Advances in Social Sciences Research Journal – Vol. 9, No. 7
Publication Date: July 25, 2022
DOI:10.14738/assrj.97.12670.
Muis, K. R., Denton, C. A., & Dube, A. (2022). Identifying CRAAP on the Internet: A Source Evaluation Intervention. Advances in Social Sciences Research Journal, 9(7). 239-265.

Identifying CRAAP on the Internet: A Source Evaluation Intervention

Krista Renee Muis
McGill University

Courtney A. Denton
McGill University

Adam Dubé
McGill University

ABSTRACT

Individuals of all ages struggle to determine the reliability of information on the internet. To address this common issue, many educational institutions have endorsed the CRAAP test (i.e., currency, relevance, authority, accuracy, and purpose) as an effective approach to support identification of unreliable information. The present study extended the CRAAP test by incorporating a modeling component on how to evaluate and integrate multiple sources of varying quality on the internet and evaluated the efficacy of this source evaluation training intervention. Eighty-two participants across Canada were recruited to evaluate six authentic webpages and then construct an argument on a specific topic. Half the sample received training to examine the currency, relevance, authority, accuracy, and purpose (i.e., CRAAP) of the webpages before completing the online activity. Results revealed that the intervention group provided more accurate rank-ordering of the webpages, but no differences were found between groups on source integration via an argumentative essay. These findings suggest that the CRAAP test is effective in improving individuals' evaluations of online sources but is not effective in promoting better quality source integration.

Keywords: epistemic cognition, source evaluation, intervention, digital literacy

The purpose of the present study was to examine the efficacy of a source evaluation training intervention implemented in an online context with adults across Canada. The training aimed to improve their source evaluation accuracy and reasoning during an online inquiry task about mental illness. Without instruction, previous research has found that people rely on content- or design-focused criteria to justify their evaluation decisions, including comprehensibility and task relevance criteria (Braasch et al., 2013; Kąkol et al., 2017; Kiili et al., 2008). To account for adults' use of various source evaluation criteria, we adopted the Apt-AIR framework (Barzilai & Chinn, 2018) and CRAAP test (currency, relevance, authority, accuracy, purpose; Blakeslee, 2004; Meriam Library of CSU Chico, 2010) to develop the training. The CRAAP test has been incorporated into numerous library websites across North America, with some informal attempts to extend the tool beyond its original checklist approach (Fielding, 2019). Although many interventions have incorporated aspects of the CRAAP test (e.g., establishing source's

authority, Barzilai, Mor-Hagani, et al., 2020; Mason et al., 2014; Pretorius, 2018), the CRAAP test has not been fully implemented and empirically evaluated, especially through the lens of the Apt-AIR framework. Moreover, critics of the CRAAP test have indicated shortcomings in delivery (Breakstone et al., 2019; Duby, 2018; Fielding, 2019) including a lack of modeling of the specific skills being taught. Given that previous research has demonstrated improved epistemic strategy use when explicitly taught and modeled (Muis & Duffy, 2013), we extended the CRAAP test by incorporating a modeling component during training and assessed whether training on source evaluation also resulted in improved argumentation via an argumentative essay.

THE CRAAP TEST

The CRAAP test is a popular framework for evaluating sources that can be found on library websites throughout North America (Batchelor, 2017; Smallwood, 2015). It typically includes 25 questions or criteria that can be used to assess the Currency, Relevance, Authority, Accuracy, and Purpose of a source (CRAAP; Blakeslee, 2010). The test has been adapted to include other aspects of online source evaluations, such as the presence of advertisements (Green, 2019) and the quality of linked webpages (Queen's University Library, 2020). Taken together, the CRAAP test reflects content-focused and epistemic criteria to assess a source's knowledge claims, author, and argument through vertical (i.e., focusing on information from one source) and lateral (focusing on information from multiple sources) reading.

Currency is used as an indicator of the source's timeliness. The associated criteria relate to the source's publication or revision date, up-to-dateness in relation to the task, and hyperlink functionality (Meriam Library of CSU Chico, 2010). Currency evaluations, in coordination with accuracy assessments (see below), can be leveraged to assess the appropriateness of the source's claims based on the freshness of the data used as evidence (Kąkol et al., 2017) and the quality of the evidence linked to it (Chiagouris et al., 2008). Here, these criteria are placed in relation to other sources. Of particular relevance, assessing currency has been found to be a significant predictor of selecting high-quality information (Barnes et al., 2003).

Relevance criteria also provide useful indicators of the source's importance for the user's task. This category includes task relevance, target audience, and an overall reliability judgment (Kiili et al., 2008). As we elaborate in the next section, relevance assessments contribute task and context cues with which epistemic ideals and processes can be adapted (e.g., type of webpage, Barzilai & Chinn, 2018). Authority criteria are also common features of source evaluation tools or trainings. In the CRAAP test, authority criteria include the author's transparency, their credentials and qualifications, and the URL. To establish reliability, the original CRAAP questions can be used as a starting point to investigate the author and venue's reputation. Beyond expertise, the author's previous work and notoriety can be used to indicate their qualifications to write on a topic. Examination of an author or publishing venue's background also provides insight about potential biases that emerge when examining the source's purpose. The accuracy criteria focus on the truthfulness of the source's content. Accuracy criteria include the presence of evidence and the ability to verify claims and arguments using other sources or prior knowledge. Indicators related to writing quality, style, and peer review process are also included. Evaluations of the source influence how claims and arguments are interpreted (Britt & Aglinskas, 2002; Wineburg, 1991). A shift from a criteria to a process focus would impact the CRAAP test's accuracy category greatest. Here, the trail of information must be followed, with

subsequent source evaluations used to examine the quality of the evidence included, rather than solely checking that evidence is presented. For familiar topics, users may rely on their prior knowledge to verify information, and for unfamiliar topics they may rely on the source's content (Barzilai, Thomm, et al., 2020). In coordination with purpose evalautions, tone may be used as an indicator of potential biases (e.g., labelling bias).

Lastly, the purpose criteria reflect measures of the source's motivations for creating the content. Purpose criteria include categorizing the source's purpose (e.g., sell, inform, persuade), identifying their level of objectivity, and assessing biases. Funding is an indicator of a source's purpose. For example, the presence of advertisements signifies that the venue receives ad revenue, which may compromise the quality of information in favour of clicks. The author's purpose may also be assessed by looking at the perspectives presented in their argument. Taken together, the CRAAP criteria address key aspects of online source evaluations.

Critiques of the CRAAP test have focused on the tool's reliance on information within the source being evaluated (i.e., vertical reading) rather than information from multiple sources (i.e., lateral reading, Breakstone et al., 2018). However, the relevance and accuracy criteria specifically support source integration, an intricate epistemic process where multiple perspectives are weighed and reconciled (Barzilai & Zohar, 2014, 2018). Further, the relevance criteria also encourage users to establish reliability based on the other criteria before making relevance-based judgments. These are all key epistemic processes, which we describe next.

The Apt-AIR Framework

In their Apt-AIR framework, Barzilai and Chinn (2018) delineated the role of epistemic cognition in source evaluations. Epistemic cognition refers to thinking about the acquisition, justification, and use of knowledge (Greene et al., 2018). During source evaluations, epistemic cognition is used to examine the epistemic properties of knowledge claims (e.g., accuracy), sources (e.g., authority) and arguments (e.g., purpose, Barzilai and Zohar, 2014), as seen in the CRAAP test. Further, Barzilai and Chinn (2018) highlighted the role of epistemic ideals and reliable epistemic processes during source evaluations. Epistemic ideals describe the criteria used to justify accepting or rejecting a knowledge claim (Chinn et al., 2011, 2014). A variety of criteria are used to justify reliability assessments with content- or design- focused criteria used most frequently (Bates et al., 2006; Kąkol et al., 2017; Ulyshen et al., 2015).

Epistemic ideals have been spontaneously observed during source evaluations (Gerjets et al., 2011; Mason et al., 2011) with mixed findings about the role epistemic ideals play in accurate source evaluations (Mason et al., 2014, 2018). Under reliable epistemic processes, Barzilai and Chinn (2018) described creative processes as procedures to construct an epistemic product such as a written argument that integrates multiple perspectives (Chinn et al., 2011, 2014). Importantly, more accurate source evaluations have been associated with more complex source integration in the form of written arguments (i.e., two-sided argument, Braasch et al., 2013; Mason et al., 2014). Based on the Apt-AIR framework, each category of the CRAAP test plays a role in source evaluation and integration. As such, one would expect that a training intervention designed to improve accuracy in source evaluation across multiple dimensions should translate into better source integration via an argumentative essay, regardless of the topic chosen.

Argument refers to a product that is constructed in written or oral form by an individual, which consists of a claim and one or more supporting reasons or evidence that are connected to the claim with warrants (Toulmin, 2003). Argumentation refers to the dynamic process that captures what is done to create the argument (Kuhn, Hemberger & Khait, 2016). Argumentation can be captured via dialogic methods (Kuhn, 2018) or via argumentative essay writing (Latifi et al., 2019; Noroozi et al., 2018; Valero Haro et al., 2019). For example, high-quality argumentative essays encompass a clear claim supported by evidence and reason, followed by acknowledgements of counter-arguments against the original claim, and integration of the arguments and counter-arguments which eventually lead to the final conclusion (Nozoori et al., 2016). The goal is to provide strong evidence to support one argument over another by weakening the other position (Kuhn, 2018). This requires appropriate epistemic ideals.

Barzilai and Chinn (2018) argued for a situated view of epistemic cognition where epistemic ideals and processes must be adapted to different domains, contexts, and tasks. To some degree, the CRAAP test accounts for these differences through the relevance criteria, which provides task and context clues to calibrate epistemic cognition. Following Sosa (2015), Barzilai and Chinn (2018) asserted that apt epistemic cognition could be assessed by accurate judgments of information supported by competent use of epistemic ideals. Thus, an apt source evaluation that is accurate and may be based on the appropriate use of content-focused criteria and epistemic ideals. Beyond task and context, researchers have also identified that prior knowledge and attitude toward the topic influence source evaluations (Bråten et al., 2016; Maier & Ritcher, 2017). Specifically, low prior knowledge (E. H. Jung et al., 2016; W. S. Jung et al., 2018) and negative attitudes about a topic (Taylor et al., 2015; Van Strien et al., 2014) often result in inaccurate reliability attributions. Within the CRAAP test, low prior knowledge may impact the user's ability to verify the webpage's accuracy using their knowledge (i.e., knowledge-based validation, Ritcher & Schmid, 2010), and their attitudes may influence their use of the accuracy and purpose criteria (e.g., accept belief-consistent information, Meppelink et al., 2019). Accordingly, to evaluate whether training results in improved source evaluation and integration, it is important to consider prior knowledge and attitudes toward the topic.

Teaching Source Evaluation Skills

Within learning contexts, targeted interventions have demonstrated that students can be trained to use important epistemic ideals related to authority and accuracy during source evaluations (Britt & Aglinskas, 2002; McGrew et al., 2019; Pretorius, 2018). Like the CRAAP test, some interventions have used declarative knowledge or modeling to train students to assess the source's authoritativeness (e.g., author's expertise or purpose) and content accuracy (e.g., plausibility, corroboration, Mason et al., 2014; Wiley et al., 2009; Hämäläinen et al., 2020). For example, Mason and colleagues (2014) investigated the efficacy of SEEK (Source, Evaluation, Explanation, Knowledge) training by asking secondary school students to read sources from an offline digital environment that ranged in authoritativeness and viewpoint, rank the reliability of each website, and provide a justification for their ranking. Students also constructed a written argument in a transfer task. Beyond prior knowledge and argumentative reasoning skills, Mason and colleagues (2014) found that students who received training more accurately rank-ordered the most reliable websites and used SEEK criteria to justify their ranking decisions. Students also demonstrated better source integration in their written essays; specifically, individuals in the intervention group with higher prior knowledge

produced more complex arguments. What is unclear is whether training led to improved argumentative essays or whether superior essays were a function of higher prior knowledge.

Moreover, much of the current understanding of source evaluations has been shaped by studies conducted in offline environments, as seen in Mason and colleagues' work (2011, 2014, 2018). Due to the situated nature of epistemic cognition (Chinn & Sandoval, 2018; Sandoval, 2017), findings in offline environments may not generalize to source evaluations on the internet. Drawing from the Apt-AIR framework, Barzilai, Mor-Hagani, and colleagues (2020) developed an offline learning environment with epistemic scaffolds to support students' source evaluations and integrations. Like Mason and colleagues (2014), the effectiveness of the training was determined through students' ratings of sources, their justifications, and their written essays without scaffolds. After using the scaffolded environment, students who received training provided more epistemic criteria to justify their source rating, produced more two-sided arguments, and reconciled more multiple perspectives compared to those who did not receive training. The researchers classified the training's impact on students' source evaluations as a positive, small effect (Barzilai, Mor-Hagani, et al., 2020). Despite increases in justification based on epistemic ideals, Mason and colleagues (2014) and Barzilai, Mor-Hagani and colleagues (2020) reported that use was low during source evaluations. In contrast, McGrew (2020) observed large improvements in students' justifications after training, documenting a shift from irrelevant justifications to epistemic ideals for nearly half of the participants (46%). Despite these positive effects, to date, few studies measuring epistemic cognition have been conducted in an authentic environment and, to our knowledge, none have explored the general adult population's source evaluation behaviors. Will training result in improved source evaluation, greater use of epistemic ideals, and better argumentative essays?

The Present Study

The purpose of the present research was to examine the efficacy of a source evaluation intervention aimed at improving adults' accuracy and reasoning during online source evaluations. To emphasize the components that describe lateral reading practices, the online CRAAP training included modeling use of each category within a single webpage and across multiple webpages. The following research questions guided the study:

1. Do participants who receive source evaluation training evaluate sources more accurately than those in a control group who receive no training?
2. Do participants who receive source evaluation training use more epistemic ideals when justifying their evaluations than those who do not receive source evaluation training?
3. Do participants who receive source evaluation training produce better quality written arguments on a topic compared to participants who do not receive training?

In accordance with previous research (Mason et al., 2014; Wiley et al., 2009), we hypothesized that adults who receive targeted source evaluation training will more accurately rank-order authentic sources from the internet. We also hypothesized that participants in the intervention group would employ more epistemic ideals to justify their choices (Mason et al., 2014). Lastly, we hypothesized that participants in the intervention condition would construct more complex arguments on the topic compared to those in the control group (Barzilai & Eshet-Alkalai, 2015; Braasch et al., 2013).

METHODS

Participants

We conducted a priori power analysis using G*Power 3 (Faul et al., 2009), which revealed that a minimum of 82 participants would be needed to detect a medium effect size with power (1−β) set at .80 and α set at .05. Two-hundred and two adults from across Canada participated. After reviewing participants' responses, 93 participants were excluded from further analyses because they (a) did not complete the survey, (b) failed at least one attention check, or (c) were under 18 years of age. Of the remaining 109 participants, 41 had been randomly assigned to receive training and 68 to the control group. Following Tabachnick and Fidell's (2013) recommendations for unequal groups, we randomly selected 41 control participants with similar educational backgrounds for further analyses.

These participants (n = 82, 32 female, 1 non-binary) represented a wide range of age groups and educational backgrounds. Participants' ages ranged from eighteen to sixty-five years of age (M = 30.7 years, SD = 10.6). They reported completing secondary (31%), college (9%), undergraduate (45%), and master's (15%) level schooling. Participants self-identified as Caucasian (56%), Asian (29%), Black or African Canadian (6%), Latinx (4%), multiracial (3%), or Indigenous (2%). Most participants reported a political affiliation with the Liberal Party (50%), followed by the Conservative Party (18%), New Democratic Party (17%), among others (15%). All participants volunteered to take part in the research.

Procedure

Ethics was first obtained from the Research Ethics Board at the authors' university. The study was then completed on each participant's personal device using an online survey platform. After providing electronic informed consent, participants completed topic knowledge and attitude measures. Next, the training group received training while participants in the control group directly began the inquiry activity. In the inquiry activity, participants were instructed to take 30 minutes to read and examine each webpage about mental illness in preparation to rank-order the webpages. After examining the webpages, participants rank-ordered the webpages and explained why each webpage was assigned its rank. Next, participants wrote arguments about the relationship between mental illness and violent behaviour. Lastly, participants completed a demographics survey.

Online Training

Participants in the training group watched a 35-minute video that described the CRAAP criteria and modelled criteria use while examining sample webpages' reliability. The topic chosen for the sample activity was medically assisted death regulations in Québec, a common topic addressed in introductory psychology courses. The epistemic aim of the sample activity was understanding (Greene et al., 2014, 2018). To ensure treatment fidelity, participants' time to complete the study was recorded and they were asked to answer three questions about the video's content before moving on to the inquiry task. Participants who completed the study in less than 35 minutes or incorrectly answered one or more of the questions were excluded from analyses. Table 1 provides a description of how each category was addressed in the video.

Materials for the Inquiry Activity

Six authentic webpages about the relations between mental illness and violent behavior were selected and displayed in a Google search results format as the "top six results" using the search

terms "mental illness + violence" (see Figure 1). The webpages held distinctive perspectives about this relationship and provided evidence from sources with varying levels of reliability (see Table 2 for webpage descriptions). To avoid interference related to webpage order, the search engine results page was randomly ordered. Sources listed on the search engine results page appeared with the (1) webpage title, (2) URL linked to the original webpage, and (3) brief description of the content with keywords in bold. Participants were instructed to take 30 minutes to read and examine the webpages in any order, and to use external sources to help them assess each webpage's reliability. For the purpose of rank-ordering the articles based on their reliability, two independent raters evaluated each webpage using the CRAAP test. Their agreement, as measured by Cohen's kappa coefficient, was established at .81, with almost perfect agreement (Landis & Koch, 1977). Despite minor disagreements, both raters ranked the webpages in the same order. Each disagreement was discussed and resolved.

Measures of Epistemic Cognition

Overall Rank-ordering of Webpages

Participants were asked to rank-order the webpages from most to least reliable (Mason et al., 2014; Walvaran et al., 2013). Rankings were compared with the expert rank-order and scored based on their proximity to the correct ranking. For example, ranking the most reliable webpage third received a score of 2 because the assigned ranking was two ranks away from the experts' ranking. Scores were summed to serve as an indicator of participants' accuracy in rank-ordering the webpages with a lower score reflecting a more accurate ranking.

Justification for Webpage Rank-ordering

Participants were also asked to justify their rank-ordering of the two most reliable and two least reliable webpages in a written response (Braasch et al., 2013; Mason et al., 2014). Justifications (47 – 411 words, M = 110 words) were qualitatively analyzed by two independent raters and participants were given one point for each CRAAP criteria used to justify their rank-ordering. The coding scheme was tested using 14 participant responses (17% of sample), and the raters' agreement was established at .83, which is considered excellent (Landis & Koch, 1977). Each disagreement was carefully examined and discussed to update the coding scheme. See Table 3 for selected criteria with illustrative examples. The second author and a trained research assistant coded the remainder of the justification responses. Participants' criteria use was summed for each macro-level category (i.e., currency, relevance, authority, accuracy and purpose) and a separate sum was calculated for epistemic ideal use.

Source Integration in Essays

After completing the inquiry activity, participants were asked to use information from the webpages and their prior knowledge to construct an essay to address the question: Are individuals with mental illness more violent than individuals without mental illness? Participants were instructed to present and justify their position regarding the relationship between mental illness and violent behaviours (47 – 389 words, M = 130 words). Two independent raters qualitatively analyzed participants' argument structure using 15 participant responses (15% of sample). Agreement, as measured by Cohen's kappa coefficient, was established at .76, with substantial agreement (Landis & Koch, 1977). The raters discussed their disagreements and updated the coding scheme to code the remaining responses. Following Chevrier and colleagues (2019), participants' written responses were given a score

between 0 and 5, with 0 representing no argument and 5 representing a two-sided argument with source evaluation(s). Illustrative examples of each response level are presented in Table 4, with the full coding scheme presented in Table 5.

Measures of Individual Differences

To mitigate the influence of individual differences on source evaluations (Bråten et al., 2016; Meppelink et al., 2019; Maier & Ritcher, 2017), participants completed measures of their prior knowledge about mental illness and attitudes toward those with mental illness. These measures were used to assess differences between the intervention and control groups.

Prior Knowledge Test

A knowledge test was developed based on common misconceptions about mental illness to capture participants' topic knowledge before completing the inquiry activity. The test consisted of 16 multiple choice items that covered central topics related to mental illness, including key terminology, statistics, and examples (CMHA, 2018a, 2018b). Correct answers were given a score of 1 and incorrect answers were given a score of 0. Scores were added to create a total sum (out of 16 possible) and were then converted to a percentage, which was used as an indicator of participants' prior knowledge about mental illness. Reliability was borderline acceptable with $\alpha = .66$. The broad range of questions and low prior knowledge may account for the score, which was solely used to assess differences between the groups' prior knowledge.

Attitudes about Individuals with Mental Illness

Participants' attitudes about mental illness were collected using an adapted version of Kobau and colleagues' (2010) instrument, which is designed to assess individuals' attitudes about central topics related to mental illness on two subscales: negative stereotypes, and recovery and outcomes. The validated instrument consists of 11 items to be rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Participants' responses were averaged for both subscales and used as indicators of their attitudes toward individuals with mental illness. Cronbach's alpha reliability scores were poor for both subscales with $\alpha = .60$ for negative stereotypes and $\alpha = .56$ for recovery outcomes. Like the prior knowledge test, these subscales were used solely to assess equivalence in attitudes between the intervention and control groups.

Demographic Survey

Participants were asked to provide information about their educational background and demographic information including their age, sex, political affiliations and online habits.

RESULTS

Preliminary Analyses

Prior to conducting the full analyses, we inspected each variable for skewness, kurtosis, and outliers. Following Tabachnick and Fidell's (2013) recommendations, we used acceptable ranges of |3| for skewness and |8| for kurtosis to assess the relative normality of the distributions. Analyses revealed that the distribution for participants' overall rank-ordering of webpages and their essay scores were normally distributed. However, all macro-level justifications and some micro-level essay scores were positively skewed and leptokurtic. Participants' scores for stating a claim, supporting their claim, and evaluating claims were negatively skewed. Given that these variables represented behavioral frequencies with

meaningful zero points, no transformations were performed. To examine univariate outliers, we converted each variable to a standardized *z*-score. We classified *z*-scores exceeding |3.3| as outliers (Tabachnick & Fidell, 2013). Analyses revealed univariate outliers for four macro-level justifications: currency (n = 2, z = 3.55 to 4.23), relevance (n = 1, z = 3.56), authority (n = 1, z = 3.91), and accuracy (n = 1, z = 4.31). All cases were retained because the values were moderate and represented less than 2% of cases for each variable (Cohen et al., 2013).

Individual characteristics were then compared across the two groups to ensure equivalence in prior knowledge, attitudes, and other demographic characteristics. No statistically detectable differences emerged for prior knowledge about mental illness, attitudes about negative stereotypes, attitudes about recovery and outcomes, education level, political party affiliation, or daily media use. Given the study design, we also confirmed that the training group took longer (M = 73 mins, SD = 31) to complete the study than their control counterparts (M = 50 mins, SD = 25).

Source Evaluation

To assess whether participants in the intervention condition evaluated sources more accurately than participants in the control condition (research question 1), an ANOVA was conducted with the overall scores of the webpage rank-ordering as the dependent variable. Results revealed a statistically detectable difference in overall rank-ordering, $F(1,80) = .4.91$, $p = .03$, $\eta_p^2 = .06$, and in rank-ordering of the most and least reliable webpages, $F(1,80) = 3.87$, $p = .04$, $\eta_p^2 = .05$. Consistent with hypotheses, individuals in the intervention condition were more accurate in their rankings compared to the control group.

For the second research question, we examined whether participants in the intervention condition used more epistemic ideals when making their ranking decisions. A MANOVA was conducted with the macro-level justification scores for the webpage rank-ordering as dependent variables. Results revealed a multivariate effect of condition, Pillai's Trace = .64, $F(5, 76) = 27.34$, $p < .001$, $\eta_p^2 = .64$. Univariate tests revealed statistically detectable differences in participants' use of the currency criteria, $F(1, 80) = 9.58$, $p = .003$, $\eta_p^2 = .11$, and the accuracy criteria, $F(1, 80) = 8.94$, $p = .004$, $\eta_p^2 = .10$. The training group used the currency criteria more often to justify their rank-ordering, whereas the control group employed the accuracy criteria more often than the intervention group. No other differences in epistemic ideals were found. Descriptive statistics for each variable are presented in Table 6.

Source Integration

Whereas participants' justifications served as indicators of their ability to evaluate the trustworthiness of divergent information, their essay responses were used as an indicator of their ability to compare and reconcile information with varying degrees of trustworthiness. Accordingly, to answer the third research question, whether individuals who received source evaluation training produced better quality written arguments on the topic of mental health compared to individuals in the control group, an ANOVA was conducted with participants' total essay score as the dependent variable. Results revealed no statistically detectable difference between the groups' essay scores, $F(1, 80) = 3.41$, $p = .07$, although the difference was trending toward significance with the control group performing slightly better than the intervention group. Descriptive statistics for the essays are presented in Table 7.

DISCUSSION

The present study aimed to evaluate the efficacy of a source evaluation training intervention for adults across Canada. We investigated their evaluations and integrations of information from authentic online sources with varying levels of reliability. Through the lens of the Apt-AIR framework, the CRAAP test encouraged users to reflect on important epistemic ideals, such as author's expertise, adequacy of evidence, and publisher's motivations, and engage in reliable epistemic processes, including knowledge-based verification and corroboration. Although widely endorsed by educators (Batchelor, 2017), the CRAAP test has been criticized for being impractical and providing misleading evaluation results as the tool does not explicitly encourage individuals to leave the webpage they are evaluating (Breakstone et al., 2019; Duby, 2018; Fielding, 2019). By modeling use of the CRAAP test online, training emphasized engagement in both vertical and lateral reading of sources—an extension of the original tool. Although the intervention group provided a more accurate rank-ordering of the webpages using currency compared to the control group, there were no differences found in source integration via quality of argumentative essays between the two groups.

Given that the intervention group was better at ranking the websites compared to the control group, this suggests that individuals in the intervention group improved their source evaluation skills. Consistent with previous research, currency has been found to be a significant predictor of selecting high-quality information (Barnes et al., 2003). This, however, did not translate into improved performance on source integration via the argumentative essay. Although previous research has shown that improved source evaluation leads to superior argumentative essays (Braasch et al., 2013; Mason et al., 2014), this was not the case for our sample.

Moreover, despite being trained to use five categories to evaluate sources, participants in the intervention condition employed only a few CRAAP categories, which they used to justify their assessments of each source. The control group's use of accuracy criteria more frequently than the intervention group suggests that they may have engaged in lateral reading to verify evidence. In contrast, the intervention group's use of currency criteria may indicate that they relied on vertical reading, which is supported by their limited source integrations in the essay. Although the CRAAP test cued users to engage in important epistemic processes, which require lateral reading, the majority of the questions (88%) did not require users to leave the webpage to be answered, as Breakstone and colleagues (2018) argued. Drawing from epistemic cognition research, the CRAAP test can be improved by shifting the focus from checklist items to criteria and interacting processes aimed at evaluating claims, sources, and arguments across websites.

Additionally, previous research has noted that not all epistemic criteria are used spontaneously or are salient in a given situation, like source evaluation and integration (Greene et al., 2010; Hammer & Elby, 2002). It may also be the case that individuals in our study required explicit training and modeling of how to write an argumentative essay. As previous research has shown, undergraduate and graduate students typically perform below the expected level for argumentative essay writing tasks (Kellogg & Whiteford, 2009). Our sample was no different, with an average of just under 50%. Clearly, beyond focusing on source evaluations and epistemic ideals, adults in the general population could benefit from direct instruction on argumentative writing. One method that has been effective in improving individuals's argumentative essay writing is via scaffolding through adaptive fading (Noroozi et al., 2018),

and worked examples and peer feedback (Latifi et al., 2019; Noroozi & Hatami, 2019; Valero Haro et al., 2019).

The findings from the present study and the suggestions for improving the CRAAP tool also support McGrew and colleagues' recent work (McGrew, 2020; McGrew et al., 2019; Wineburg & McGrew, 2017) that explored the lateral reading behaviours of fact checkers. Wineburg and McGrew (2017) noted fact checkers' focus on embedding their evaluation within the larger context of the topic by reading laterally. Unlike previous research (Barzilai, Mor-Hagani, et al, 2020; Mason et al., 2014), however, the control group produced more complex source integrations, often exhibiting multiple perspectives with some degree of claim evaluation or reconciliation. The control groups' use of purpose criteria, which emphasizes objectivity and balanced arguments, may have supported their ability to represent and reconcile multiple perspectives in their essays. Multiple documents research has extensively accounted for the processes that occur during source integrations, including the cognitive and metacognitive aspects of linking sources to content and each other (Salmerón et al., 2018). However, further research is needed to better understand which criteria support online source evaluations and integrations and whether this kind of training does indeed transfer to a more complex task like argumentative essay writing.

Limitations

This study is not without limitations which will aid future research on source evaluations. First, although we embedded indicators of treatment fidelity in the training group's design, study completion time and attention checks may not have ensured that participants watched the training video fully to apply the CRAAP criteria during the inquiry activity. Given the additional time requirement, it is also possible that the training group may have become fatigued from the effort required to complete the tasks. To overcome this limitation, future online tutorials may benefit from an interactive design that allows participants to apply their skills through brief practice questions or a discovery approach, as seen in Pretorius' work (2018).

Second, we did not measure participants' baseline source evaluation and integration skills. Although we established that both groups were similar in age, and educational background, they may have had dissimilar experience with source evaluations. For example, the training group may have been familiar with some aspects of the CRAAP test and been prompted to use the currency and authority criteria after watching the video tutorial. By measuring participants' baseline skills, future research can better assess the impact of this training on participants' behaviours.

Lastly, the webpages selected for evaluation were published by organizations across North America, including the US, which influenced several participants' evaluations. Although only 10% of the sample justified their evaluations based on the publishing country, given the political climate at the time the study was conducted, this factor may have influenced Canadian participants. To minimize this impact, future research with authentic sources should assess the potential impact of political factors on evaluations of sources about socio-scientific issues.

CONCLUSIONS

Access to the internet has changed the way we interact with information and how we should teach important source evaluation skills. Based on the Apt-AIR framework, the CRAAP test

includes assessments of authority and accuracy, which are related to desirable epistemic ideals and processes. This study makes methodological contributions to understanding source evaluations, providing online access to participants rather than a curated offline environment. Further, this study provides evidence of the CRAAP test's shortcomings. This widely referenced evaluation tool requires explicit instruction to embed source evaluations within a larger context where multiple sources and perspectives are reconciled to establish reliability.

References

Barnes, M. D., Penrod, C., Neiger, B. L., Merrill, R. M., Thackeray, R., Eggett, D. L., & Thomas, E. (2003). Measuring the relevance of evaluation criteria among health information seekers on the Internet. *Journal of Health Psychology*, *8*(1), 71–82. https://doi.org/10.1177/1359105303008001436

Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences*, *27*(3), 353–389. https://doi.org/10.1080/10508406.2017.1392968

Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learning and Instruction*, *36*, 86–103. https://doi.org/10.1016/j.learninstruc.2014.12.003

Barzilai, S., Mor-Hagani, S., Zohar, A. R., Shlomi-Elooz, T., & Ben-Yishai, R. (2020). Making sources visible: Promoting multiple document literacy with digital epistemic scaffolds. *Computers & Education*, *157*, 103980. https://doi.org/10.1016/j.compedu.2020.103980

Barzilai, S., Thomm, E., & Shlomi-Elooz, T. (2020). Dealing with disagreement: The roles of topic familiarity and disagreement explanation in evaluation of conflicting expert claims and sources. *Learning and Instruction*, *69*, 101367. https://doi.org/10.1016/j.learninstruc.2020.101367

Barzilai, S., & Zohar, A. (2014). Reconsidering personal epistemology as metacognition: A multifaceted approach to the analysis of epistemic thinking. *Educational Psychologist*, *49*(1), 13–35. https://doi.org/10.1080/00461520.2013.863265

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review*, 1–27.

Batchelor, O. (2017). Getting out the truth: The role of libraries in the fight against fake news. *Reference Services Review*, 143-148.

Bates, B. R., Romina, S., Ahmed, R., & Hopson, D. (2006). The effect of source credibility on consumers' perceptions of the quality of health information on the Internet. *Medical Informatics and the Internet in Medicine*, *31*(1), 45–52. https://doi.org/10.1080/14639230600552601

Blakeslee, S. (2004). The CRAAP test. *Loex Quarterly*, *31*(3), 4.

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, *38*(3), 180–195. https://doi.org/10.1016/j.cedpsych.2013.03.003

Bråten, I., Strømsø, H. I., & Andreassen, R. (2016). Sourcing in professional education: Do text factors make any difference? *Reading and Writing*, *29*(8), 1599–1628.

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28.

Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *Phi Delta Kappan*, *99*(6), 27–32.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*(4), 485–522.

Chevrier, M., Muis., K. R., Pekrun, R., Sinatra, G., & Denton, C. A. (2019). *Thinking critically about socio-scientific issues: The roles of epistemic cognition and epistemic emotions* [Unpublished manuscript]. Department of Educational and Counselling Psychology, McGill University.

Chiagouris, L., Long, M. M., & Plank, R. E. (2008). The consumption of online news: The relationship of attitudes toward the site and credibility. *Journal of Internet Commerce*, *7*(4), 528–549. https://doi.org/10.1080/15332860802507396

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, *46*(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D. N. Rapp & J. L.G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 425–453). MIT Press.

Chinn, C., & Sandoval, W. A. (2018). Epistemic cognition and epistemic development. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Resiman (Eds.), *International handbook of the learning sciences* (pp. 46–55). Routledge.

Canadian Mental Health Association (CMHA). (2018a). *Stigma and discrimination*. Retrieved from https://ontario.cmha.ca/documents/stigma-and-discrimination/

CMHA. (2018b). *Myths about mental illness*. Retrieved from: https://cmha.ca/documents/myths-about-mental-illness

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Cresswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage Publications.

Cresswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Denton, C. A., Muis, K.R., Armstrong, S., & Dubé, A. (in preparation). En-garde: Source evaluations in the digital age. *Contemporary Educational Psychology*.

Duby, H. R. (2018). What a load of CRAAP: Evaluating information in an era of "fake news". *Tennessee Libraries*, *68*(4).

Eshet-Alkalai, Y., & Chajut, E. (2009). Changes over time in digital literacy. *CyberPsychology & Behavior*, *12*(6), 713–715.

Eshet-Alkalai, Y., & Chajut, E. (2010). You can teach old dogs new tricks: The factors that affect changes over time in digital literacy. *Journal of Information Technology Education: Research*, *9*(1), 173–181.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fielding, J. A. (2019). Rethinking CRAAP: Getting students thinking like fact-checkers in evaluating web sources. *College & Research Libraries News*, *80*(11), 620.

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*(2), 220–231. https://doi.org/10.1016/j.learninstruc.2010.02.005

Green, J. (2019, Jan 22). Check yourself with lateral reading: Crash course navigating digital information #3 [Youtube video]. https://www.youtube.com/watch?v=GoQG6Tin-1E&t=157s.

Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000263

Greene, J. A., Copeland, D. Z., Deekens, V. M., & Seung, B. Y. (2018). Beyond knowledge: Examining digital literacy's role in the acquisition of understanding in science. *Computers & Education*, *117*, 141–159.

Greene, J. A., Yu, S. B., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education*, *76*, 55–69. https://doi.org/10.1016/j.compedu.2014.03.008

Halverson, K. L., Siegel, M. A., & Freyermuth, S. K. (2010). Non-science majors' critical evaluation of websites in a biotechnology course. *Journal of Science Education and Technology*, *19*(6), 612–620. https://doi.org/10.1007/s10956-010-9227-6

Hämäläinen, E. K., Kiili, C., Marttunen, M., Räikkönen, E., González-Ibáñez, R., & Leppänen, P. H. T. (2020). Promoting sixth graders' credibility evaluation of Web pages: An intervention study. *Computers in Human Behavior*, *110*, 106372. https://doi.org/10.1016/j.chb.2020.106372

Jung, E. H., Walsh-Childers, K., & Kim, H.-S. (2016). Factors influencing the perceived credibility of diet-nutrition information web sites. *Computers in Human Behavior*, *58*, 37–47. https://doi.org/10.1016/j.chb.2015.11.044

Jung, W. S., Chung, M.-Y., & Rhee, E. S. (2018). The effects of attractiveness and source expertise on online health sites. *Health Communication*, *33*(8), 962–971. https://doi.org/10.1080/10410236.2017.1323364

Kąkol, M., Nielek, R., & Wierzbicki, A. (2017). Understanding and predicting Web content credibility using the Content Credibility Corpus. *Information Processing & Management*, *53*(5), 1043–1061. https://doi.org/10.1016/j.ipm.2017.04.003

Kiili, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating Internet sources: From versatile evaluators to uncritical readers. *Journal of Educational Computing Research*, *39*(1), 75–95. https://doi.org/10.2190/EC.39.1.e

Kobau, R., DiIorio, C., Chapman, D., Delvecchio, P., & SAMHSA/CDC Mental Illness Stigma Panel Members. (2010). Attitudes about mental illness and its treatment: Validation of a generic scale for public health surveillance of mental illness associated stigma. *Community Mental Health Journal*, *46*(2), 164–176. https://doi.org/10.1007/s10597-009-9191-x

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Machackova, H., & Smahel, D. (2018). The perceived importance of credibility cues for the assessment of the trustworthiness of online information by visitors of health-related websites: The role of individual factors. *Telematics and Informatics*, *35*(5), 1534–1541. https://doi.org/10.1016/j.tele.2018.03.021

Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction*, *21*(1), 137–151. https://doi.org/10.1016/j.learninstruc.2010.01.001

Mason, L., Junyent, A. A., & Tornatora, M. C. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers & Education*, *76*, 143–157. https://doi.org/10.1016/j.compedu.2014.03.016

Mason, L., Scrimin, S., Tornatora, M. C., Suitner, C., & Moè, A. (2018). Internet source evaluation: The role of implicit associations and psychophysiological self-regulation. *Computers & Education*, *119*, 59–75. https://doi.org/10.1016/j.compedu.2017.12.009

McGrew, S. (2020). Learning to evaluate: An intervention in civic online reasoning. *Computers & Education*, *145*, 103711. https://doi.org/10.1016/j.compedu.2019.103711

McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students' web savvy: An intervention study. *British Journal of Educational Psychology*, bjep.12279. https://doi.org/10.1111/bjep.12279

Meppelink, C. S., Smit, E. G., Fransen, M. L., & Diviani, N. (2019). "I was right about vaccination": Confirmation bias and health literacy in online health information seeking. *Journal of Health Communication*, *24*(2), 129–140. https://doi.org/10.1080/10810730.2019.1583701

Meriam Library of CSU Chico. (2010). Evaluating information - Applying the CRAAP test [PDF document]. *Meriam Library of CSU Chico*. https://library.csuchico.edu/sites/default/files/craap-test.pdf

Meyers, L. S., Gamst, G., & Guarino, A. J. (2016). *Applied multivariate research: Design and interpretation*. Sage Publications.

Morse, J. M. (2003). Principles of mixed methods and multimethod research design. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 189–208). Sage Publications.

Pretorius, L. (2018). Experiential and self-discovery learning in digital literacy: Developing the discernment to evaluate source reliability. *College & Undergraduate Libraries*, *25*(4), 388–405. https://doi.org/10.1080/10691316.2018.1530626

Queen's University Library. (2020). CRAAP criteria. *Bachelor of health sciences toolbox.* https://guides.library.queensu.ca/BHSc/CRAAP

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist*, *52*(3), 148–166. https://doi.org/10.1080/00461520.2017.1322968

Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning*, *5*(1), 47–65.

Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). IAP Information Age Publishing.

Salmerón, L., Gil, L., & Bråten, I. (2018). Effects of reading real versus print-out versions of multiple documents on students' sourcing and integrated understanding. *Contemporary Educational Psychology*, *52*, 25–35. https://doi.org/10.1016/j.cedpsych.2017.12.002

Sandoval, W. A. (2017). Situating practices of epistemic cognition. In T. G. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change* (1st ed., pp. 253–260). Routledge. https://doi.org/10.4324/9781315467139-30

Smallwood, C. (2015). *The complete guide to using Google in libraries: Instruction, administration, and staff productivity* (Vol. 1). Rowman & Littlefield.

Stanford History Education Group. (2016). *Evaluating information: The cornerstone of online civic reasoning* [PDF document]. https://stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf

Stanford History Education Group. (2020). COR. *Civic Online Reasoning*. https://cor.stanford.edu

Subramaniam, M., Taylor, N. G., St. Jean, B., Follman, R., Kodama, C., & Casciotti, D. (2015). As simple as that?: Tween credibility assessment in a complex online world. *Journal of Documentation*, *71*(3), 550–571. https://doi.org/10.1108/JD-03-2014-0049

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics: International edition*. Pearson.

Taylor, W. D., Johnson, G., Ault, M. K., Griffith, J. A., Rozzell, B., Connelly, S., Jensen, M. L., Dunbar, N. E., & Ness, A. M. (2015). Ideological group persuasion: A within-person study of how violence, interactivity, and credibility features influence online persuasion. *Computers in Human Behavior*, *51*, 448–460. https://doi.org/10.1016/j.chb.2015.04.071

The News Literacy Project. (2020). *Checkology*. Checkology. https://get.checkology.org

Ulyshen, T. Z., Koehler, M. J., & Gao, F. (2015). Understanding the connection between epistemic beliefs and Internet searching. *Journal of Educational Computing Research*, *53*(3), 345–383.

Van Strien, J. L., Brand-Gruwel, S., & Boshuizen, H. P. (2014). Dealing with conflicting information from multiple nonlinear texts: Effects of prior attitudes. *Computers in Human Behavior*, *32*, 101–111.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2013). Fostering students' evaluation behaviour while searching the internet. *Instructional Science*, *41*(1), 125–146. https://doi.org/10.1007/s11251-012-9221-x

White, A. M., DeCuir-Gunby, J. T., & Kim, S. (2019). A mixed methods exploration of the relationships between the racial identity, science identity, science self-efficacy, and science achievement of African American students at HBCUs. *Contemporary Educational Psychology*, *57*, 54–71. https://doi.org/10.1016/j.cedpsych.2018.11.006

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106. https://doi.org/10.3102/0002831209333183

Wineburg, S., & McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information [PDF document]. https://pdfs.semanticscholar.org/5b26/9628f4dc29b514dfcb0b5e429e49fc0dae6d.pdf?_ga=2.134731505.1090439893.1607365627-347836626.1607210879

Wineburg, S. S. (1991). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, *28*(3), 495–519.

Table 1: CRAAP Training Overview

Webpage #	Description	Criteria Assessed
1	An authentic source was selected from a search engine results page to model how participants could assess the webpage's: a. Timestamp or revision and copyright history b. design features c. link functionality d. currency for sample task	Currency, Relevance
2	A second authentic source was selected from a search engine results page to model how participants could assess the author's credentials and experience by looking for their biography. Since no biography was available for the author, a Google search was used to find their credentials, affiliations, and previous work. The venue's purpose was also discussed by highlighting the advertisements and searching for ownership and funding details.	Authority, Purpose
1, 2, 3	Using Snopes, a fact-checking website, evidence about the sample topic was search for to demonstrate that evidence can be assessed within and across sources. Snopes reports about medically assisted death was examined to demonstrate the websites' use. Returning to websites 1 and 2, the content's references were examined and indicators of a peer review processes were identified.	Accuracy
1 & 2, 4	Using Mediabiasfactcheck, a news media rating website, the publishers of webpages 1 and 2 were searched and their bias, ownership, and accuracy reports were examined. The contents of each webpage was also examined for indicators of bias in the text.	Authority, Accuracy, Purpose

Table 2: Webpage Descriptions for Inquiry Activity

Webpage	Title	**Author Description**	**Publication Type**	**Publication Venue**	Publication Year	Main Claims
1	Violent victimization more prevalent among people with mental health-related disabilities	Marta Burczycka, Analyst	Government website	Statistics Canada	2018	People who suffer from mental health-related disabilities are **more likely to be the victims of violence**.
2	Yes, the untreated seriously mentally ill *are m*ore violent than others	D. J. Jaffe, Executive Director of the Mental Illness Policy Org.	News website	National Review	2017	**Those with untreated mental illness are more prone to engaging in acts of violence**.
3	Media's damaging depictions of mental illness	Margarita Tartakovsky, Associate Editor	Educational website	PsychCentral	2019	The portrayal of mental illness in the **media contributes to the stigma against those that suffer from mental illness**, several myths are addressed.
4	Violence, mental illness and substance use	Canadian Mental Health Association, BC Division	Not-for-profit website	Here To Help	2013	A description of mental illnesses and their **relation to other behaviours including violence and substance use**.
5	People with mental illness twice as likely to be victims of violence: study	Katya Slepian, Reporter	News website	The Columbia Valley Pioneer	2018	**Those who experience mental illness are more likely experience**

						violence and demonstrating this through data can help reduce the stigma against mental illness.
6	How mental illness is misrepresented in the media	Kristin Fawcett, Contributor	News website	U.S. News & World Report	2015	**Media depictions of mental illness are often misleading** and contribute to the stigma around the subject.

***Note*. Differences influencing epistemic evaluations are in bold**

Table 3: Selected Justification Criteria with Illustrative Examples

Micro-level category	Description	Illustrative example (Participant number)
Macro-level category: Currency		
Date	An evaluative decision about a rank based on participant's consideration of webpage's various publication dates. Participant may mention any of the following: • Original publication date • Revision date • Copyright date	"The time stamp and copyrights are up to date" (PN 227) "The currency of the article at first glance is from 2014, though on further inspection, it was last modified in October 2018 and contains relevant information from 2018." (PN 255)
Design	An evaluative decision about a rank based on participant's consideration of webpage's design or format features. Participant may mention any of the following: • Links to social media or other sharing capabilities • Pop up boxes • Videos/images/advertisements • Webpage "look" is older/newer • Format of webpage or its content	"Webpage opened with advertising all over it, pop ups at the bottom, etc." (PN 293) "The web page style hasn't been updated in a decade." (PN 232)
Macro-level category: Relevance		
Alignment	An evaluative decision about a rank based on participant's consideration of webpage's alignment with their task definition. Participant may mention any of the following: • topic relevance (e.g., central or peripheral to task) • appropriateness of evidence for task (e.g., type of evidence) • current enough to complete task	"Being a national review, it might not pinpoint exactly what you are looking for on a local level" (PN 213) "It talks quite a bit about substance abuse so it's very relative to the topic" (PN 242)
Type	An evaluative decision about a rank based on participant's consideration of the type or location of the webpage. Participant may mention any of the following: • News article or news section • Journal article • Blog • Encyclopedia	"Columbia Valley Pioneer - A published news article / blog is considered a lesser credible source of information based on the fact that there can be underlying unknown factors behind the write-up." (PN 224) "Although the website has a quite in-depth analysis and has many reliable sources, I ranked it fifth on the list. This is because it is news article. In comparison to the other sites I would place it lower in reliability." (PN 277)

Macro-level category: Authority		
Author expertise	An evaluative decision about a rank based on participant's consideration of author or organization's expertise related to the topic. Participant may mention any of the following: • author or organization's credentials (e.g., educational degree) • author or organization's affiliations • author or organization's years of experience • author or organization's funding status (e.g., what area are they funded for?)	"It's a government website which draws data from credible scientists who work for StatsCan." (PN 209) "It seems to me an important center and concerned with mental health and with a lot of experience." (PN 220)
Venue	An evaluative decision about a rank based on participant's consideration of the body that published the webpage. Participant may mention any of the following: • publisher's reputation (e.g., factual reporting, sourcing practices, fact checking system) • publisher's funding • publisher's ownership	"Because it was made by Statistics Canada which should be the more reliable since it has been made by the government." (PN 219)
Macro-level category: Accuracy		
Evidence	An evaluative decision about a rank based on participant's consideration of the sources of information provided as evidence. Participant may mention any of the following: • quality of evidence • type of evidence • source of evidence • currency of evidence • sufficiency of evidence	"Although this article does have some evidence to back up its claims, they are relatively few." (PN 203) "The article's accuracy is questionable since the sources used for the examples used aren't scientific." (PN 255)
Corroboration	An evaluative decision about a rank based on participant's consideration of whether information can be corroborated or not in other sources.	"I don't know if it's reliable or not, less statistics and goes against another one [source]." (PN 292) "I have learnt similar information at school. I also learnt from home and my own experience." (PN 302)
Macro-category: Purpose		

Argument	An evaluative decision based on participant's consideration of author or organization's argument structure (e.g., one sided or two sided argument).	"Seems to come from a very one sided source. Does not talk about positive part of mental illness victims." (PN 262) "Invalid reasoning and conclusion." (PN 300)
Bias	An evaluative decision based on participant's consideration of author or organization's bias(es). Participant may mention any of the following: 1. source selection 2. labelling (e.g., expert, spokesperson, official)	"The mediabias [rating from mediabiasfactcheck.com] was good IMO." (PN 290) "Their title attracts the wrong idea, even though are talking about how people with mental illness are not actually more violent the title of the article gives the opposite idea." (PN 294)

Table 4: Illustrative Examples of Essay Scores

Score	Illustrative example of essay response
0	To be honest, my background is not on the mental illness. However, I have seen many people having mental illness. They are more likely affected by their environment such as parents, school or work environment and people. And it could be resolved when their problematic environment treated.
1	I believe mentally ill individual can are likely to show aggressive behaviour (to different extent).
2	I believe that people who experience mental illness are not more violent compared to the rest of the population. Rather they people with mental illness are more vulnerable and prone to being a victim of violence. From the article from The Columbia Valley Pioneer, titled " People with mental illness twice as likely to be victims of violence: study", the article talked about how although some people who suffer from a mental illness may be violent, it outweighs the fact that these group of people are far more likely to be victims of violence. The article furthers talking about stats of people with mental illness who report the violence/assaults they experienced to the police, the number of people who report their assaults are only about 22%. I believe with this fact alone we can justify the misconception of mental illness associated with violence, these group of people feel as they have to defend themselves from the rest of the population to survive and have resorted to maybe a violent defensive mechanism. I believe people suffer from mental illness don't feel safe to discuss the violence they suffer as they will not get the right kind of help if they do so. I don't think people with mental illness present more violence than the rest of the population, I believe it is a stigma that people need to start stop believing. This will be one step into combating the misconception.
3	Individual with mental illness tends to be more violent than individuals without mental illness, because mental illness often come hand-in-hand with other issues, like binge drinking and heavy drug use, he added, which by themselves can make people more aggressive. Statistics Canada figures show 15 per cent of people with mental illness used drugs, compared to six per cent of those without. People living with mental health conditions may experience stigma, discrimination and social exclusion that significantly impacts on their lives. On the other side, mental illness individuals tend to be victimized more than other people. According to statistics Canada, individuals with mental health-related disabilities were less likely to report their victimization to the police (22% versus 31%). Among those who did, levels of satisfaction with police action were similar to those reported by victims with no such condition. So the reason why mental ill individual tends to be more violent than individuals without mental illness, is for multiple reason. They tend to have experienced more stuff, half (51%) of people with a mental health-related disability reported having experienced physical or sexual abuse as a child, and almost a quarter (23%) reported having experienced homelessness. Binge drinking (44%) and drug use (15%) were significantly more common among those with mental health-related disabilities. Mental ill people don't become violent just like this, it is a series of event that led to it.

4	Yes, I think mentally ill populations are more violent than other sectors of the population. I personally know family members and friends who show such violent tendencies and who have been diagnosed with some forms of mental illness. I also think that although many sufferers of mental illness become ill from exposures to forms of violence and victimization, and have been passive victims, I feel that their victimization begets violent tendencies for themselves. Its akin to the adage of the sins of the father being passed down and from one person or act onto another. I have family members and friends who due to victimization, have become more violent-prone. However, that's not to say that some sufferers of mental illness aren't docile and passive, but some sufferers do tend to be more aggressive and violent than populations that have not been diagnosed with such an illness. I also think that the severity of their violent tendencies and symptoms vary based on their specific diagnoses. For example, sufferers of PTSD and bipolar disorder may act out more acts of violence (to self and/or to others) than other relatively "normal" forms of mental illness. I feel this way because of statistics and certain news reporting, such as veterans of wars coming back home and kill themselves and/or their families. In general, I think that those who suffer from mental illness may not have such great control of their actions and thinking. Such distortions and lack of self-control therefore begets violent tendencies. I don't think this pertains to all mental illness sufferers, just some. So to sum up, yes individuals with mental illness can be more violent than those without mental illness, but I wouldn't tag this label to all sufferers of mental illness as a general statement.
5	In my opinion based on the articles read and prior understanding, individuals with mental illness are generally NOT more violent than individuals without mental illness. There is evidence that individuals with mental illness, however, experience violence more than individuals without mental illness, but this does not mean that they are violent themselves. The Statistics Canada article reported that 1 in 10 people with mental illness experience violence (i.e., are victimized vs. causing violence). The "Here to Help" article indicated that in a large US study, the risk of violence from a mentally ill person without a co-existing substance abuse problem was about the same as the general population over the next three years. Now in my response I did say "generally" since in some specific types of mental illness such as schizophrenia or psychosis, there is an increased risk of violence. There are complexities in understanding the link between violence and mental illness due to co-existing factors such as substance abuse, lower income, and marginalization. With these co-factors present, violence may be erroneously correlated to mental illness. If only 3% (estimated) of violent crime in Canada is attributed to mental illness, I stand by my opinion that in general individuals with mental illness are not more violent than those without mental illness.

Table 5: Essay Score Coding Scheme with Illustrative Examples

	1 point	0 points	Illustrative examples (Participant number)
Making a claim	Participant makes a claim or takes a position, regardless of the claim's accuracy.	Participant does not make a claim or they do not justify their claim.	"In my opinion, I think that individuals with mental illness are more violent than those without mental illness." (PN 202)
Presenting supportive grounds (One-sided argument)	Participant supports their position with valid arguments and evidence.	Participant does not support their position, or their arguments and evidence are invalid.	"Based on the presented data, it would seem that, in general, there is no direct connection between outward violence and mental illness. If anything, mentally ill are more susceptible to inward violence. That is to say their illness might cause them to be victimized more often than general population." (PN 258)
Presenting alternative perspectives (Two-sided argument)	Participant engages with an alternative perspective by identifying valid arguments and evidence in support of that perspective.	Participant only presents arguments that support one perspective. Participant may acknowledge an alternative perspective but does so without providing evidence to support that perspective.	"Now in my response I did say "generally" since in some specific types of mental illness such as schizophrenia or psychosis, there is an increased risk of violence. There are complexities in understanding the link between violence and mental illness due to co-existing factors such as substance abuse, lower income, and marginalization. With these co-factors present, violence may be erroneously correlated to mental illness." (PN 269)
Evaluating claims	Participant explains whether a claim is credible, limited or generalizable.	Participant accepts or dismisses a claim without evaluating it or explaining why it should be accepted or rejected.	"Since there are so many different kinds of mental illness and the degree of one person's mental illness can differ from someone else. One person might be stable with proper treatment and medication while others are not receiving treatment. So one person could have violent tendencies, so then media portrays all mental illness as being more violent. Instead each violent case should be examined individually because you can't stereotype one trait to an entire group of people." (PN 294)
Reconciling or integrating perspectives	Participant acknowledges the validity of arguments on both sides, which	Participant does not include a conclusion or creates a one-sided conclusion	"I agree that there are some people who suffer from mental illness that are extremely violent, that need to be confined, and can react unexpectedly. However, I believe that those are a small

	is consistent with their evaluation of claims.	that does not include an alternative perspective.	fraction of the mental health population. The media is quick to portray someone with mental health issues as violent as it adds to entertainment value. This perpetuates the stereotype. I do not believe that having a mental illness makes you a more violent person. However, I do believe and it was confirmed by the websites that people with mental illness are more likely to be victims of crime." (PN 281)

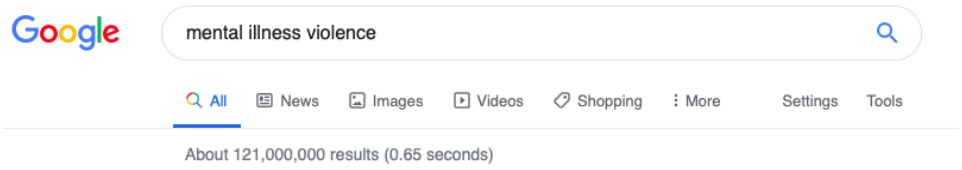
Table 6: Means and Standard Deviations of Ranking and CRAAP Criteria Used by Condition

Variables	Training group (*n* = 41)		Control group (*n* = 41)	
	M	*SD*	*M*	*SD*
Overall rank-ordering	13.27	2.99	14.63	2.58
Most & least reliable rank-ordering	8.63	2.94	9.98	2.80
Currency*	1.15	1.94	0.17	0.54
Relevance	0.80	1.19	.93	1.15
Authority	1.12	1.50	1.20	1.50
Accuracy*	1.02	1.29	1.98	1.57
Purpose	1.37	1.70	1.60	1.60

Notes. **Lower rank-ordering scores represent a more accurate rank-ordering**
***p < .05**

Table 7: *Means and Standard Errors of Essay Scores by Condition*

Variables	Training group (*n* = 41)		Control group (*n* = 41)	
	M	*SD*	*M*	*SD*
Total essay score	2.2	.21	2.73	.21
Claim	.85	.36	.95	.22
Support for a claim	.83	.38	.95	.22
Alternative perspective	.17	.38	.32	.47
Evaluate claims	.27	.45	.34	.48
Reconcile perspectives	.07	.38	.17	.38



Violent victimization of Canadians with mental health-related disabilities
https://www150.statcan.gc.ca › daily-quotidien › 181018 › dq181018b-eng
Oct 18, 2018 - **Violent** victimization of Canadians with **mental health**-related ... and older with a disability related to **mental health** experienced **violence** in the ...

Yes, the Untreated Seriously Mentally Ill Are More Violent Than Others
https://www.nationalreview.com › 2017/11 › untreated-seriously-mentally-...
Nov 9, 2017 - Yes, the Untreated Seriously **Mentally Ill** Are More **Violent** Than Others ... industry is "people with **mental illness** are no more **violent** than others.

Media's Damaging Depictions of Mental Illness - Psych Central
https://psychcentral.com › lib › medias-damaging-depictions-of-mental-illn...
by M Tartakovsky - Cited by 25 - Related articles
A man who suffers from **schizophrenia** goes on a shooting spree in Times Square and later stabs a pregnant physician in the stomach. These are the opening ...

Violence, Mental Illness and Substance Use | Here to Help
https://www.heretohelp.bc.ca › infosheet › violence-mental-illness-and-sub...
This connection can fuel fear and this fear of **violence** can be a driving force for negative attitudes toward people with **mental illnesses** and substance use ...

People with mental illness twice as likely to be victims of ...
https://www.columbiavalleypioneer.com › news › people-with-mental-illness...
Oct 29, 2018 - A leading researcher on **mental health** in Canada hopes figures showing the high rates of victimization against those with **mental illness** will ...

How Mental Illness is Misrepresented in the Media | Wellness ...
https://health.usnews.com › Health › Wellness
Apr 16, 2015 - Unless you majored in psychology or attended medical school, chances are the bulk of your knowledge about **mental illness** comes from the ...

Figure 1: Search Engine Results Page for Inquiry Activity