Critical Appraisal of Information on the Web in Practice: Undergraduate Students’ Knowledge, Reported Use, and Behaviour

Évaluation critique de l’information sur la toile : une vision pratique : les connaissances des étudiants de premier cycle, leur utilisation déclarée et leur comportement

Neil Hogan, University of Alberta
Connie Varnhagen, University of Alberta

Abstract
Undergraduates use a wide range of information resources for academic and nonacademic purposes, including web sites that range from credible, peer reviewed, online journal sites, to biased and inaccurate promotional web sites. Students are taught basic critical appraisal skills, but do they apply these skills to make decisions about information in different web sites? In an experimental setting, undergraduate students examined pairs of web sites containing conflicting information based on different aspects of critical appraisal, namely credibility of the author of the information, purpose of the web site, and last update of the site, and answered multiple choice questions about the conflicting information. Results indicated that students failed to use critical appraisal criteria, and that while knowledge of and self-reported use of these criteria were related to each other, they were not related to behaviour. This research demonstrates the need for alternative strategies for critical appraisal instruction and assessment.

Résumé
Les étudiants de premier cycle consultent une vaste gamme de sources d’information à des fins universitaires et non universitaires, y compris des sites Web allant de revues en ligne crédibles et évaluées par des pairs à des sites Web promotionnels partials et inexacts. On enseigne aux étudiants des méthodes de base d’évaluation critique, mais mettent-ils ces méthodes en pratique pour prendre des décisions relativement à l’information tirée de différents sites Web? Dans un cadre expérimental, les étudiants de premier cycle ont étudié des paires de sites Web contenant des informations contradictoires en se fondant sur différents aspects de l’évaluation critique, notamment la crédibilité de l’auteur de l’information, l’intention du site Web et la dernière mise
à jour du site, et ont répondu à des questions à choix multiples concernant les informations contradictoires. Les résultats indiquent que les étudiants n’ont pas utilisé les critères d’évaluation critique et que si les connaissances et l’utilisation de ces connaissances déclarée par les étudiants étaient reliées, cette relation ne correspondait toutefois pas au comportement observé. Cette recherche démontre la nécessité de stratégies de rechange en matière d’enseignement de l’évaluation critique et son évaluation.

**Introduction**

Undergraduate students surf the World Wide Web on a daily basis, both for academic and for non-academic purposes. Because the information on the web is seemingly limitless and largely unfiltered, students must decide which sources of information they will use and how they will use the information. These decisions may simply involve relying on the first resource encountered in a web search (Bar-Ilan, Keenoy, Levene, & Yaari, 2009; Biddix, Chung, & Park, 2011; Jansen, Spink, Bateman, & Saracevic, 2000; Pan, et al., 2007) or they may involve any number of other criteria based on critical appraisal of the information contained within the resources (Idling, Crosby, Auernheimer, & Klemm, 2009; Varnhagen, 2002; Walraven, Brand-Gruwel, & Boshuizen, 2009). The information literacy skills of students finding, appraising, and citing information from the web for academic purposes are of significant interest and, potentially, significant concern (Biddix et al, 2011; Idling et al., 2009; Metzger, Flanagin, & Zwarun, 2003; Thompson, 2003; Walraven et al, 2009).

Critical appraisal of information, regardless of medium, requires evaluating the credibility of the author and/or source of the information, identifying the purpose of the resource, and appraising the currency, accuracy, and comprehensiveness of the information (Idling et al, 2009; Varnhagen, 2002; Varnhagen, McFall, Savage, & Welker, 2011; Walraven et al, 2009). Published journal articles are perceived as more credible sources of academic information than are web resources because journal articles have undergone a process of peer review (Hilligoss & Rieh, 2008; Rieh & Belkin, 2000). Students need to be more vigilant in critically appraising web-based informational resources than they do peer reviewed journal articles because the peer review process serves as a filter for less credible information. Evaluating credibility of the author and/or source of web-based information requires determining the credentials of the author (e.g., an author who conducts peer reviewed research in the topical area is more credible than a lay person blogging about the topic). Identifying the purpose of the resource is important for determining bias (e.g., a web site promoting some product or service is more likely to be biased and therefore less credible than an encyclopedia article on the same topic). Appraising the quality of the information is more difficult than assessing the author/source and the purpose of the information. For example, to evaluate accuracy and comprehensiveness, the reader may need to refer to other sources of information (e.g., additional articles, an encyclopedia, an expert in the field or professor, etc.) or simply ask whether his or her own questions have been answered by the source. Although determining the currency of the information is relatively straightforward (e.g. by searching for the last update date or checking for recent references to other resources), the critical reader must have some knowledge about the topic and must systematically confirm the veracity of the information with additional resources. Students may also be encouraged to consider the domain of a web site (e.g. “.gov” vs. “.com” or “.org”) when evaluating the quality of an online information resource.
A number of early studies have found undergraduate students’ critical appraisal skills for web-based information to be lacking. Metzger et al. (2003) found that students relied heavily on the web for academic information but they rarely evaluated the accuracy or credibility of the information. In surveys of post secondary students in 1999, 2000, and 2001, Metzger and Flanagan (Flanagan & Metzger, 2000, 2007; Metzger, 2007) found that students “rarely” or “occasionally” applied critical appraisal criteria to evaluate information they found on the web; the students were most likely to report performing relatively effortless appraisals of information currency and completeness. Mittermeyer (2005) found that only one-quarter of first year university students could even correctly identify the correct criteria for critically appraising informational resources.

In response to research such as this and anecdotal evidence that undergraduate students cite non-credible web resources in academic papers, researchers and educators have attempted to develop checklists or guidelines to help in the appraisal of information (e.g., Barker & Kupersmith, 2010; Varnhagen, 2002). However, given that students do not tend to perform exhaustive appraisal of web sites, these criteria may be too cumbersome to be practical (Flanagan & Metzger, 2000, 2007; Meola, 2004; Metzger, 2007; Metzger et al, 2003). Others have developed instructional methods for teaching critical appraisal and other information literacy skills (e.g, Hepworth & Walton 2009; Metzger, 2007; Polkinghorne & Wilton, 2010; Varnhagen, et al, 2004; Varnhagen, 2005; Warschauer, 2007). Some of these projects have demonstrated gains in critical appraisal knowledge as a result of the intervention (Julien, Detlor, Serenko, Willson, & Lavallee, 2009; Meola, 2004; Metzger, 2007; Samson, 2010; Varnhagen, 2005).

Critical appraisal knowledge is not the same as critical appraisal behaviour, however. Students may know how to find scholarly resources or how to critically appraise the author and credibility of a web site, but they may not put this knowledge into action (Metzger et al., 2003; Metzger, 2007). While the current study includes an assessment of undergraduate students’ explicit knowledge of critical appraisal criteria, more importantly, it also evaluated if knowledge of these criteria actually influences students’ decision making in a practical situation (i.e. when they have not been told to use them). The main research questions in this study were: (a) are students making decisions about web-based information based on critical appraisal, (b) is there an association between students’ knowledge and reported use of critical appraisal skills, and (c) how do knowledge and reported use relate to actual critical appraisal behaviour? We also asked the students additional questions about their prior knowledge and interest in particular subject areas, to control for their potentially confounding effects.

To address the first, practical question, we selected and modified extant web sites to contain conflicting information. Although all sites were reputable, based on critical appraisal criteria, one web site was more credible than the other in each pair. We modified three pairs of web sites each for three different criteria, namely credibility of the author/source, purpose of the site, and currency of the site. For the credibility of the author/host condition, the more credible informational sites were selected from well-known associations or internationally known authorities on the asked-for information. The less credible informational sites were selected from other associations or authorities that are less well known with respect to the asked-for information. For the purpose of the site condition, we selected informational web sites as our more credible web sites and persuasive and/or testimonial web sites as our less credible web sites. For the currency of the site condition, we selected or modified sites with a recent “last
updated” statement as our more credible sites and sites with no update or copyright date as our less credible sites.

We embedded discrepant information into approximately similar locations within each site in the pair. Students were then asked to examine the pairs of sites and find information to answer a multiple choice question about that information with three options, namely the information from the more credible site, information from the less credible site, and a “not sure” option. We hypothesized that, if students are applying critical appraisal criteria, they would respond with the option corresponding to the more credible site and not to the option corresponding to the less credible site or “not sure.”

We also had a control condition where the sites were identical in credibility, purpose, and currency, but only one of the sites contained the asked-for information. This allowed us to rule out guessing and random responding.

Figure 1. Screen capture of the Canadian Mental Health Association web site. The third bullet point contains the asked-for information.
The following example was used under the credibility condition and is representative of the nature of our questions. The two web sites that students were asked to consider are shown in Figures 1 and 2. Figure 1 shows an excerpt of the first page of the Canadian Mental Health Association with mental health prevalence statistics. Figure 2 shows an excerpt from the first page of deal.org, a site developed by the Youth Engagement Section of the Royal Canadian Mounted Police to empower youth and prevent crime. Although both are reputable hosts, the Canadian Mental Health Association is a more credible source of statistics on mental illness than is the Royal Canadian Mounted Police. The Canadian Mental Health Association (CMHA) site indicates, “In Ontario for example, one in six kids and teens (aged 4-16 years) suffers from a mental illness.” We modified the article in the deal.org ezine to incorrectly state, “In Ontario alone, about one in five 4 to 16 year olds suffer from some type of psychiatric disorder.” Otherwise, the web pages were left intact. The students were asked to answer the question, “How many kids in Ontario suffer from mental illness/psychiatric disorder?” The options were: “1 in 6” (the statistic provided by the CMHA site), “1 in 5” (the statistic provided by deal.org), and “Not sure. Possibly somewhere between 1 in 5 and 1 in 6.” Based on our hypothesis, students using their critical appraisal knowledge should answer the question with the multiple choice option, “1 in 6.” Two more examples of pairs of web sites that differed in credibility are: (1) the official web site of the American Kennel Club and an unverified online blog, both discussing the numbers of a particular dog breed in America and (2) the web site of a provincial motor association and an individual’s personal web site, both providing suggestions for winter car care.

To address the second and third questions, we used an existing critical appraisal test (Daniels, 2007). We analyzed the two components of Daniels’s test, namely knowledge of and self-reported use of critical appraisal strategies, to determine if there is a relationship between them. Considering the many attempts that have already been made to increase students’ knowledge of critical appraisal skills, including those described above, undergraduate students today are likely more aware of critical appraisal criteria than when Mittermeyer (2005) conducted his research; are undergraduates today also more likely to report using these criteria than Metzger’s (Flanagan & Metzger, 2000, 2007; Metzger, 2007; Metzger et al., 2003) students were a decade ago? We also compared the two components of Daniels’s (2007) test to performance on the practical task, to determine if they are related to behaviour.

Additionally, Daniels (2007) found that knowledge of critical appraisal, coupled with domain knowledge and motivation to critically appraise, predicted likelihood of critically appraising health-related web sites. Based on these findings, we also asked about knowledge of and interest in each of the topics in the study.
Figure 2. Screen capture of the deal.org web site. The first sentence contains the modified asked-for information.

Method

Participants

University students enrolled in first year introductory psychology courses participated in this study as part of their course requirements. The 44 participants ranged in age from 18 to 24 years old (\( M = 19.5 \) years, \( SD = 1.4 \) years) and in year of studies from 1\(^{st}\) year to 5\(^{th}\) year (46\% 1\(^{st}\) year, 36\% 2\(^{nd}\) year, and 18\% 3\(^{rd}\) to 5\(^{th}\) year). Thirteen of the participants were male (30\%) and 31 were female (70\%).
*Materials*

Materials consisted of demographics and background knowledge/interest scales, web site archives and accompanying experimental questionnaire, and Daniels’s (2007) critical appraisal knowledge and use test.

The participant demographics questions were gender, age, and year in university. The background knowledge/interest scales consisted of 5-point Likert scales for self-rated knowledge (1 = not at all knowledgeable, 5 = very knowledgeable) and interest (1 = not at all interesting, 5 = very interesting) in the topics of the 12 different web resources used in the experiment.

The experimental manipulation consisted of 12 pairs of web sites on different topics (e.g., prevalence of mental illness, melatonin, and electric cars). In some cases, one web site in each pair was the extant web site. The other web site was modified (with permission of the web site author or host whenever possible) to provide different information according to the criteria manipulation (see Figures 1 and 2). In other cases, both web sites were modified to provide information corresponding to the criteria manipulation. We had three pairs of web sites each for the three experimental manipulations, namely, credibility of the author/host, purpose of the site, and currency of the site. For the credibility of the web site condition, we selected two web sites containing similar information and modified the information on the less credible site. For purpose of the web site we selected an informational and a persuasive or commercial web site and modified the information on the persuasive site. For currency of the web site, we selected two web sites and then modified or deleted the last update of one web site and modified the information on the undated site. We also had a control condition in which only one site in the pair contained the requested information. We created three pairs of sites for each of the four conditions.

For each pair of web sites, we created a question regarding the manipulated information. The questions were multiple choice with one answer being the information from the more credible site, the other answer being the information from the less credible site, and a third answer being “not sure.” For the control condition, given that the asked-for information was only addressed on one of the web sites, we created a fictitious answer as one option.

The critical appraisal knowledge and self-reported use test was adapted from Daniels (2007). The test assesses six aspects of critical appraisal, namely how to determine the credibility, purpose, currency, accuracy, objectivity, and comprehensiveness of a web resource. The first part of the questionnaire assesses knowledge about how to appraise these six aspects of critical appraisal. The questions are open ended; the instructions are to write down up to three different ways to appraise each of the six criteria. The number of distinct responses comprised the knowledge score; the maximum potential knowledge score for this part of the test is 18. The second part of the critical appraisal test consists of a 4-point scale of self-reported frequency of use for each of the stated critical appraisal behaviours, with 1 = never use to 4 = use all the time. The maximum score for this part of the test is 4 times the number of ways recorded, to a potential maximum of 72.

*Procedure*

Upon arriving at the experimental session the participants were given a brief introduction to the tasks they were to complete. They were told that the purpose of the study was to evaluate the role
of web site design in students’ abilities to find information on web sites, since a goal of the study was to monitor their behaviour in the absence of explicit instructions to critically appraise web sites. Next, the participants completed the demographic and topic knowledge/interest questions. Participants were given the instructions and experimental questionnaire and were assigned to individual computers, in private rooms, without an active internet connection. The web site archives were preloaded onto these computers. For each pair of web sites, the participant opened the web sites and answered the associated question. After completing this second stage of the study, the participants completed the critical appraisal knowledge test.

Results

To address the first research question, regarding the relationship between students’ behaviour and critical appraisal criteria, we examined correctness on the practical task as a function of critical appraisal criterion manipulation. We summed the number correct for each type of critical appraisal manipulation for our analysis of differences between the four critical appraisal manipulation conditions. The mean number correct (out of a maximum of 3) is shown in Figure 3 for the different critical appraisal conditions. We analysed these differences in a repeated measures analysis of variance with one repeated factor (the four critical appraisal manipulations) and found a significant effect of critical appraisal manipulation, $F(3, 129) = 26.41$, $p < 0.001$, partial $\eta^2 = 0.39$ (large effect according to Cohen, 1988). Tukey’s honestly significant difference post hoc analyses revealed significantly more correct responses for the control compared with all experimental manipulations and significantly fewer responses for the purpose manipulation as compared with the credibility manipulation, HSD = 0.43, $p < 0.05$.

We also examined students’ incorrect responses. We calculated the proportion of “not sure” responses of incorrect responses as a function of critical appraisal criterion. This measure provides an indirect measure of students’ confusion about critical appraisal. Because half of the students correctly chose the correct information for all three examples in at least one of the criteria conditions (almost always the control condition), we could not analyze the proportion of “not sure” responses using analysis of variance. Instead, we calculated one-sample $t$-tests, comparing the proportion of “not sure” responses to guessing between “not sure” and the incorrect answer. We found statistically significant effects for the purpose condition, $M = 0.28$ ($SE = .06$), $t(43) = -3.51$, $p < 0.005$, and for the credibility condition, $M = 0.34$ ($SE = 0.6$), $t(38) = -2.44$, $p < 0.05$. These results indicate that, for the purpose and credibility conditions, students were more likely to select the incorrect information than they were to answer that they were not sure which information was correct. The mean proportion of the date condition was $M = 0.48$ ($N = 43$, $SE = 0.7$), indicating that participants were equally likely to select “not sure” as they were the incorrect information.

To examine the second research question, regarding the relationship between critical appraisal knowledge and use, we examined the correlation between participants' scores on the critical appraisal knowledge test and their self-reported use of the critical appraisal criteria they identified. The mean score on the critical appraisal knowledge test was $M = 9.6$ ($SD = 3.0$), with a range of 3 to 15 (max = 18). The mean score on self-reported critical appraisal use was 19.9 ($SD = 7.4$), with a range of 6 to 37 (potential max = 72). We found a significant Pearson product moment correlation between critical appraisal knowledge and self-reported critical appraisal use, $r(42) = 0.71$, $p < 0.001$. Examining relationships between critical appraisal knowledge and self-
reported use and participant age, sex, and year of university, we found no significant correlations between these two scores and demographic characteristics of the participants (with all but one correlation less than 0.20).

![Figure 3](image.png)

**Figure 3.** Means (and standard error bars) for correct answers to the questions for the different conditions. The maximum score is 3.

To address the third research question, we examined correlations between general critical appraisal (knowledge and use) and scores on the information identification test. We found no significant correlations between critical appraisal knowledge or self-reported use and correct identification of the information on any of the web sites, all $r < 0.25$.

Regarding the relationship between participants' behaviour and their knowledge of and interest in the topics, we first examined the relationship between reported knowledge of and interest in the different web sites and correct identification of the information on those sites. Mean self-reported knowledge about each of the topics ranged from $M = 1.2$ ($SD = 0.8$) to $M = 2.9$ ($SD = 1.2$). Mean self-reported interest in each of the topics ranged from $M = 1.3$ ($SD = 0.6$) to $M = 2.7$ ($SD = 1.2$). Correlations between knowledge and interest were significant for all web sites, $r(42) = 0.34 - 0.79$ (with most $r > 0.50$), $p < 0.005$. Correlations between participants' self-reported knowledge or interest in the topic and their responses on the information identification test revealed no significant correlations, all $r < 0.34$.

**Discussion**

Despite the modest sample size, this study produced strong results. Although the university students in this study demonstrated good knowledge and reported high levels of use of critical appraisal criteria, their self-report data was not consistent with their behaviour. Compared with a control condition, where participants identified the correct information over 75% of the time, participants were able to identify the correct information using critical appraisal criteria only
around 30% of the time. Supporting these findings, when students could not identify the correct information, they were more likely to identify the incorrect information in the purpose and credibility critical appraisal conditions as opposed to indicating that they were not sure about the correct information. These results indicate that university students are not using basic criteria, such as date of publication or last update, purpose of the web site, or credibility of the source, to critically appraise web sites. While we did not ask the students directly how or why they chose the answers that they did, this would undoubtedly be interesting, and may be pursued in future studies. Nonetheless, this apparent lack of applied critical appraisal skills is potentially a major problem, considering that the majority of information found on the World Wide Web has not undergone a scientific peer review process. Our experimental findings are also consistent with older survey research indicating that undergraduate students either do not know or do not generally use critical appraisal criteria when searching the web for information (Flanagan & Metzger, 2000, 2007; Metzger, 2007; Mittermeyer, 2005).

Our findings indicate that we need more discriminating assessment of critical appraisal knowledge and skills. Instructors and librarians have traditionally assessed students’ abilities to list critical appraisal criteria or provide self-reports of use as indicators of information literacy (Flanagan & Metzger, 2000, 2007; Metzger, 2007; Mittermeyer, 2005; Project SAILS, 2000-2011). Based on measures such as these, students in our study demonstrated good knowledge of critical appraisal and good intentions to use their knowledge of critical appraisal when evaluating web sites. However, the students’ knowledge and intentions did not translate into behaviour. Thus, our experimental approach indicates that more traditional paper-and-pencil tests are not adequate indicators of real-world performance. More direct measures of critical appraisal skill and knowledge need to be developed to assess the outcomes of information literacy instruction.

There are a number of possible explanations for students’ failure to use critical appraisal skills in practice. For example, it is possible that motivation is essential for engaging in critical appraisal. Daniels’s (2007) study of undergraduates found that evaluation of a web resource tends to be done quickly and that motivation is one of the factors that can determine whether a consumer will critically appraise a web resource. In a study of high school students, Harouni (2009) also found that motivation and interest in the subject matter were necessary for students to engage in critical appraisal. Metzger (2007) developed a dual process model of web site appraisal that starts with motivation to actually engage in critical appraisal. Others, such as Keller (e.g. Visser & Killer, 1990) have developed methods of increasing motivation for other educational purposes, which may prove useful in increasing students’ use of critical appraisal skills. However, we assessed knowledge and interest in the different topics and found no correlation between these components of motivation and students’ actual appraisal behaviour, which indicates that, while motivation may be necessary, it is not sufficient for critical appraisal.

Our finding that students made significantly more correct responses to the control condition indicates that students can search for and find information on a web site. They were also better able to identify correct information as a function of general credibility of the author/host of the web site than the purpose of the web site, with currency of the web site falling somewhere in between. This suggests that students are engaging in some critical appraisal. They can determine, for example, that the Canadian Mental Health Association is a more credible source of information about mental illness than an ezine from another credible – though not necessarily an authoritative – source. Students are particularly poor at determining credibility based on more
subtle aspects such as whether the purpose of the web site is to inform or promote. Evaluating the content of the web site is more complicated and requires greater processing and validation of the information than does surface appraisal of the source (Metzger, 2007).

Clearly, continued efforts to educate university students about information literacy and the importance of critical appraisal are needed. University and library web sites include lists of critical appraisal criteria and links to other, similar lists of criteria (e.g., Barker & Kupersmith, 2010; http://www.vuw.ac.nz/staff/alastair_smith/evaln/evaln.htm). Simply providing students with the tools is not enough, however; students need to be guided in bridging academic knowledge and practical application (Julien et al, 2009; Varnhagen et al, 2004). Metzger (2007) suggested that credibility rating systems and expert-vetted web site directories could be developed to provide credible information to students. Perhaps, instead of leaving critical appraisal to experts as Metzger suggested, new trends in social media can be harnessed to encourage critical appraisal. Kim (2010) interviewed participants in Yahoo! Answers, a social question and answer forum, and found that, while users described using a wide range of critical appraisal criteria in evaluating the different answers provided to their questions, they also relied on answer ratings in their evaluation. Users of Yahoo! Answers award points that reflect quality of answers much like commercial sites allow customers to provide reviews of products. Browser add-ons, such as Xmarks allow for user reviews of web sites. Future research is needed into browser tools such as these that harness social approaches for critical appraisal of web-based information (Metzger, Flanagan, & Medders, 2010).

Along with providing new methods of assessment and developing additional and socially-mediated instruction for critical appraisal skills, we may need to design instructional environments that allow students greater freedom to develop critical appraisal skills. As McDowell (2002) pointed out, lecturers tend to assign directed readings. While some instructors do give assignments that require students to consult the literature for appropriate resources, consulting the assigned texts and the instructor’s lectures are still the main requirements of many undergraduate courses. In the interest of training students to actively critically appraise information sources consistently, instructors may consider increasing the number or proportion of assignments that require students to find information resources themselves.

In conclusion, while the participants in this study revealed that they possess at least superficial knowledge of critical appraisal skills and indicated that they use these skills to evaluate informational web sites, they demonstrated a disconnect between knowledge and behaviour. The results of this study have revealed a need to improve instruction to help and encourage students to make better sense of the web.

Acknowledgements

Acknowledgements are due to the students of the Applied Development and Instructional Technology Lab at the University of Alberta for their involvement - particularly Gabriela Corabian, Kristen Welker, and Clement Wong, whose assistance was instrumental in the successful completion of the experimental sessions.
References


**Authors**

Neil Hogan. University of Alberta. Neil Hogan is a graduate student in Clinical Psychology at the University of Saskatchewan. As an undergraduate psychology student at the University of Alberta, he spent three years as part of the Applied Development and Instructional Technology Lab under the supervision of Dr. Varnhagen. Email: nhogan@ualberta.ca

Connie Varnhagen, University of Alberta. Connie K. Varnhagen is a Professor in the Department of Psychology and the Academic Director of the Undergraduate Research Initiative at the University of Alberta. She studies educational technology and development of critical appraisal of web-based information. Email: varn@ualberta.ca

This work is licensed under a Creative Commons Attribution 3.0 License.