources

Students generally found the *Washington Post* article to be the most reliable, followed by *NPR*, *Huffington Post*, *The Blaze*, and then *Daily Kos*. *Daily Kos* was the only source where the mean score actually dropped after online research. Every other source scored at least slightly higher in its mean score after the students were given an additional two minutes to research (Figure 1).

Difference between initial mean score and mean score after research:
NPR +0.38
Huffington Post +0.15
The Blaze +0.07
Washington Post +0.44
Daily Kos -0.42

The source that had the greatest change in mean score after the students completed research was the *Washington Post*, with an overall increase in mean score of +0.44. *Daily Kos* was the only source to have a decrease in overall mean score after the students completed research, with a decrease of 0.42. *The Blaze* was the mean score that changed the least between the initial scoring and the scoring after research, with a difference of only +0.07.

Another measure that sheds light on how students understood the different sources is the standard deviation, or how much variation or dispersion occurred in the data set. The lower the standard deviation, the more the students tended to rate the source close to the mean rating. The higher the standard deviation, the more extreme (higher or lower) the students tended to rate the source from the mean.

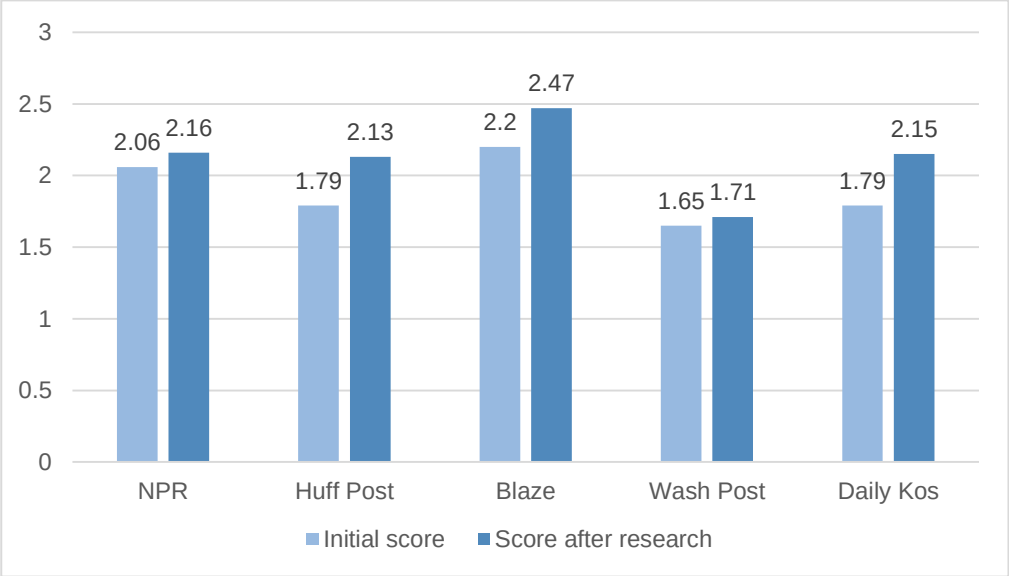


Figure 2: Standard deviation of student scores for sources

In terms of the standard deviation, the source with the greatest amount of deviation amongst students was *The Blaze*, followed by *NPR*, *Daily Kos*, and *Huffington Post*. *Washington Post* had the lowest standard deviation. Across all sources, the standard deviation was consistently lower on the initial scoring than the second scoring (Figure 2). This means that online research created a greater extreme in how students scored the sources, suggesting that students were more likely to score an article on the extreme low or high end of the spectrum after research.

Difference in standard deviation between initial score and score after research
NPR +0.10
Huff Post +0.34
Blaze +0.27
Wash Post +0.06
Daily Kos +0.36

In addition to having the lowest overall standard deviation, the *Washington Post* also had the least change in standard deviation between the initial and second scoring, with only a +0.06 change between the two scores. *NPR* had the next least change in standard deviation, followed by *The Blaze*, *Huffington Post*, and then *Daily Kos*.

Researchers next analysed the source ratings by performing paired t-tests on the initial scores with the scores after research. There were significant differences found in the two sets of ratings for *NPR*, *Washington Post*, and *Daily Kos*, but no significant differences found between the sets of ratings for *Huffington Post* and *The Blaze*.

Table 3: Repeated measure t-test results

	Mean	SD	P-value
NPR	-0.3482	1.22	0.0083
Huff Post	-0.1682	1.332	0.2363
Blaze	-0.0452	1.712	0.8042
Washington Post	-0.4542	1.212	0.0007
Daily Kos	0.3862	1.2829	0.0058

Boldface indicates significant difference

The above values are calculated based on the paired/repeated measure t-test procedure. The mean was calculated by first calculating the difference between the students' initial rating of each source and the students' rating of the sources after research, and then taking the average of those differences. The standard deviation was calculated in the same way. The p-value comes from a two-tailed t-distribution with n-1 degrees of freedom, where n is the number of respondents per question.

4.3 Source evaluation behaviours

Throughout the administration of the test, each page encouraged students to use think-aloud (TA) protocols as they scored each source, completed research, and provided details about what made each source more and less reliable. Students' screens were also recorded.

While students could and did use many behaviours multiple times throughout the test, researchers were interested to compare overall which behaviours were used by each student at least once during the test. Figure 3 below illustrates the number of students that researchers observed using each evaluation behaviour.

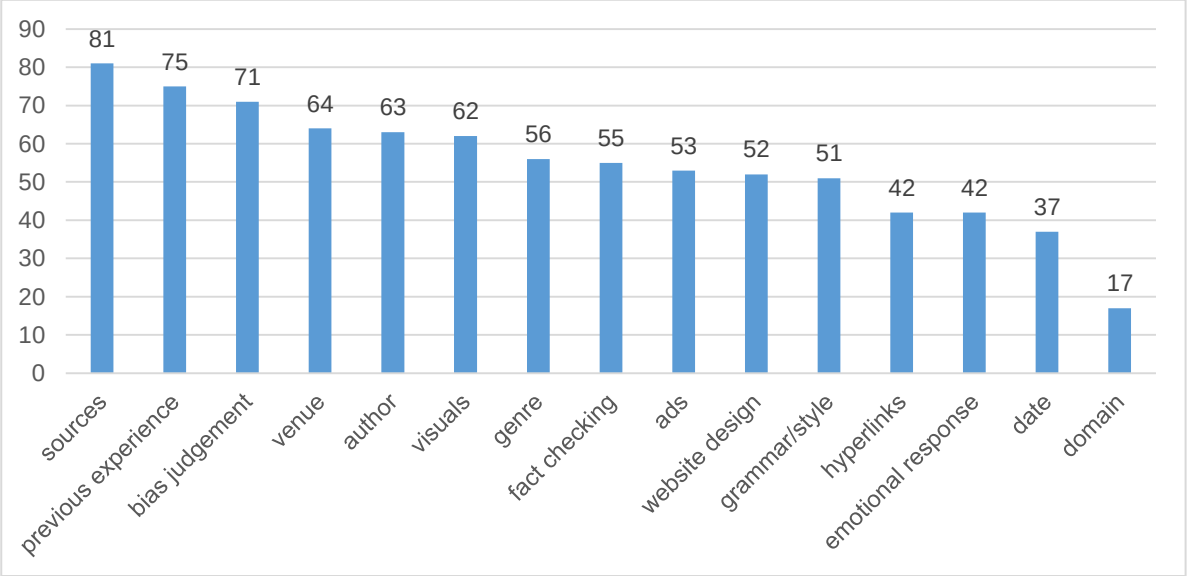


Figure 3: Most used evaluation behaviours (per 84 students who completed test)

Out of 84 subjects reported on, 81 used sources in the article as a way to evaluate its reliability. This could take a variety of directions, from commenting on what facts were sourced, whether specific organisations were named as the basis for evidence, or just that an opinion was not backed up by a source.

Other high-use behaviours that were observed affecting students' analysis of a source's reliability were previous experience (75 students) and bias judgement (71 students). Previous experience might include prior use of a publication venue or prior knowledge of the subject matter of the article. Students mentioned other articles they had read from the website and classes or books they had read on a topic. Overall, students seemed more familiar with the *Huffington Post* and *Washington Post*, showing brand recognition and sharing examples of whether or not they regularly read or agreed with the source. *NPR* and *The Blaze* were less familiar for most students, and students in general had not interacted with *Daily Kos*.

The research team defined bias judgement as a behaviour where students made judgements based on their political, social, economic, or religious bias, often discussing the way the article conformed or did not conform with their views. For example, in one of the articles, some students reacted to the discussion of the safety of birth control for teens by explaining that they agreed or disagreed with the ability of teens to access birth control, and then used their personal views as the basis of their evaluation of the article's credibility. In another article, many students' views on global warming distracted them from actually rating the article's reliability – they focused on relating their views and justifying why the article fit or did not fit with their views.

Fact-checking behaviours, including independently verifying statistics used in the articles or looking into claims made about the subject matter, were exhibited at least once throughout the study by 55 out of the 84 students. Low-use behaviours were criteria like the source's domain (17 students) and date of publication (37 students).

Another way to understand which evaluative behaviours are most common in students is to examine how frequently the students used each behaviour as they evaluated the five articles.

Whereas with Figure 3, where researchers examined which behaviours were used even once during the course of the study, Table 4 describes how often students used the behaviours as they looked at the five sources. If the students used a behaviour five times, researchers considered that student falling into the 'every' frequency range, since there were five sources to evaluate. Many of the students fell into the frequency range of 'few' (one to two) or 'most' (three to four) when considering how many times they used specific behaviours. The frequency 'never' varied the most, with only three students not using the evaluation behaviour 'sources' at all, but well over half of participants not using the behaviours date or domain at any point in the study.

Table 4: Observed behaviour frequencies

	Observed Frequency			
	Never (0)	**Few (1–2)**	**Most (3–4)**	**Every (5)**
Sources	3	17	29	35
Previous Experience	9	28	32	15
Bias Judgement	13	35	22	14
Venue	20	34	23	7
Author	21	35	16	12
Visuals	22	47	14	1
Genre	28	42	13	1
Fact checking	29	30	16	9
Ads	31	25	18	10
Website Design	32	36	10	6
Grammar/Style	33	28	16	7
Hyperlinks	42	20	8	14
Emotional Response	42	31	11	0
Date	47	30	4	3
Domain	67	11	5	1

The top three behaviours used five times over the course of the test are sources, previous experience, and bias judgement – a finding that directly corresponds to the top behaviours used at least once throughout the test. Hyperlinks (pointing out that the article has hyperlinks or clicking on hyperlinks), a behaviour that was relatively low when considering the number of students who used it at least once, falls at the higher end of the spectrum in terms of students who used it five times. Visuals and genre were two behaviours that fell in the middle of the spectrum when examined by how many students used them at least once, but had only one student in each category use these behaviours every time throughout the study.

The behaviours ads, grammar/style, author, and fact checking were used with most of the sources (three to four times) by 16–18 students. These behaviours fell towards the upper middle of the evaluation behaviours in this case, suggesting that if these behaviours are used by students, they will be used more than once. About half the behaviours (visuals, genre, website design, emotional response, date, and domain) were used by relatively few of the students more than 1-2 times, and

half the students did not use hyperlinks or emotional response even once. Date and domain fared even worse, with 47 and 67 (out of 84) students not using them at all.

4.4 Observed behaviour vs. what students say is important

Towards the end of the test, students were asked to select from a list any of the evaluative techniques that they thought would be important in considering a source. For almost all of the criteria listed, at least half of the students marked it as important.

One of the most significant findings of the research was found in comparing the difference between what students said was important with what researchers were actually able to observe them doing when participants were given a specific research task. Figure 4 below illustrates what students selected as being important in evaluating credibility (dark blue) with what evaluative behaviours researchers actually observed them using (light blue).

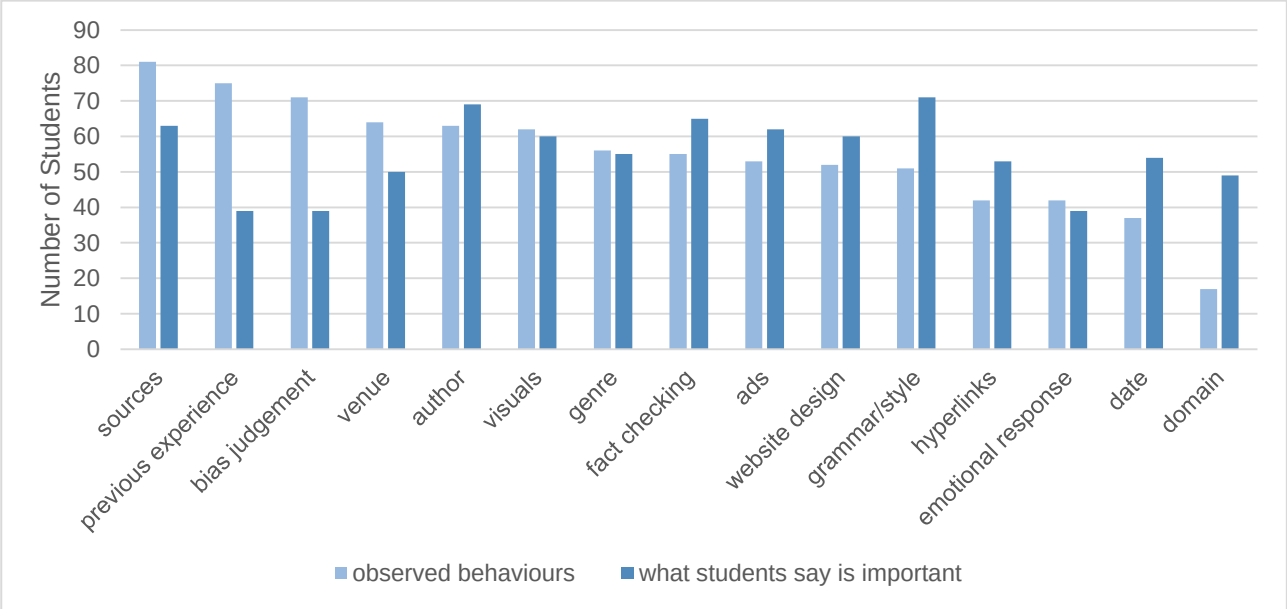


Figure 4: Evaluation behaviours compared (per 84 students who completed test)

Student behaviour varied the most from the behaviours they identified as important in two specific categories, previous experience and bias judgement. Just under half of participants identified previous experience or bias judgement as an important evaluative behaviour, but researchers observed much higher levels of use of these two behaviours throughout the study. Seventy-five of 84 students were observed using previous experience and 71 students using bias judgement. Examples of behaviours that students said were important in evaluating a source but that students did not actually use as much were criteria like domain, grammar/style, date, and fact checking.

As stated in the methods section, students self-reported on the research behaviour of ‘emotional response’ at the end of the study, which researchers then split out into three related but distinct categories of relying on previous experience, bias judgement, and emotional response while transcribing and coding. While researchers admit this may have swayed the results, in some ways it made them more conservative since if all of these behaviours would have been lumped together, there would have been a much higher frequency of ‘emotional response.’ In other words, though students did not see themselves as having personal responses to these articles, most of them demonstrably did, and used those responses as a way of judging information credibility.

5. Discussion

At the outset of the study researchers wished to understand the most and least used behaviours of first-year students while evaluating web sources. They also wished to understand what students thought the most important source evaluation techniques were in these contexts, and if those student opinions matched up with how students actually assessed source material. The below trends drawn from the data analyse these questions, and other issues like how research changed student perception of a source, the role of previous experience in evaluating material, and the role of bias in source evaluation

5.1 Why did students' mean scores not change drastically from before researching the articles to after researching?

Researchers who conducted this study assumed that, given the wide variety of source types, genres, topics, and political tones present in the study, (some of which were believed to be familiar to students and others that were expected would not be), the average ratings of at least some of the sources would more drastically change after research. As the mean scores did not change significantly after researching the articles, the authors assume that either students were persuaded by their initial 'gut reactions' of the source, or that their source evaluation behaviours were not sophisticated enough to help them commit to a much higher or lower score – even for the sources that warranted such extremes. While the standard deviations of the scores after research did show more extremes than the before research scores, the fact that the overall average ratings did not change drastically is telling. Even though the sources were widely different, students tended to be non-committal scorers and had difficulty differentiating between different source types and articulating the different qualities of sources via their ratings. Other literature corroborates this trend as students had trouble 'explain[ing] exactly what criteria they were using' while evaluating information (Faix 2014, p. 627).

5.2 What might explain students' most and least used research behaviours?

The study's findings corroborate that students rely on surface features to determine reliability. For example, the most used evaluative behaviour in the current study was 'sources', where a participant simply had to mention the article's use (or lack of) source material to be counted as a behaviour. While this is a productive impulse on the part of the subjects, many fewer students actually clicked on sources (the research behaviour coded as 'hyperlinks') or corroborated information (the research behaviour coded as 'fact checking'). This shows that students may recognise the importance of features like sources, but do not expend the effort needed to actually check up on those sources or corroborate whether information is truly viable.

What is meaningful in this finding is that students are recognising the importance and value of sourced material, and even articulating that importance. However, students seem to have more difficulty checking the source material itself for bias, reliability, or value. As information experts understand, cherry picking sources for information is easy, as is hyperlinking to sources that seem to corroborate a point of view, but interpreting sourced material is difficult and time-intensive. If the results of this study are any indication, college students need to engage in more critical thinking about what kinds of sourced material lend authority to the article in which it is cited.

5.3 What role does previous experience play in how novices evaluate popular information?

Researchers found that after the research behaviour 'sources,' 'previous experience' was the next most frequently used behaviour, though students did not self-report being persuaded by previous experience with much frequency. While related, researchers delineated between 'previous experience', 'bias judgement', and 'emotional response' with the first two behaviours used frequently in the study by subjects. The authors acknowledge that experts also utilise 'previous experience' as they make judgements about information sources (Brand-Gruwel et al., p.236).

However, these results show that novice students using previous experience to help them make value judgements about sources were unable to pick out highly biased information sources from fact-based sources. Therefore, without domain expertise, students relying on their previous experiences with these sources or topics did not necessarily help them evaluate them any better, suggesting that previous experience must be strong and domain-specific in order to help make valuable information-evaluation judgements.

5.4 What role does bias play for the novice?

Bias judgement was the third most used behaviour in the study both in terms of high frequency and in terms of it being used at least once by a large percentage of students throughout the study. Some students were aware of their biased reactions to sources. In the words of one of the student subjects about the article on global warming in *The Blaze*: 'I'd give it an 8 at start just cause, funny [sic] enough I just feel like the article supports my own opinions, and feelings about the subject. So that makes me, right off the bat, think that it's a more reliable source.' Most students, however, were not aware of their bias, as is apparent by the self-reporting mechanism. Bias is clearly an issue students need more training in, especially in how it relates to previous experience and objectively assessing a source's credibility.

5.5 How does what students *say* is important compare with what students actually *do* when evaluating sources?

Researchers found that there were some important differences in how students are actually evaluating sources when compared with what students say or believe is important. In the self-reporting at the end of the test, students identified grammar/style, author, and fact-checking as the top three criteria in judging a source's credibility (Figure 4). Yet when looking at the behaviours that researchers observed most frequently, none of these three behaviours ranked nearly as high.

Similarly, amongst the top three behaviours that researchers observed during the course of the test (sources, previous experience, and bias judgment), only the behaviour sources was also reported as important by students, with 63 students marking it a criterion of reliability. Clearly, helping students find a way to recognise the actual criteria they are using when engaging with a source and having more integrity between what they know is important and how they actually evaluate need to be a priority in teaching source evaluation skills to students.

6. Limitations

While there are a few limitations apparent in the study methods, researchers believe that the overall study is sound and has gathered important information for library professionals.

First, students were not given a specific research context in which to evaluate these sources. Some studies give students specific research contexts such as a class project or a personal information need. For this study, researchers simply presented students with the sources and asked them to rate their credibility. This was done self-consciously by researchers as they wanted to find out how students were organically evaluating information sources. However, some findings might have changed had students been given a specific research need.

Second, while researchers did ask participants some basic demographic questions, the test did not attempt to correlate the students' department or course of study with their results. With such a small sample size, researchers could not have drawn conclusions on whether a specific discipline was more likely to rate sources differently or use different evaluative skills, but this is certainly an area for further study and future research.

Third, the student body at the researchers' university was not as diverse as researchers had hoped, with more than half of participants reporting that they identified as politically conservative (and the other half identifying as unaffiliated, independent, or liberal). As no data about the effect of

political affiliation on research behaviours has been reported on in this article, researchers do not believe that this is a salient reason to discount study findings.

Finally, researchers used the word ‘reliability’ to ask students to rate the articles from 1–10. However, the literature on this subject often uses the words ‘credibility’ or ‘believability’ to describe this objective value of a source. Reliability has a slight connotation of truthfulness or trustworthiness which is why the researchers chose the term – they wanted to know whether a student would trust the information found therein.

To reiterate, though there were some limitations to the study, researchers are confident with their findings. Through pre-testing, careful review of the literature, and extensive recording, transcribing, and coding, the authors of the study believe that the results of the study are sound.

7. Implications and conclusion

The presented study has serious implications for further research as well as implications for IL instruction librarians as they interact with early university students.

As this study gives significant insight into most and least used source evaluation techniques by students, follow-up studies might articulate why certain techniques are used by students over others. Also, follow-up studies which would test the effect of IL instruction on how students are using different techniques could be telling.

In terms of teaching strategies based on findings, first, librarians must self-consciously teach novice students how to build off their initial reactions to trust surface features like sources and to actually corroborate those sources. Part of this instruction may focus on the differences between surface features and the more critical thinking that goes into other features of an article, but also the affordances and strategic ways to corroborate information.

Instruction in domain expertise as it relates to prior experience and information-evaluation behaviours could be a productive area on which to focus. As the study found students utilizing prior experience often, and yet still having difficulty evaluating sources, IL librarians can help students understand what kinds of previous experience are helpful in evaluating sources and how to develop useful previous experience in background research to objectively evaluate.

In order to combat personal bias (or at least acknowledge it as a way individuals interact with information) IL instructors can develop lesson plans which help students understand their own biases and think critically about how this helps or hinders their relationship to information sources. Ultimately, these critical thinking approaches will help students as they interact with sources throughout their college careers and throughout their lives. Getting students to think self-consciously earlier in the source evaluation process is perhaps the best way to open up other healthy research strategies.

Finally, it’s important to note that researchers did observe some students demonstrating a self-awareness of their evaluation shortfalls as they described their methods. Some students were surprisingly honest, such as this participant explaining how their bias affected their rating: ‘I’m sure I do judge [this source] slightly based on whether I agree with the opinion…because if I disagree with it, I’m trying harder to find faults in it.’ Given the budding self-awareness of university students, IL librarians can leverage this maturation into more proactive and healthy source evaluation behaviours.

Overall, IL librarians thinking critically about how students are actually evaluating sources and aligning their lesson plans with actual behaviours will allow for better learning outcomes and a more informed, critical citisenry.

References

Brand-Gruwel, S., Kammerer, Y., Meeuwen, L., & Gog, T. (2017). Source evaluation of domain experts and novices during web search. *Journal of Computer Assisted Learning, 33(3)*, 234–251. https://doi.org/10.1111/jcal.12162

Chinn, C., & Rinehart, R. (2016). Commentary: Advances in research on sourcing-source credibility and reliable processes for producing knowledge claims. *Reading & Writing*, *29(8)*, 1701–1717. https://doi.org/10.1007/s11145-016-9675-3

Clawson, L. (2017, April 24). There's a growing crisis in care for disabled and elderly people. Oh, and it's a jobs crisis, too. *Daily Kos*. Retrieved from https://www.dailykos.com/stories/2017/4/24/1655792/-There-s-a-growing-crisis-in-care-for-disabled-and-elderly-people-Oh-and-it-s-a-jobs-crisis-too

Connaway, L. S., Dickey, T. J., & Radford, M. L. (2011). "If it is too inconvenient I'm not going after it:" Convenience as a critical factor in information-seeking behaviours. *Library & Information Science Research*, *33(3)*, 179–190. https://doi.org/10.1016/j.lisr.2010.12.002

Davenport, C. (2017, March 30). Elon Musk's SpaceX makes history by launching a 'flight-proven' rocket. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2017/03/30/elon-musks-spacex-makes-history-by-launching-a-flight-proven-rocket/?noredirect=on&utm_term=.f7fd63296ca6

Dubicki, E. (2015). Writing a research paper: Students explain their process. *Reference Services Review, 43(4)*, 673–688. https://doi.org/10.1108/RSR-07-2015-0036

Faix, A. (2014). Assisting students to identify sources: An investigation*. Library Journal, 63(8–9)*, 624–636. https://doi.org/10.1108/LR-07-2013-0100

Flanagin, A. J., & Metzger, M. J. (2008). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. *Digital Media, Youth, and Credibility,* 5–27. 10.2196/jmir.3991

Harris, F. J. (2008). Challenges to teaching credibility assessment in contemporary schooling. *Digital Media, Youth, and Credibility,* 155–179. 10.1162/dmal.9780262562324.155

Haskins, J. (2017, April 2). Global warming fail: Study finds melting sea ice is actually helping Arctic animals. *The Blaze*. Retrieved from https://www.theblaze.com/news/2017/04/02/global-warming-fail-study-finds-melting-sea-ice-is-actually-helping-arctic-animals

Hersher, R. (2017, March 16). Over-the-counter birth control pills would be safe for teens, researchers say. *NPR*. Retrieved from https://www.npr.org/sections/health-shots/2017/03/16/520268440/over-the-counter-birth-control-pills-would-be-safe-for-teens-researchers-say

Johnson, F., Rowley, J., & Sbaffi, L. (2015a). Modelling trust formation in health information contexts. *Journal of Information Science*, *41*(4), 415–429. https://doi.org/10.1177/0165551515577914

Johnson, F., Rowley, J., & Sbaffi, L. (2015b). Exploring information interactions in the context of Google. *Journal of the Association for Information Science and Technology*, *67(4)*, 824–840. https://doi.org/10.1002/asi.23443

Johnson, F., Sbaffi, L., & Rowley, J. (2014). Assessing trustworthiness in digital information. In: *International Data and Information Management* (IDIMC) 2014 Conference Proceedings. LISU, pp. 68–76. http://e-space.mmu.ac.uk/id/eprint/6169799

Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58(13)*, 2078–2091. https://doi.org/10.1002/asi.20672

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies, 59(B)*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Ostenson, J. (2014). Reconsidering the checklist in teaching internet source evaluation. *Portal: Libraries and the Academy, 14(1)*, 33–50. https://doi.org/10.1353/pla.2013.0045

Pickard, A. J., Shenton, A. K., & Johnson, A. (2014). Young people and the evaluation of information on the world wide web: Principles, practice and beliefs. *Journal of Librarianship and Information Science, 46(1)*, 3–20. https://doi.org/10.1177/0961000612467813

Rowley, J., Johnson, F., & Sbaffi, L. (2015). Students' trust judgements in online health information seeking. *Health Informatics Journal*, *21(4)*, 316–327. https://doi.org/10.1177/1460458214546772

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004). Trust and mistrust of online health sites. *Proceedings of CHI 2004*. New York: ACM Press, 663–670. https://doi.org/10.1145/985692.985776

Stanford History Education Group (2016). Evaluating information: The cornerstone of civic online reasoning. Robert R. McCormick Foundation. 1–29. Retrieved from https://stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf

Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, *31(2)*, 176–203. https://doi.org/10.1080/07370008.2013.769994

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital Media, Youth, and Credibility,* 73–100. 10.1162/dmal.9780262562324.073

Taylor, A., & Dalal, H. A. (2014). Information literacy standards and the World Wide Web: results from a student survey on evaluation of Internet information sources. *Information Research: An International Electronic Journal*, *19(4)*. Retreived from https://eric.ed.gov/?id=EJ10504755

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359(6380)*, 1146–1151. https://doi.org/10.1126/science.aap9559

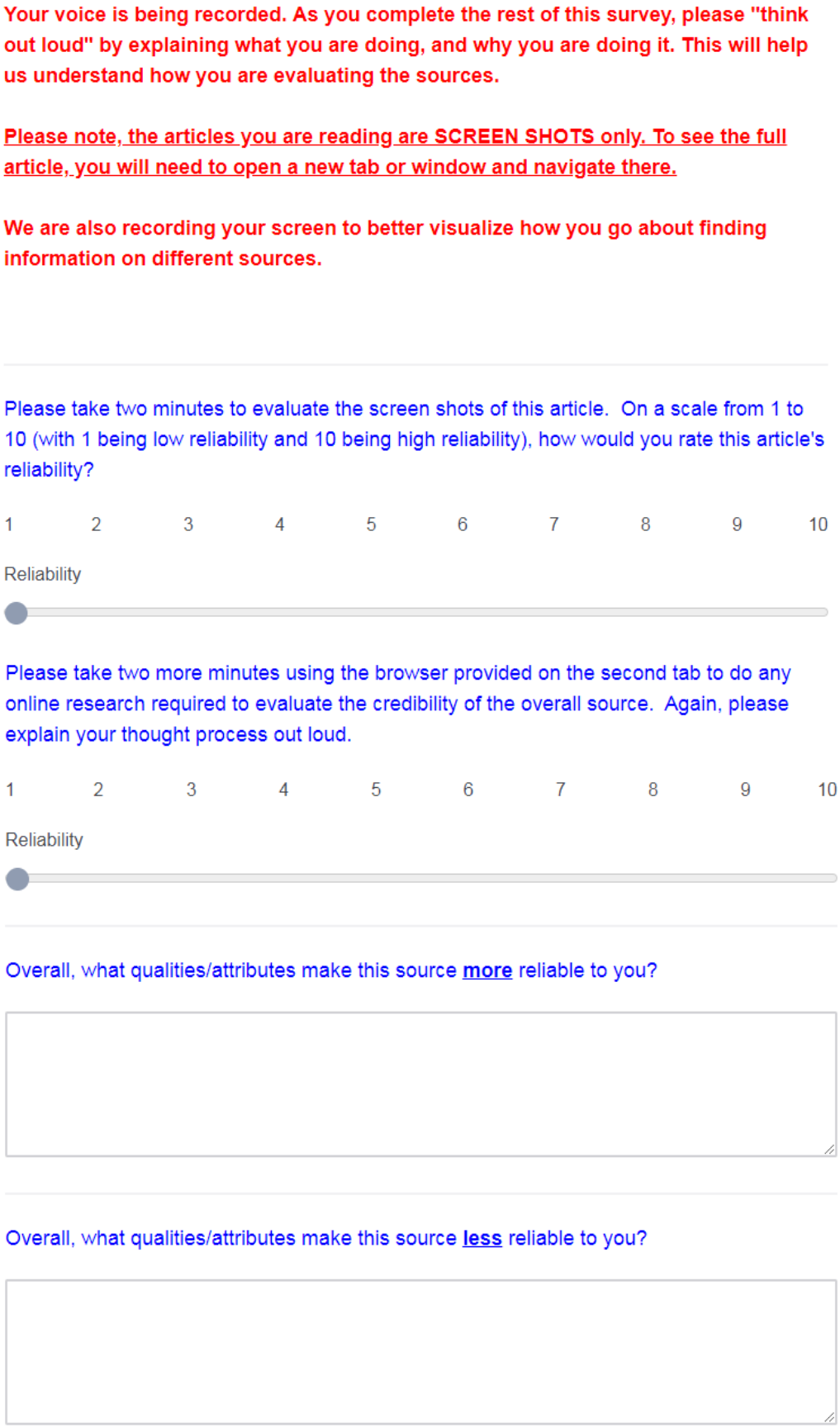
van Strien, J. L. H, Kammerer, Y., Brand-Gruwel, S., & Boshuizen, H. P. A. (2016). How attitude strength biases information processing and evaluation on the web. *Computers in Human Behaviour, 60*, 245–252. https://doi.org/10.1016/j.chb.2016.02.057

von der Mühlen, S., Richter, T., Schmid, S., Schmidt, E. M., & Berthold, K. (2016). The use of source-related strategies in evaluating multiple psychology texts: a student–scientist comparison. *Reading and Writing*, *29(8)*, 1677–1698. https://doi.org/10.1007/s11145-015-9601-0

Wallace, P. (2017, March 19). A third way for universities. *Huffington Post*. Retrieved from https://www.huffingtonpost.com/entry/a-third-way-for-universities_us_58ce8a71e4b07112b6472ec44

Appendix I

Figure 5: Test instructions



Your voice is being recorded. As you complete the rest of this survey, please "think out loud" by explaining what you are doing, and why you are doing it. This will help us understand how you are evaluating the sources.

Please note, the articles you are reading are SCREEN SHOTS only. To see the full article, you will need to open a new tab or window and navigate there.

We are also recording your screen to better visualize how you go about finding information on different sources.

Please take two minutes to evaluate the screen shots of this article. On a scale from 1 to 10 (with 1 being low reliability and 10 being high reliability), how would you rate this article's reliability?

1 2 3 4 5 6 7 8 9 10

Reliability

Please take two more minutes using the browser provided on the second tab to do any online research required to evaluate the credibility of the overall source. Again, please explain your thought process out loud.

1 2 3 4 5 6 7 8 9 10

Reliability

Overall, what qualities/attributes make this source **more** reliable to you?

Overall, what qualities/attributes make this source **less** reliable to you?

Figure 6: Self-reported research behaviours