Corroborating students’ self-reports of source evaluation

Alexandra List & Patricia A. Alexander

To cite this article: Alexandra List & Patricia A. Alexander (2018) Corroborating students’ self-reports of source evaluation, Behaviour & Information Technology, 37:3, 198-216, DOI: 10.1080/0144929X.2018.1430849

To link to this article: https://doi.org/10.1080/0144929X.2018.1430849

Published online: 03 Feb 2018.

Submit your article to this journal

Article views: 125

View Crossmark data
Corroborating students' self-reports of source evaluation

Alexandra Lista and Patricia A. Alexander

ABSTRACT

Students were asked to report their typical practices with regard to source evaluation using the Credibility Assessment Scale (CAS). Students’ reports were validated against behavioural and cognitive indicators of source evaluation. Specifically, while researching a social science prompt, students’ source use behaviours, related to text evaluation, were logged. Following task completion, students were asked to rank the trustworthiness of the information sources they accessed and to justify their rankings. The criteria students cited for rankings of text trustworthiness were considered to be cognitive-based indicators of source evaluation, and mapped onto CAS items. Limited correspondence was found between students’ reported engagement in verification-related behaviours and either the behaviours manifest during task completion or the criteria for source evaluation cited at post-task. At the same time, a correspondence was found among behavioural and cognitive aspects of source evaluation, within the context of a specific task. This study is unique in directly corroborating self-reported, behavioural, and cognitive measures of source evaluation and examining these within the context of a rich and naturalistic multiple text task. Implications for research and practice are discussed.

ARTICLE HISTORY

Received 1 March 2017
Accepted 17 January 2018

KEYWORDS

Heuristic evaluation; cognition; hypertext

In a survey of college students, Wiley et al. (2009) found 79% of undergraduates to report never having received training in how to evaluate sources when searching for information on the Internet. This percentage helps to explain the limitations in undergraduates’ source evaluation documented widely in the empirical literature (Britt and A格林skas 2002; Tillotson 2002; Meola 2004; Stadtler and Bromme 2007; Bråten, Stromso, and Britt 2009; Currie et al. 2010; Bråten, Strømso, and Salmerón 2011; Macedo-Rouet et al. 2013; Salmerón, Macedo-Rouet, and Rouet 2016; Stadtler et al. 2016; Kiili et al. 2017). But, survey and empirical findings on undergraduates’ source evaluation have not always aligned. For instance, Head and Eisenberg (2010) found over 70% of surveyed undergraduate students to report considering the author of a text when evaluating information. Yet, Stadtler and Bromme found only 20% of students in a control group to include document information, like author, in their written notes, based on multiple texts, as compared to 32% of students in a treatment group receiving an intervention to promote source evaluation (2007). Given the preponderance of methodological approaches that have been used to examine undergraduates’ source evaluation and the discrepancies among them, there is a need to better understand the correspondence between reported and observed measures of source evaluation. The purpose of the present study is to examine the alignment among (a) survey, (b) behavioural, and (c) cognitive measures of source evaluation in capturing students’ judgements of texts.

Text evaluation

Text evaluation has been defined as students’ positive or negative judgements of some specific aspect of text or texts overall (Wolfe and Goldman 2005). Although judgements of texts may be determined according to any number of criteria, including relevance, usefulness, reliability, or reputability (Rouet et al. 1997; Henry 2006; Sanchez, Wiley, and Goldman 2006; Kiili, Laurinen, and Marttunen 2008; Wiley et al. 2009; Macedo-Rouet et al. 2013; Stadtler et al. 2013), most commonly students have been asked to evaluate texts based on either their credibility (Wiley et al. 2009; Mason, Boldrin, and Ariasi 2010) or trustworthiness (Rouet et al. 1996; Britt and A格林skas 2002; Bråten, Strømso, and Britt 2009; Bråten, Strømso, and Salmerón 2011).

While students’ credibility judgements have been investigated as based on content-related factors, like message convincingness, plausibility, or believability (Mason, Boldrin, and Ariasi 2010; Lombardi and Sinatra 2012; Metzger and Flanagan 2013), trustworthiness
evaluations have been conceptualised more narrowly as based on students’ judgements of document information. Document information, like author or publisher, refers to text metadata or source characteristics that can be used to ascertain text origin. For instance, Bråten, Stremö, and Salmerón (2011), in investigating students’ ratings of text trustworthiness, examined the extent to which these were based on document information like author, document type (e.g. textbook, newspaper article), publisher, and date of publication, as well as on source content. Considering aspects of the author, in particular, Stadtler and Bromme (2014) suggest that students evaluate source trustworthiness based on two primary considerations: (a) author expertise, or authoritativeness and qualifications to provide correct information, and (b) author benevolence, or intentions to provide accurate and unbiased content.

A variety of methodological approaches have been adopted to examine students’ trustworthiness evaluations (e.g. Brem, Russell, and Weems 2001; Britt and Aglinskas 2002; Bråten, Stremö, and Salmerón 2011; Stadtler and Bromme 2014; Mason et al. 2017). Generally, these may be said to fall into three categories: (a) survey and other self-report approaches, (b) behavioural measures of source evaluation, and (c) cognitive approaches to understanding how students judge texts.

### Surveys and source evaluation

Many self-report-based approaches have been used to assess students’ source evaluation. In one example, Burton and Chadwick (2000) asked college students about their preferred criteria when evaluating texts during information search. Students reported prioritising source availability, ease of understanding, and ease of location when evaluating both library sources and websites on the Internet. Although students also reported considering author credentials, argument originality, and a source being peer reviewed in evaluating texts, these criteria ranked much lower among students’ priorities. As a contrast, Kim and Sin (2007) found students to deem source accuracy or trustworthiness to be a top priority, followed by source accessibility, when rating the criteria they prioritised when evaluating texts. Purdy (2012) asked students to identify the standards they used in determining their favourite or most preferred sources. By far the most popular reason cited for source preference was ease of use, forwarded by 36.90% of students, followed by considerations of quality. Discrepancies in preference criteria reported were large; ease of use was cited as a criterion for source preference twice as frequently as was any other reason.

A number of survey measures have also been developed to assess students’ source evaluation. Flanagin and Metzger (2000, 2007) developed a questionnaire to determine the frequency with which students engaged in a variety of verification behaviours when looking for information on the Internet. This nine-item survey was administered to undergraduate samples across three years. Overall, students reported most commonly considering whether information was current and whether sources represented fact or opinion; verifying authors’ qualifications or credentials was the least commonly cited verification strategy. Likewise, Hargittai et al. (2010) developed an eight-item, behaviourally focused Credibility Assessment Scale (CAS). Mirroring results from Metzger (2007), the verification behaviour students reported most frequently engaging in was considering information currency, with visiting the about us page constituting the evaluation behaviour least frequently employed, indicating students’ limitations in considering author credentials when evaluating texts.

While survey measures of source evaluation have typically been found to have high reliability, initial efforts to validate self-report data against students’ source use behaviours have fallen short. Hargittai et al. (2010) collected observational and interview data on the source use of a subset of undergraduate students completing the Credibility Assessment Scale. Despite many students reporting frequently considering author expertise when evaluating sources, Hargittai et al. found only 10% of students to actually comment on author credentials during Internet search. Beyond this identified discrepancy, Hargittai et al. were limited in directly connecting students’ self-reported source evaluation practices with manifest behaviour. Flanagin and Metzger (2007) attempted to validate a self-report source verification scale by examining its association with observed evaluation-related behaviours (i.e. the number of additional links students accessed, beyond a single source). Although investigating only a single behavioural indicator, Flanagin and Metzger found students’ self-reported engagement in verification behaviours to be negatively correlated with the number of external websites they visited when searching for information on the Internet. Findings from both Hargittai et al. and Flanagin and Metzger suggest the need to further validate students’ reported evaluation practices against a broader repertoire of source evaluation behaviours manifest during Internet search.

### Behavioural measures of source evaluation

A number of behavioural approaches have also been adopted to capture students’ text evaluations in-situ. These have included both descriptive approaches, like
eye-tracking to monitor students’ attendance to document information (Gerjets, Kammerer, and Werner 2011; Vidal-Abarca et al. 2011; Kammerer and Gerjets 2012), and interventions, aimed at designing user interfaces to foster sourcing (i.e. the evaluation of document information, like author; Brit and Aglinskas 2002; Stadtl and Bromme 2007, 2008).

Gerjets, Kammerer, and Werner (2011) used eye-tracking, alongside think-aloud data, to examine students’ source evaluations across two conditions. In the experimental condition, students were explicitly instructed to evaluate texts whereas in the control condition no specific evaluation instructions were provided. The number and duration of fixations on document information (i.e. author, publisher) were examined when students evaluated both search results and websites accessed. However, the two conditions were not found to differ on either eye-tracking metric (i.e. number and duration of fixations).

At the same time, students’ attendance to document information, as assessed through eye-tracking, has been found to be associated with its salience, or explicitness of presentation on a webpage (Kammerer et al. 2009). This may be why user interfaces designed to promote students’ attendance to document information have been found to be effective in encouraging source evaluation (e.g. Britt and Aglinskas 2002; Stadtl and Bromme 2007; 2008). Britt and Aglinskas (2002) designed the Sourcer’s Apprentice to both train students in source evaluation and to help them associate document information (e.g. author, point of view, date) with notecards written based on information presented across multiple texts. Through three experiments, the Sourcer’s Apprentice intervention was found to significantly improve students’ task performance, as compared to a control group, and, specifically, to support the correct identification of texts’ document information at post-test. Comparing performance on a transfer task, essays written by students in the Sourcer’s Apprentice training condition received higher grades from teachers and included significantly more document information (e.g. citations) than did essays from students in the control group.

An additional fount of behavioural data on source evaluation has been students’ writing. Commonly, students’ written products have been examined for evidence of either spontaneous or instructed citation or direct referencing to sources (Davis 2002; Knight-Davis and Sung 2008; Strømsø et al. 2013; Strømsø and Bråten 2014; List, Alexander, and Stephens 2017). Davis (2002) examined undergraduates’ citation behaviours across a seven-year period. Over time, he found students to reference significantly fewer scholarly sources when writing research papers, despite citing significantly more sources, overall, pointing to students supplementing traditional academic texts with web materials of varying quality. Looking more specifically at the particular sources students cited, McClure and Clink (2009) evaluated references used in students’ research papers according to their authority, bias, and timeliness. However, such evaluations were researcher-determined rather than student-generated, highlighting the need to supplement behavioural evidence of source evaluation with student report. Indeed, a number of methodological approaches have focused on examining students’ reasoning about source evaluations, as either stand-alone analyses or in association with behavioural metrics.

**Cognitive measures of multiple source use**

A variety of cognitively focused methodological approaches have been used to assess source evaluation. These have included interviews, think-alouds, and tasks asking students to rate or rank texts according to particular dimensions, like trustworthiness or reliability (Rouet et al. 1996; Hofer 2004; Bråten, Strømsø, and Britt 2009; Walraven, Brand-Gruwel, and Boshuizen 2009; Wiley et al. 2009; Head and Eisenberg 2010; Bråten, Strømsø, and Salmerón 2011; Mason, Ariasi, and Boldrin 2011; Barzilai and Zohar 2012; Rowley and Johnson 2013). Methods such as these can be characterised according to two dimensions: (a) temporality and (b) cuing. The temporality dimension captures when students’ text evaluations took place, either during or following task completion. Cuing reflects the extent to which students were explicitly prompted to evaluate texts or whether spontaneous judgements are examined.

During task completion, source evaluation has most commonly been examined using think-aloud methodologies. Students’ utterances have been analysed for evidence of explicit source evaluation or epistemic cognition (i.e. manifestations of students’ beliefs about what constitutes valid sources of knowledge), associated with text evaluation. Such investigations have generally come to two conclusions. First, they have found students to explicitly evaluate texts only rarely (Kiili, Laurinen, and Marttunen 2008; Goldman et al. 2012; Bråten et al. 2014). Second, they have suggested that when students do evaluate texts, these evaluations can be thought of as falling along a continuum, reflecting more or less sophisticated understandings of where knowledge comes from and how it may be substantiated (i.e. epistemic cognition; Mason, Ariasi, and Boldrin 2011; Barzilai and Zohar 2012; Kammerer, Amann, and Gerjets 2015; Greene, Sandoval, and Bråten 2016). For example, Mason, Boldrin, and Ariasi (2010) coded students’
think-aloud utterances as aligning with learners’ beliefs about sources of knowledge. Students’ evaluations of texts were ordered in sophistication from judgements made based on the (a) popularity of the source or the (b) authoritativeness or expertise of the author, to judgements rendered based on the (c) scientific nature of the text.

While Mason, Boldrin, and Ariasi (2010) did not explicitly prompt students to evaluate texts, instead capturing students’ spontaneous epistemic cognitions associated with source evaluation, other in-task measures have specifically instructed students to evaluate texts during multiple source use. These more directed approaches to understanding source evaluation have included efforts to compare think-aloud data when students have been explicitly instructed to evaluate texts or not (Gerjets, Kammerer, and Werner 2011) and interfaces designed to explicitly prompt students to evaluate texts during information use (Stadtl and Bromme 2007; 2008; Kammerer and Gerjets 2010; List, Alexander, and Stephens 2017). As may be expected, studies that have compared spontaneous vis-à-vis instructed text evaluation have found students explicitly directed to evaluate texts to do so to a much more considerable extent and to have better recall for document information (Stadtl and Bromme 2007; Gerjets, Kammerer, and Werner 2011), indicating greater source evaluation.

While measures of text evaluation collected during multiple text use have primarily been focused on capturing spontaneous evaluations (e.g. Kiili, Laurinen, and Marttunen 2008; Mason, Boldrin, and Ariasi 2010; Mason, Ariasi, and Boldrin 2011; Bråten et al. 2014), assessments of text evaluation following task completion have tended to be more explicit in asking students to judge texts. Specifically, such studies have commonly asked students to rate or rank texts according to a variety of criteria and to justify their evaluations to varying degrees. Specifically, we were interested in examining the extent to which students’ reports of source evaluation behaviours aligned with their demonstrated behaviours during a multiple text task and with the justifications for source evaluation they reported following task completion. We had the following research questions:

In addition to prompted text evaluations, students’ spontaneous judgements of texts following task completion have also been examined. Walraven, Brand-Gruwel, and Boshuizen (2009) conducted group interviews with students after they had completed a research task requiring the use of multiple texts. Through interviews, Walraven, Brand-Gruwel, and Boshuizen (2009) captured the criteria students spontaneously mentioned as the bases for source evaluation. Most commonly, students reported considering the title and rank of search results, during search, and judging the usability of websites, during text access.

In work by Walraven, Brand-Gruwel, and Boshuizen (2009), as well as in much other work (e.g. Rouet et al. 1996; Wiley et al. 2009; Strømsø and Bråten 2014), behavioural indicators of source evaluation have been examined alongside students’ cognitions about judgements of texts. For instance, Gerjets, Kammerer, and Werner (2011) examined both think-aloud data (i.e. cognitive) and students’ fixation patterns (i.e. behavioural) for evidence that students looked at document information, like title and author, to evaluate texts. Furthermore, Goldman et al. (2012) in comparing the differences between extreme groups of students, those who learned disproportionately much or little from a multiple text task, found them to differ in both behavioural and cognitive metrics of source evaluation. Specifically, these two groups differed in both their visits and revisits to reliable versus unreliable websites (i.e. behavioural) and in the number of think-aloud utterances they produced in association with evaluating the credibility and reliability of texts (i.e. cognitive).

Present study

Overall, behavioural and cognitive indicators of source evaluation have largely been consistent with one another in highlighting students’ limitations with regard to source evaluation (Walraven, Brand-Gruwel, and Boshuizen 2009; Gerjets, Kammerer, and Werner 2011; Goldman et al. 2012). At the same time, these metrics have rarely been directly juxtaposed with one another or jointly compared to students’ reported source evaluation practices. The purpose of the study is to directly investigate the alignment between students’ intended, demonstrated, and reported indicators of source evaluation. Specifically, we were interested in examining the extent to which students’ reports of source evaluation behaviours aligned with their demonstrated behaviours during a multiple text task and with the justifications for source evaluation they reported following task completion. We had the following research questions:
1. To what extent do students report engaging in a variety of source evaluation behaviours when conducting research on the Internet? 
2. What is the correspondence between students’ reported source evaluation behaviours (i.e. self-reported indicator) and those manifest while completing a task requiring the use of multiple texts (i.e. behavioural indicator)? 
3. What is the correspondence between students’ reported source evaluation behaviours (i.e. self-reported indicator) and criteria reported when justifying rankings of text trustworthiness (i.e. cognitive indicator)? 
4. What is the correspondence between students’ evaluation-related behaviours, demonstrated during task completion (i.e. behavioural indicator), and justifications cited for rankings of text trustworthiness (i.e. cognitive indicator)? 
5. To what extent do self-reported, behavioural, and cognitive measures of source evaluation predict citation in students’ responses?

**Methods**

**Participants**

Participants were 197 undergraduate students at a large, Mid-Atlantic university (female: 65.48%, n = 129; male: 29.95%, n = 59). The sample was, on average, 20.47 years of age (SD = 2.08). Students were racially and ethnically diverse. Specifically, 49.23% of students were White (n = 97), 19.29% were Asian (n = 38); 16.24% were African American (n = 32), 3.55% were Latino (n = 7), and 6.09% of students (n = 12) reported being biracial or multiracial. Nine participants did not provide demographic information.

**Measures**

Data were collected across three phases. In the first phase, participants completed a variety of individual difference measures. These included self-report scales capturing students’ frequency of source use (i.e. Information Source Credibility Scale, Metzger, Flanagan, and Zwarun 2003) and their frequency of engagement in a variety of source evaluation behaviours (i.e. Credibility Assessment Scale, Hargittai et al. 2010). In the second phase, participants complete a multiple text task. This involved researching the topic of the Arab Spring in Egypt and composing a written response based on information provided in a library of six digital texts. During the multiple text task, log data were captured about students’ text use, including text revisits and whether or not students accessed document information for each text used. In phase three, participants were asked to rank the texts they accessed in terms of trustworthiness and to justify their rankings.

While phase one was completed by participants online, at a time and location of their choosing, phases two and three were completed online, in a computer lab with a researcher present. Data for this study come from a larger study examining students’ practices with regard to source evaluation (List, Alexander, and Stephens 2017) but are uniquely presented in this manuscript.

**Individual difference measures**

Participants completed an academic behaviours questionnaire to capture their typical habits and practices when researching information on the Internet. Included in the questionnaire were the Information Source Credibility Scale and the Credibility Assessment Scale.

**Information Source Credibility Scale**

The Information Source Credibility Scale (Metzger, Flanagan, and Zwarun 2003) presented students with eight different document types (e.g. journal articles, newspapers, textbooks, Wikipedia), including those featured in this study, and asked them to report (a) how frequently they used each in completing schoolwork and (b) how credible they considered each to be. Specifically, for each document type, students were asked to endorse: (a) how frequently do you use each of these information sources when completing schoolwork and (b) how credible do you consider each of these information sources to be. Students were asked to rate frequency of source use on a 5-point scale, from never to all the time, and source credibility on a 7-point scale, from not at all credible to very credible. Cronbach’s alpha reliability for the eight-item scale, capturing the frequency of students’ reported use of different document types, was 0.55, indicating variability in students’ frequency of source use. Reliability for the eight-item scale, reflecting students’ credibility judgements of various document types, was 0.54, likewise indicating discrimination in students’ evaluations of different sources.

**Credibility Assessment Scale**

The Credibility Assessment Scale (Hargittai et al. 2010) included nine items that asked students to endorse the frequency with which they engaged in a variety of behaviours associated with source evaluation when looking for information on the Internet. This included asking participants to rate how frequently they would: check to see if the information is current, check to see who the author is, or check the qualifications or credentials of
the author. Students rated the frequency of each behaviour on a 7-point scale from never to very often. Cronbach’s alpha reliability for our sample was 0.86; this is comparable to other reliability values previously reported in the literature (e.g. α = 0.92; Metzger, Flanagan, and Zwarun 2003).

Multiple text task
The multiple text task included a research phase and a response phase.

Research phase
During the research phase, participants were presented with a library of six digital texts and asked to use these to respond to the prompt, Should the United States support General el Sisi and the military regime or Mohamed Morsi and the Muslim Brotherhood. The six texts were presented according to document type (i.e. blog, analysis essay, newspaper, public opinion poll, Twitter, Wikipedia) in the digital library. The library featured both traditional document types, with print analogs (e.g. newspaper), as well as new, digital document types, created on the Internet, for online consumption (e.g. Twitter). Texts ranged in their author expertise, the type of evidence provided, and the point of view forwarded, in support of or opposing Mohamed Morsi or General el Sisi, two political opponents vying for control of Egypt.

Log data of students’ behaviours during the multiple text task were captured, including texts visited and order and duration of text access. When students visited a source, only the text of the source would appear. At the top, students were presented with the option to Click here to learn more about this source. Clicking this button resulted in students’ accessing a text’s document information, including title, document type, author, publisher, date and location of publication, and URL. Students were then, further, presented with the option to Click here to learn more about the author. Clicking this link resulted in students receiving information about author background, position, and expertise. For instance, author credentials for the newspaper article stated that Maggie Michel, the author of the article, was the Associated Press Senior Reporter, Cairo Bureau.

Whether or not students elected to access texts’ document information and author credentials was also logged. Students were free to visit none, some, or all of the sources in the library as well as to revisit texts. Participants were also told that there was no time limit for the task and that while they were able to take notes during the research phase, they would not be able to return to the library texts when composing their responses.

Response phase
Following the research phase, participants were asked to align themselves with General el Sisi, Mohamed Morsi, or an Other option and to compose a written response justifying their position. Students’ responses were coded for, among other indicators, the number of citations, or direct references to specific library sources, included.

Source ranking and justification
Following the multiple text task, participants were presented with a ranking and justification task. Specifically, students were asked to rank each text accessed according to its trustworthiness, with a one designating the most trustworthy text and a six designating the least trustworthy text. Then, participants were asked to justify their rankings of each text accessed. A bottom-up coding approach was mapped onto coding categories previously established in the literature. A subset of the justification criteria students cited were examined in this study, in association with specific items on the Credibility Assessment Scale. These included justifications for trustworthiness rankings based on considerations of author’s credentials and whether a document was based on fact or opinion. Students’ justifications for trustworthiness rankings were collapsed across document types; therefore, whether or not students proposed a particular type of ranking justification for any of the texts they evaluated was examined.

Aligning indicators of text evaluation
Reports on the Information Source Credibility scale were validated in two ways. First, the frequency with which students reported using a variety of document types for their school work was compared to manifest text access. Second, students’ ratings of the credibility of different document types were compared to students’ rankings of various document types, according to trustworthiness, following task completion. A summary of validation measures used for the Information Source Credibility Scale is included in Table 1.

Table 1. Validation of the Information Source Credibility Scale (Metzger, Flanagan, and Zwarun 2003).

<table>
<thead>
<tr>
<th>Frequency of document type use</th>
<th>Frequency of document type access</th>
<th>Post-task trustworthiness rankings of document types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credibility of various document types</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Students’ endorsements of various items on the Credibility Assessment Scale were validated against behavioural measures, demonstrated during text use, and against the justification criteria students cited for source evaluation following task completion, used as a cognitive indicator of source evaluation. Each of the nine items on the Credibility Assessment Scale was validated to some extent. Behavioural validation was undertaken by mapping three groupings of items on the Credibility Assessment Scale against behaviours manifest during the multiple text task. Cognitive validation examined the correspondence between students’ endorsement of each item on the Credibility Assessment Scale and its citation as a justification criterion for students’ rankings of text trustworthiness. A summary of behavioural and cognitive measures used to validate the Credibility Assessment Scale is included in Table 2.

### Behavioural validation

Students’ responses to the Credibility Assessment Scale were validated against three types of multiple text use behaviours. First, we examined the correspondence between students endorsing items on the credibility scale associated with accessing texts’ document information (e.g. check to see if the information is current; check to see who the author is) and whether or not students accessed document information for texts visited during the multiple text task. Second, students’ endorsements of items associated with considering author background or credentials (e.g. check the qualifications or credentials of the author; consider the author’s goals/objectives for posting information) were validated against students accessing information about author credentials during text use. The third set of items concerned students’ validation of information against content presented in other texts (e.g. seek out other sources to validate the information; check to see if the information is complete/comprehensive). This was validated against two behaviours: (a) students accessing multiple sources, which almost all of the participants in our study did, and (b) students revisiting texts, considered to be an indicator of conceptualising or re-evaluating information in a text, in light of content presented in other texts.

### Cognitive validation

Students’ endorsement of each item on the Credibility Assessment Scale was further validated against the criteria students cited in justifying source trustworthiness rankings following task completion. For instance, the extent to which students endorsed the item consider the author’s goals/objectives for posting information was compared to students citing judgements of author motivation as a justification for their rankings of text trustworthiness.

Students’ trustworthiness ranking justifications were coded using both bottom-up and top-down processes. The literature was examined a-priori to determine a number of justification categories for text trustworthiness (e.g. author credentials, document type, date of

---

**Table 2. Validation of the Credibility Assessment Scale (Hargittai et al. 2010).**

<table>
<thead>
<tr>
<th>Self-report measure of source evaluation (pre-task)</th>
<th>Behavioural measure of source evaluation (during task)</th>
<th>Cognitive measure of source evaluation (post-task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check to see if the information is current</td>
<td>Access text document information</td>
<td></td>
</tr>
</tbody>
</table>
| Check to see who the author is                      | Access authors’ credentials                           | **Currency:**
| Visit the ‘about us’ page on a Web site             | Check the qualification or credentials of the author | ‘The newspaper was most trustworthy because it was live information.’ |
| Consider the author’s goals/objectives for posting information | Seek out other sources to validate the information | **Author:**
| Check to see what other sites link to the web site you are viewing | Number of multiple texts accessed | ‘It was probably written by a credible author’ |
| Check to see if the information is complete/comprehensive | |
| Consider whether the views represented facts or opinions | |

**Author:**

- ‘A reporter is usually objective and states situations as they are seen’
- ‘The analysis essay was written by a professor at a university, which means the author is certainly knowledgeable’

**Author Objectives:**

- ‘Most of the authors were of Egyptian descent and may have been biased towards the coup d’état’
- ‘News stations can be biased towards specific groups or views’

**Corroboration:**

- ‘Its information was completely different from both the essay and the newspaper article; it was the outlier’
- ‘The Wikipedia article … provided consistent information with the other credible sources.’

**Information Completeness:**

- ‘Gave a comprehensive analysis of the situation’
- ‘Seemed factually based and unbiased’

---

**Fact/Opinion:**

- ‘A blog tends to be just opinions not facts’
- ‘Seemed factually based and unbiased’

Students’ endorsements of various items on the Credibility Assessment Scale were validated against behavioural measures, demonstrated during text use, and against the justification criteria students cited for source evaluation following task completion, used as a cognitive indicator of source evaluation. Each of the nine items on the Credibility Assessment Scale was validated to some extent. Behavioural validation was undertaken by mapping three groupings of items on the Credibility Assessment Scale against behaviours manifest during the multiple text task. Cognitive validation examined the correspondence between students’ endorsement of each item on the Credibility Assessment Scale and its citation as a justification criterion for students’ rankings of text trustworthiness. A summary of behavioural and cognitive measures used to validate the Credibility Assessment Scale is included in Table 2.
publication; Rouet et al. 1996; Bråten, Strømsø, and Britt 2009). Furthermore, a bottom-up coding approach was used to identify justification categories uniquely appearing in our data (e.g. critical evaluation; writing style).

For the present analyses, specific justification categories were identified as corresponding to items on the Credibility Assessment Scale. In some cases, items on the scale directly corresponded to justification categories (e.g. currency); in other cases, several justification categories were collapsed to correspond to a single CAS item. For instance, students’ endorsement of the item check the qualifications or credentials of the author was aligned with students’ justifications for source trustworthiness based on (a) whether or not the author seemed knowledgeable, (b) whether or not the source was scholarly or academic, and (c) whether the author offered a first-hand or second-hand account of events. Two items on the CAS did not have corresponding justification categories associated with them (i.e. Visit the ‘about us’ page on a website; Check to see what other sites link to the website you are viewing). Table 2 includes the correspondence between items on the CAS and the justification categories students cited in explaining their trustworthiness rankings. Because students were fairly limited in the justifications for trustworthiness rankings they provided, each justification category was binary coded according to whether or not students cited that justification in explaining their trustworthiness rankings of any source evaluated.

Results

Research question 1: students’ reported source evaluation practices

The first research question examined descriptive information about students’ reported perceptions of various document types (i.e. Information Source Credibility Scale) and the frequency with which they reported engaging in a variety of source evaluation behaviours (i.e. Credibility Assessment Scale). Table 3 includes descriptive information about the frequency with which students reported using each of eight document types when completing schoolwork.

As may be expected, students reported most commonly using journal articles (M = 3.81, SD = 2.44) and textbooks (M = 3.68, SD = 0.99) when completing schoolwork. Indeed, 68.45% of students reported frequently or always using journal articles to complete schoolwork; 62.03% reported commonly using textbooks, while 50.27% reported frequent Wikipedia use. Students were least likely to use blogs (M = 2.11, SD = 1.07) or Twitter (M = 1.71, SD = 1.14) for completing academic work.

Table 4 provides descriptive information about how credible students considered each type of information source to be. Students’ perceptions of source credibility closely mirrored their frequency of source use for academic tasks. Students considered the textbook (M = 6.29, SD = 0.90) to be the most credible source, alongside journal articles (M = 6.27, SD = 1.05). While many students reported using Wikipedia frequently, it was rated as a moderately credible source (M = 3.80, SD = 1.48); indeed, only 34.67% of students considered it to be a credible or very credible source (n = 65). At the same time, 41.71% of students rated Wikipedia as not credible or not at all credible (n = 78); as such, Wikipedia may represent a controversial source for students. Both blogs (M = 2.60, SD = 1.24) and Twitter (M = 2.21, SD = 1.36) were rated as lowest in credibility.

Table 5 includes descriptive information about the frequency with which students reported engaging in a variety of source evaluation behaviours. The verification behaviour students reported most commonly engaging

<table>
<thead>
<tr>
<th>Table 3. Reported frequency of source access.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>Journal Article</td>
</tr>
<tr>
<td>Textbook</td>
</tr>
<tr>
<td>Wikipedia</td>
</tr>
<tr>
<td>Research Centre Report</td>
</tr>
<tr>
<td>Newspaper</td>
</tr>
<tr>
<td>Public Opinion Survey</td>
</tr>
<tr>
<td>Blog</td>
</tr>
<tr>
<td>Twitter</td>
</tr>
</tbody>
</table>
in was checking information currency \( (M = 5.26, SD = 1.57) \), followed by considering whether information in text represents fact or opinion \( (M = 5.07, SD = 1.65) \) and considering the extent to which information was complete or comprehensive \( (M = 5.07, SD = 1.43) \). Interestingly, considering who the author was or evaluating author credentials (e.g. check to see who the author is) did not fall into students’ more frequently reported verification behaviours. Only 28.34% of students reported checking the author of a source often or very often. The least commonly employed verification strategy was checking author qualifications or credentials \( (M = 3.64, SD = 1.74) \). In fact, 48.13% of students reported verifying author credentials rarely or never.

**Research question 2: behavioural validation**

**Behavioural comparison**

Patterns of source access during multiple text task completion were examined, relative to the frequency with which students reported using a variety of document types for completing academic assignments. These are presented in Table 6. However, because the texts students reported using most frequently for academic tasks (i.e. textbook, journal article) were not available to students and because the library limited the variety of sources students had access to, these are presented only descriptively. The two notable findings to emerge from this descriptive information were that the newspaper was used much more frequently during task completion than its typical reported use for completing academic tasks, as was the Wikipedia article.

Students’ ratings of the credibility of various document types were compared to document type rankings, completed at post-task. See Table 7. Examining descriptive information identified both important consistencies and inconsistencies in students’ evaluations of different document types, reported independent of and embedded within an academic task. Specifically, digital source types, including Wikipedia, Twitter, and blog posts, were all rated low in terms of perceive credibility on the Information Source Credibility Scale and ranked

### Table 4. Reported credibility ratings of different document types.

<table>
<thead>
<tr>
<th>Source</th>
<th>1: Not at all credible</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7: Very credible</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textbook</td>
<td>0.00%</td>
<td>0.53%</td>
<td>0.53%</td>
<td>4.28%</td>
<td>8.56%</td>
<td>35.83%</td>
<td>50.27%</td>
<td>6.29</td>
<td>0.90</td>
</tr>
<tr>
<td>Journal Article</td>
<td>0.53%</td>
<td>1.07%</td>
<td>1.07%</td>
<td>3.74%</td>
<td>9.09%</td>
<td>31.02%</td>
<td>53.48%</td>
<td>6.27</td>
<td>1.05</td>
</tr>
<tr>
<td>Research Centre Report</td>
<td>2.67%</td>
<td>1.60%</td>
<td>2.67%</td>
<td>6.95%</td>
<td>16.58%</td>
<td>33.16%</td>
<td>36.36%</td>
<td>5.78</td>
<td>1.39</td>
</tr>
<tr>
<td>Newspaper</td>
<td>1.60%</td>
<td>2.67%</td>
<td>7.49%</td>
<td>14.44%</td>
<td>24.06%</td>
<td>31.02%</td>
<td>18.72%</td>
<td>5.25</td>
<td>1.40</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>6.95%</td>
<td>13.90%</td>
<td>20.86%</td>
<td>23.53%</td>
<td>20.86%</td>
<td>12.83%</td>
<td>1.07%</td>
<td>3.80</td>
<td>1.48</td>
</tr>
<tr>
<td>Public Opinion Poll</td>
<td>8.02%</td>
<td>26.20%</td>
<td>22.46%</td>
<td>19.25%</td>
<td>15.51%</td>
<td>5.88%</td>
<td>2.67%</td>
<td>3.36</td>
<td>1.49</td>
</tr>
<tr>
<td>Blog</td>
<td>19.79%</td>
<td>31.55%</td>
<td>27.27%</td>
<td>13.90%</td>
<td>4.81%</td>
<td>2.67%</td>
<td>0.00%</td>
<td>2.60</td>
<td>1.24</td>
</tr>
<tr>
<td>Twitter</td>
<td>39.57%</td>
<td>29.95%</td>
<td>11.23%</td>
<td>11.76%</td>
<td>4.81%</td>
<td>2.14%</td>
<td>0.53%</td>
<td>2.21</td>
<td>1.36</td>
</tr>
</tbody>
</table>

### Table 5. Reported engagement in source verification behaviours.

<table>
<thead>
<tr>
<th>Source</th>
<th>1: Never</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7: Very often</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check to see if the information is current</td>
<td>2.14%</td>
<td>6.95%</td>
<td>5.35%</td>
<td>11.76%</td>
<td>18.18%</td>
<td>33.69%</td>
<td>21.93%</td>
<td>5.26</td>
<td>1.57</td>
</tr>
<tr>
<td>Consider whether the views represented are fact or opinions</td>
<td>4.28%</td>
<td>2.67%</td>
<td>13.37%</td>
<td>10.16%</td>
<td>23.53%</td>
<td>22.99%</td>
<td>22.99%</td>
<td>5.07</td>
<td>1.65</td>
</tr>
<tr>
<td>Check to see if the information is complete or comprehensive</td>
<td>1.07%</td>
<td>3.74%</td>
<td>11.76%</td>
<td>14.44%</td>
<td>23.53%</td>
<td>29.95%</td>
<td>15.51%</td>
<td>5.07</td>
<td>1.43</td>
</tr>
<tr>
<td>Seek out other sources to validate the information</td>
<td>5.88%</td>
<td>5.88%</td>
<td>11.76%</td>
<td>13.37%</td>
<td>22.99%</td>
<td>21.39%</td>
<td>18.72%</td>
<td>4.81</td>
<td>1.74</td>
</tr>
<tr>
<td>Check to see which other sites are linked to the website you are viewing</td>
<td>8.02%</td>
<td>10.16%</td>
<td>15.51%</td>
<td>16.58%</td>
<td>20.86%</td>
<td>16.04%</td>
<td>12.83%</td>
<td>4.32</td>
<td>1.79</td>
</tr>
<tr>
<td>Check to see who the author is</td>
<td>9.09%</td>
<td>12.83%</td>
<td>12.30%</td>
<td>16.58%</td>
<td>20.86%</td>
<td>16.58%</td>
<td>11.76%</td>
<td>4.24</td>
<td>1.83</td>
</tr>
<tr>
<td>Visit the ‘about us’ page on a website</td>
<td>11.76%</td>
<td>16.58%</td>
<td>15.51%</td>
<td>13.90%</td>
<td>13.90%</td>
<td>16.58%</td>
<td>11.76%</td>
<td>3.98</td>
<td>1.94</td>
</tr>
<tr>
<td>Consider the author’s goals and objectives for posting information</td>
<td>12.30%</td>
<td>13.37%</td>
<td>13.37%</td>
<td>17.65%</td>
<td>24.06%</td>
<td>13.37%</td>
<td>5.88%</td>
<td>3.91</td>
<td>1.76</td>
</tr>
<tr>
<td>Check the qualifications or credentials of the author</td>
<td>12.30%</td>
<td>19.25%</td>
<td>16.58%</td>
<td>16.58%</td>
<td>19.79%</td>
<td>10.16%</td>
<td>5.35%</td>
<td>3.64</td>
<td>1.74</td>
</tr>
</tbody>
</table>
lowest in trustworthiness following task completion. In this study, the newspaper article was considered to be more trustworthy than the analysis essay, introduced to students as a report written by a professor for a research centre; in an a-priori survey measure, students rated these documents reciprocally in terms of their relative credibility. Students may have preferred the newspaper article over the analysis essay, within the context of the task used in this study because the prompt assigned to students was a current events task. This suggests that students’ trustworthiness evaluations are impacted by both task-specific and document type factors. 

Finally, students’ endorsements of various items on the Credibility Assessment Scale were compared to performance. Based on students’ responses to the Credibility Assessment Scale, three different types of source evaluation behaviours were examined: (a) whether or not students accessed texts’ document information, (b) whether or not students considered authors’ credentials, and finally (c) whether or not students revisited texts, as a proxy for corroboration.

**Document information access**

Whether or not students viewed document information for each source accessed was first compared to students’ endorsement of three items on the Credibility Assessment Scale: (a) check to see if the information is current, (b) check to see who the author is, and (c) visit the ‘about us’ page on a website. However, the Cronbach’s alpha reliability for this three-item scale was only modest (α = 0.63). Students’ average scores on this subscale were correlated with the proportion of texts for which students accessed document information. However, no significant association was found, r(187) = −0.08, p = .27.

To further examine the correspondence between reported and observed document information access, the accessing of document information was dichotomised. Specifically, students were split into two groups according to whether or not they had accessed document information for any source used during task completion. An independent samples t-test was used to examine whether students viewing document information for any source accessed differed in their reported examination of document information; however, the two groups were not found to significantly differ, t(182) = 0.31, p = .58.

**Accessing author credentials**

The second set of survey items validated the extent to which students reporting to access author credentials did so during task completion. Two items on the credibility assessments scale were used as an author credentials subscale: (a) check the qualifications or credentials of the author and (b) consider the author’s goals/objectives for posting information. The two-item scale had a Cronbach’s alpha reliability of 0.66.

Correlation analysis did not determine a significant association between students’ reporting to and demonstrating accessing of author credentials during task completion, r(185) = 0.10, p = .17. Because so few students accessed author credentials during source use, this variable was dichotomised, to reflect students either viewing author credentials for any source accessed or not. Independent sample t-tests were used to examine whether

---

**Table 6.** Comparing reported and demonstrated text access.

<table>
<thead>
<tr>
<th>Document type</th>
<th>Mean (7-point scale)</th>
<th>Percent accessing often or always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper</td>
<td>2.44</td>
<td>24.06% (n = 45)</td>
</tr>
<tr>
<td>Research Centre Report</td>
<td>2.80</td>
<td>32.62% (n = 61)</td>
</tr>
<tr>
<td>Public Opinion Poll</td>
<td>2.14</td>
<td>10.70% (n = 20)</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>3.41</td>
<td>50.27% (n = 94)</td>
</tr>
<tr>
<td>Twitter</td>
<td>1.71</td>
<td>11.23% (n = 21)</td>
</tr>
<tr>
<td>Blog</td>
<td>2.11</td>
<td>11.76% (n = 22)</td>
</tr>
</tbody>
</table>

**Table 7.** Comparing reported and demonstrated trustworthiness rankings.

<table>
<thead>
<tr>
<th>Document type</th>
<th>Mean (7-point scale)</th>
<th>Percent rating sources as credible or very credible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper</td>
<td>5.25</td>
<td>73.80% (n = 138)</td>
</tr>
<tr>
<td>Research Centre Report</td>
<td>5.78</td>
<td>86.10% (n = 161)</td>
</tr>
<tr>
<td>Public Opinion Poll</td>
<td>3.36</td>
<td>24.06% (n = 45)</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>3.80</td>
<td>34.76% (n = 65)</td>
</tr>
<tr>
<td>Twitter</td>
<td>2.21</td>
<td>7.49% (n = 14)</td>
</tr>
<tr>
<td>Blog</td>
<td>2.21</td>
<td>7.49% (n = 14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document type</th>
<th>Mean trustworthiness ranking (6-point scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper Article</td>
<td>1.93 (SD = 1.10)</td>
</tr>
<tr>
<td>Analysis Essay</td>
<td>2.49 (SD = 1.31)</td>
</tr>
<tr>
<td>Public Opinion Poll</td>
<td>2.96 (SD = 1.30)</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>3.41 (SD = 1.26)</td>
</tr>
<tr>
<td>Twitter</td>
<td>5.21 (SD = 1.17)</td>
</tr>
<tr>
<td>Blog</td>
<td>3.41 (SD = 1.26)</td>
</tr>
</tbody>
</table>

Note: For mean trustworthiness rankings, lower scores correspond to higher rankings of source trustworthiness.
students viewing document information for any source accessed reported considering author credentials to a greater extent on the Credibility Assessment Scale. The two groups did not significantly differ, \( t(183) = -0.78, p = .44 \).

**Corroboration**

The third set of source use behaviours, compared in reference to students’ self-reports, concerned students’ corroboration or comparison of information presented across texts. To examine students’ self-reported engagement in corroboration, three items from the Credibility Assessment Scale were examined: (a) seek out other sources to validate the information, (b) check to see what other sites link to the web site you are viewing, and (c) check to see if the information is complete/comprehensive. The three-item scale was found to have a Cronbach’s alpha reliability of 0.68.

Students’ scores on the corroboration subscale of the Credibility Assessment Scale were compared to three behavioural metrics: (a) the number of unique sources students visited; (b) the total number of sources students visited, including revisits, and (c) the number of sources students revisited at least once. The association between students’ scores on the corroboration items of the Credibility Assessment Scale and the total number of unique sites visited was not significant, \( r(187) = 0.07, p = .32 \); furthermore, there was no significant correlation between students’ scores on the corroboration items and the total number of texts they visited, including revisits, \( r(197) = -0.11, p = .15 \).

Because the majority of students revisiting texts were only revisiting a single source, this variable was categorised into three groups: (a) students revisiting no sources, (b) students revisiting one source, and (c) students revisiting more than one source. We examined the extent to which these three groups differed in their reported scores on the corroboration subscale; however, an analysis of variance comparing the three groups on CAS scores was not significant, \( F(2, 184) = 0.84, p = .44 \).

Examining three distinct behavioural components of credibility assessment: (a) considering document information, (b) examining author credentials, and (c) corroborating information across sources, determined that none of these behaviours were meaningfully associated with students’ reports on a questionnaire survey.

**Research question 3: cognitive validation**

Students’ reports on the Credibility Assessment Scale were also compared to the justification criteria they cited for rankings of text trustworthiness, following task completion. Students were asked to rank and provide a ranking justification for every text accessed. In comparing responses on the CAS, we examined whether stronger endorsements of particular CAS items corresponded to concordant types of justifications cited for any text accessed. Seven CAS items were examined.

Independent sample \( t \)-tests were used to examine whether or not students providing each type of justification for trustworthiness rankings differed in their endorsement of corresponding CAS items. However, no significant differences were found for any of the items: (a) currency, \( p = .96 \); (b) other sources to validate information, \( p = .68 \); (c) facts or opinions, \( p = .16 \); (d) check the author, \( p = .65 \); (e) author qualifications or credentials, \( p = .96 \); (f) information is complete or comprehensive, \( p = .92 \); (g) consider author goals or objectives, \( p = .39 \). There was further no significant correlation between students’ average scores on the CAS and the total number of justifications, of each type, provided for rankings of text trustworthiness, \( p = .41 \).

**Research question 4: matching behavioural and cognitive metrics of trustworthiness**

As a final validation measure, four multiple text use behavioural measures (i.e. accessing document information, accessing author credentials, the total number of sources visited, and text revisits) were compared to the justification categories students cited for rankings of text trustworthiness. Independent sample \( t \)-tests were used to examine whether participants demonstrating particular behaviours during multiple text task completion, or not, differed in the justifications they cited for rankings of trustworthiness. Descriptive information on behavioural and cognitive measures of source evaluation is included in Table 8.

| Table 8. Behavioural and cognitive indicators of source evaluation. |  |
|---|---|---|---|---|---|---|---|
| Justification category (Cognitive) | \( M (SD) \) | Percent of students citing | Behavioural measures | \( M (SD) \) | Percent of students |
| Fact/opinion | 0.80 (1.05) | 45.69% (n = 90) | Total unique sources | 4.68 (1.54) |
| Author goals | 0.78 (1.03) | 42.64% (n = 84) | Proportion doc info. access | 0.44 (0.42) | 59.39% (n = 117) |
| Info. completeness | 0.58 (0.89) | 35.03% (n = 69) | Proportion author info. access | 0.28 (0.39) | 40.61% (n = 80) |
| Author | 0.50 (0.95) | 27.41% (n = 54) | Total sources (w/Revisits) | 5.50 (2.20) | 29.44% (n = 58) |
| Author credentials | 0.39 (0.71) | 26.40% (n = 52) | | | |
| Corroboration | 0.13 (0.56) | 8.12% (n = 16) | | | |
| Currency | 0.04 (0.23) | 3.52% (n = 7) | | | |
**Document information access**

Analyses were conducted to examine whether document information access was associated with the types of justifications students cited following task completion. The number of justifications students produced related to either author credibility or author goals was compared across students accessing document information or not. While the number of author-related justifications was not found to differ according to whether or not students accessed document information (\(p = .67\)), there were significant differences in the number of justifications related to authors’ goals that students provided, \(t(115.53) = 2.03, p < .05\), Cohen’s \(d = 0.15\), corresponding to a small effect size. Specifically, students accessing document information during task completion considered authors’ goals in justifying trustworthiness rankings significantly more frequently (\(M = 0.99, SD = 1.22\)) than did students not accessing this information (\(M = 0.65, SD = 0.87\)).

**Accessing author credentials**

The number of justifications associated with author qualifications was found to coincide with students accessing author credentials, \(r(132.55) = 2.69, p < .01\), Cohen’s \(d = 0.20\), corresponding to a medium effect size. While students accessing author credentials produced an average of 0.57 justifications (\(SD = 0.61\)) associated with author qualifications, those not accessing author credentials cited significantly fewer such justifications (\(M = 0.27, SD = 0.61\)). More generally, the total number of justifications based on author qualifications that students cited was significantly associated with the proportion of sources for which students accessed author credentials, \(r(183) = 0.22, p < .01\).

**Corroboration**

Students justifying trustworthiness rankings based on accessing multiple sources, or not, did not differ in either the number of unique sources accessed (\(p = .28\)) or in the number of total sources accessed, including text revisits (\(p = .36\)).

**Research question 5: predicting citation**

A step-wise multiple regression was used to examine the effectiveness of self-report, behavioural, and cognitive factors in predicting the number of citations in students’ responses, as a measure of multiple source use performance associated with source evaluation. Students’ average scores on the Credibility Assessment Scale were entered at Step 1. Behavioural factors were entered at Step 2; specifically, binary variables for (a) whether or not students viewed document information for each text accessed, (b) whether or not students considered author credentials for each text accessed, and (c) whether or not students’ revisited texts were entered as predictors in the model. Finally, the total number of justifications students produced for rankings of text trustworthiness, corresponding to CAS items, were entered at Step 3.

The final model was significant, \(F(5, 167) = 4.87, p < .001, R^2_{adj} = 0.10\); Step 1 to Step 2: \(\Delta R^2 = 0.11, p < .001;\) Step 2 to Step 3: \(\Delta R^2 = 0.01, p = .25\). Coefficients based on Step 3 of the model are presented in Table 9.

**Discussion**

The aim of the present study was to examine students’ reported practices with regard to source evaluation and validate these in reference to demonstrated behaviours during task completion and criteria for source evaluation cited at post-task. Overall, no associations were identified between students’ reported source evaluation practices and evaluative behaviours and cognitions. This is concerning given the wide use of self-report instruments in understanding students’ beliefs and behaviours with regard to source evaluation (Burton and Chadwick 2000; Flanagin and Metzger 2000, 2007; Fogg et al. 2001; Tillotson 2002; Metzger, Flanagin, and Zwarun 2003; Escoffery et al. 2005; Head 2007; Kim and Sin 2007; Rainie and Tancer 2007; Selwyn 2008; Shanahan 2008; Gross and Latham 2009, 2011; Hargittai et al. 2010; Head and Eisenberg 2010; Timmers and Glas 2010).

The dubious association between students’ reported and demonstrated source evaluation behaviours during task completion has, likewise, been identified in prior research (Fogg et al. 2001; Freeman and Spyridakis 2004; Ivanitskaya, Boyle, and Casey 2006; Gross and Latham 2007, 2009, 2012; Pinto 2010). Ivanitskaya, Boyle, and Casey (2006) compared students’ health-

**Table 9. Model summary predicting number of citations based on self-report, behavioural, and cognitive measures of source evaluation.**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>(B)</th>
<th>Std. Error (B)</th>
<th>(\beta)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average CAS score</td>
<td>0.16</td>
<td>0.12</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document information Access</td>
<td>0.87</td>
<td>0.29</td>
<td>0.22</td>
<td>0.003**</td>
</tr>
<tr>
<td>Author credentials</td>
<td>0.33</td>
<td>0.31</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Revisits</td>
<td>0.97</td>
<td>0.31</td>
<td>0.23</td>
<td>0.002**</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total justifications</td>
<td>-0.08</td>
<td>0.07</td>
<td>-0.09</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Note: \(F(5, 167) = 4.87, p < .001, R^2_{adj} = 0.10\); Step 1 to Step 2: \(\Delta R^2 = 0.11, p < .001;\) Step 2 to Step 3: \(\Delta R^2 = 0.01, p = .25\). Coefficients based on Step 3 of the model.
related information-seeking behaviours, assessed through both self-report and performance items, including those related to information search and source evaluation. In brief, they concluded that: ‘While the majority of students think that their research skills are good or excellent, many of them are unable to conduct advanced information searches, judge the trustworthiness of health-related websites and articles, and differentiate between various information sources.’ These same kinds of conclusions may be drawn from the present study, as well.

**Study importance**

While some prior work has examined the validity of self-report data relative to behaviours manifest during task completion (e.g. Flanagan and Metzger 2007; Hargittai et al. 2010), the present study uniquely contributes to the literature on source evaluation and efforts to validate questionnaires of verification behaviours in at least three ways. First, we validate the commonly used Credibility Assessment Scale (Hargittai et al. 2010) against a broad set of behavioural metrics. Whereas prior work has examined a comparatively limited set of source use behaviours (Flanagan and Metzger 2007; Hargittai et al. 2010), in this study, the repertoire of behaviours examined included students’ accessing of texts and text revisits as well as their consideration of document information. Second, beyond examining behavioural manifestations of source evaluation, cognitive aspects of students’ source judgements were likewise considered in this study and corroborated against students’ self-reported behaviours. In particular, we considered the extent to which the verification behaviours students reported performing frequently (e.g. considering currency) were reflected in the criteria they cited when justifying their rankings of source trustworthiness or actually evaluating sources during information use. Such methods of cognitively based corroboration for survey data have not been undertaken in prior work. However, when examining students’ manifest behaviours, no associations were found between the verification-related factors participants reported considering frequently and those they actually cited when evaluating sources within the context of an academic multiple text task. Third, the behavioural and cognitive validations of the CAS undertaken in this study are particularly compelling because they are grounded in a rich academic task. Students’ source use behaviours and criteria cited for source evaluation were examined when participants were asked to research and write about a controversial social science topic, emblematic of the types of tasks students are commonly asked to perform in academic settings. This task stands in contrast to studies of verification behaviours that have asked students to evaluate websites in a de-contextualised manner, independent of responding to a specific prompt (e.g. Fogg et al. 2001).

**Self-Reported, behavioural, and cognitive measures of source evaluation**

Given the lack of correspondence between self-reported and behavioural and cognitive measures of source evaluation, at least three different frameworks can be used to better understand or contextualise students’ survey responses. First, it may be the case that the Credibility Assessment Scale, rather than capturing students’ typical practices with regard to source evaluation, better reflects students’ information literacy self-efficacy (Kurbanoglu, Akkoyunlu, and Umay 2006; Kilic-Cakmak 2010; Akkoyunlu and Yilmaz 2011). Information literacy has been broadly defined (Maughan 2001; Eisenberg, Lowe, and Spitzer 2004; Kurbanoglu, Akkoyunlu, and Umay 2006; Katz 2007; Eshet-Alkalai and Chajut 2010) as including ‘higher-order abilities such as: assessing search results for quality and relevance; evaluating the reliability, validity, authority, and timeliness of retrieved information; and applying new information to the planning and creation of scholarly and professional pursuits’ (Pinto 2010, p. 87); indeed, it is reflected in many of the verification-related behaviours included on the Credibility Assessment Scale.

In the present study, as in prior work, students were found to be fairly optimistic about their skills with regard to source evaluation (Maughan 2001; Gross and Latham 2007; Metzger 2007; Head and Eisenberg 2010), indicating high efficacy. Although students’ own behavioural ratings were not found to correspond to manifest behaviours or cognitions, a critical disconnect, these results may nevertheless be viewed somewhat optimistically. Increased levels of efficacy have been associated with positive performance outcomes across a number of domains (instruction: Bandura 1993; college learning: Chemers, Hu, and Garcia 2001; math: Hackett and Betz 1989; Pajares and Graham 1999; writing: Pajares and Johnson 1996; job performance: Stajkovic and Luthans 1998). Nevertheless, it should be noted that within the domain of information literacy, although Gross and Latham (2007) found students’ perceptions of their abilities and demonstrated performance to be positively correlated, this association was not strong to a statistically significant extent. More generally, students’ self-reports of the frequency of their engagement in verification-related behaviours may simply reflect students’ knowledge of the criteria and behaviours they ought to be engaging when searching for and using information.
online. These desired behaviours have commonly been emphasised in information literacy curricula and research skills trainings (Scholz-Crane 1998; Coiro 2003; Meola 2004; Sanchez, Wiley, and Goldman 2006).

Second, the analysis provided by Gross and Latham (2009), when validating a measure of students’ reported source evaluation practices relative to manifest behaviours, seems to apply to our study as well. Specifically, Gross and Latham (2009) suggest that the disconnect between students’ reported and demonstrated source evaluation behaviours represents a disconnect between judgements based on process or product. Questionnaires, like the CAS, aiming to capture the frequency with which students perform a host of source evaluation behaviours, are focused on process or the methods whereby students may ascertain the quality and accuracy of sources. As a contrast, when students search for information, they may be focused on product, or whether or not they are able to effectively resolve their specific queries of interest. When asked to report on their process of information use, students may over-report the evaluative behaviours they engage in because they perceive themselves to be successful at formulating a task product. In other words, because students are typically able to appropriately resolve search queries, they may incorrectly perceive themselves to be engaging in quality information use processes. This misalignment in focus may explain the discrepancies between students’ reported and demonstrated source evaluation practices. Indeed, such an explanation seems to be supported by behavioural metrics constituting the only significant predictors of source evaluation performance (i.e. citation).

Finally, the discrepancy between students’ reported and manifest behaviours may reflect a difference in scope. Numerous studies have suggested that students can evaluate information according to both source-related features (e.g. author) and content quality (e.g. argument strength; Chu and Kamal 2008; Hahn, Harris, and Corner 2009; Mason, Boldrin, and Ariasi 2010; Stadtl er and Bromme 2014). At the same time, students have been found to evaluate texts according to source-related factors relatively infrequently (Britt and Aglinskas 2002; Gerjets, Kammerer, and Werner 2011; Braasch et al. 2012; Stadtl er and Bromme 2014). Although under-examined in prior work, this may, rather, indicate that students instead evaluate texts based on content. Surveys and self-report measures may not be sufficiently sensitive to detect when students are evaluating texts according to source or content features. So, when responding to specific survey items, students may be reporting their general dispositions towards source evaluation rather than the frequency with which they engage in the specific behaviours the questionnaire items are targeting. This interpretation suggests the need to use cognitive interviews (Willis 2004) or other qualitative methods to better understand students’ survey responses.

**Self-report measures**

Of course, limitations typically associated with the use of self-report measures may also apply. These include items cuing students’ over-reporting and students’ responses reflecting a desirability bias (Fisher 1993; Van de Mortel 2008). Additionally, accurately reporting typical source evaluation behaviours on a survey requires that students be self-regulated and meta-cognitively engaged during information search. However, we know that students are often not meta-cognitively active when looking for information on the Internet and often need explicit prompting to engage in or monitor their source evaluation (Hofer 2004; Azevedo 2005; Stadtl er and Bromme 2007, 2008; Lazonder and Rouet 2008; Mason, Boldrin, and Ariasi 2010). Indeed, the low validity ascribed to questionnaire data in this study may simply reflect students’ difficulties with accurately tracking their engagement in verification-related behaviours, even at the undergraduate level.

As a final note, there is reason to not be entirely pessimistic about the use of surveys in assessing students’ source evaluation, a necessary methodology for large-scale studies (e.g. Burton and Chadwick 2000; Fogg et al. 2001; Metzger 2007; Hargittai et al. 2010). First, survey results from this study were found to be highly consistent with findings in prior work (Flanagin and Metzger 2000, 2007; Metzger, Flanagin, and Zwarun 2003; Metzger 2007), indicating promising reliability. For instance, in looking at students’ credibility assessments of different document types, our study corroborates prior work emphasising students’ beliefs in the trustworthiness of textbook sources (Bråten, Stømso, and Britt 2009; Bråten, Stømso, and Salmeron 2011). Furthermore, in our study as well as in five prior waves of data collection, dating back to 1999, students reported checking the currency of information and attempting to distinguish fact and opinion much more frequently than they did verifying authors’ credentials or qualifications, the least commonly reported verification behaviour, across samples (Metzger 2007).

The CAS was further able to capture differences in the frequency with which students reported engaging in various verification behaviours, like visiting the about us page on a website or checking the completeness of information. Students reported engaging in the latter behaviour much more frequently than they did the former. This distinction in the frequency with which students cited engaging in various source evaluation behaviours
is notable. Bråten, Strømsø, and Salmerón (2011) found students with lower levels of prior knowledge to be distinguishable from their higher-knowledge counterparts by the justification criteria they cited for source evaluation. Specifically, students low in prior knowledge had a more difficult time distinguishing between relevant and irrelevant source evaluation criteria, whereas high-knowledge students based their trustworthiness ratings on factors like publisher and document type (e.g. textbook, newspaper) more frequently than they did on currency-related criteria (i.e. date of publication). Given similar work indicating that students high and low in performance on a multiple text task may be distinguishable according to the criteria they report using when evaluating sources (Wiley et al. 2009; Goldman et al. 2012), it may be that between-subjects differences in students’ survey responses are important to consider in understanding source evaluations, even given their low correspondence to demonstrated verification behaviours.

Performance

In this study, the number of citations in students’ responses was used as an indicator of source evaluation performance. Elsewhere, the number of citations has commonly been used as a metric of both multiple text task performance and evaluation (e.g. Britt and Aglinskis 2002; Wiley and Voss 1999). Although not a primary focus of this study, it is noteworthy that only behavioural measures were found to directly correspond to performance. This seems to indicate that while students’ actual engagement in verification behaviours may be supported by a variety of factors (e.g. cognitions about source trustworthiness, personal definitions of credibility), it is only behaviour that ultimately corresponds to task performance. Nevertheless, future work should consider how both behavioural and cognitive aspects of source evaluation support various facets of multiple text task performance.

Limitations

A number of limitations must be considered in association with this study. First, we were interested in validating an existing measure commonly used to tap students’ behaviours associated with source evaluation, the Credibility Assessment Survey (Hargittai et al. 2010). While emblematic of other self-report measures (Cameron, Wise, and Lottridge 2007; Katz 2007; Gross and Latham 2009; Pinto 2010), the CAS is only one instrument; other measures of source evaluation have been developed and require further validation. Additionally, because the CAS was an existing measure, not all behaviours manifest during task completion (e.g. citation) could be captured in the survey; conversely, not all items on the CAS were represented in the behaviours logged during students’ completion of a multiple text task.

The CAS was also a survey limited in focus on the behaviours students engage in when searching for and evaluating information. A broader survey framework may be needed to capture a wider set of constructs associated with source evaluation, such as students’ efficacy for information verification (Kurbanoglu, Akkoyunlu, and Umay 2006). In particular, Hilligoss and Rieh (2007) introduce an expansive framework for understanding students’ credibility assessments. They consider students’ judgements of source credibility to include: students’ (a) definitions of the construct of credibility, (b) heuristics commonly used when evaluating sources, and (c) interactions between text-based and contextual features that may result in students evaluating information in more deliberate (i.e. non-heuristic) ways. The CAS may be considered to tap only aspects of the second and third dimensions Hilligoss and Rieh (2007) identify. Meanwhile, some of these source evaluation components, for instance students’ perceptions of the construct of credibility, may better lend themselves to self-report-based assessment than do students’ behaviours (Rieh and Danielson 2007).

Further consideration should be given to the scales or units used when asking students to report the frequency with which they engage in verification behaviours. Findings from this study suggest that asking students to endorse Likert-scale items corresponding to frequency ratings may negatively contribute to measure validity. Asking students whether or not they typically engage in behaviours associated with credibility assessment on a more restricted or binary scale (e.g. yes/no) may improve the validity of questionnaire data and curb students’ tendencies to over-estimate their engagement in verification behaviours. Similar checklist measures have been used in early studies of text evaluation (e.g. Fritch and Cromwell 2001; Meola 2004; Doyle and Hammond 2006).

Finally, as suggested previously, more qualitative methods may be effective in improving our understanding of the dimensions or constructs captured by the CAS and similar surveys. These may also help to understand the contextual or task-embedded nature of source evaluation (Hilligoss and Rieh 2007). Using a phenomenological approach, Gross and Latham (2011) characterised differences in students’ source evaluation when individuals pursued imposed, or assigned, vis-a-vis self-generated, or personal, queries. Among the differences, Goss and Latham identified between searches in each of these contexts was students’ adherence to standards of...
information quality. When searching for information for personal reasons, students could freely choose whether to consider the quality of information sources or not; as a contrast, academic tasks imposed standards of information quality on students’ search. Further work, examining students’ source evaluations across a variety of contexts and tasks, may be needed to additionally validate the CAS alongside other measures of source evaluation. Nevertheless, this study represents an important step in integrating self-reported, behavioural, and cognitive approaches to understanding source evaluation.

Conclusion

In this study, we corroborate the association between students’ reported engagement in source evaluation, manifest source evaluation behaviours during a multiple text task, and cognitively based justifications for rankings of text trustworthiness following task completion. No association was found between self-report and behavioural and cognitive measures of source evaluation (RQ2 & RQ3). However, a limited but important association was identified between behavioural measures of source evaluation (i.e. accessing document information; accessing author credentials) and the types of justifications students cited when ranking text trustworthiness (RQ4). This indicates that while students’ behaviours and cognitions about source evaluation may be somewhat aligned, these are not necessarily well-captured by self-report measures. In predicting the number of citations in students’ responses, as a performance-based measure of evaluation, only behavioural indicators proved significant (RQ5). Collectively, these findings indicate at least three implications for research and practice. First, when using self-report measures of source evaluation, it is advisable to corroborate these against behavioural and cognitive measures. Moreover, data from this study suggest that relying exclusively on self-report to assess students’ source evaluation may be an insufficient metric – a key implication for further research in this field. Second, only teaching students about effective strategies for source evaluation may be insufficient to spark behavioural change. In other words, it appears to be the case that students, while knowing what to do to evaluate sources, do not practice what is preached (Paul et al. 2017). Third, supporting students’ reasoning about source trustworthiness, a cognitive facet of source evaluation, may be a promising avenue for increasing students’ engagement in source evaluation behaviours. The association between behavioural and cognitive measures of source evaluation identified in this study shows further promise in providing a framework for how students’ competencies in source evaluation may be assessed. It seems that asking students to reason about the trustworthiness of various texts corresponds to the extent to which they may engage in such reasoning during the course of text processing and, indeed, can be used to stimulate source evaluation behaviours during multiple text use. As a whole, this study provides evidence for a multifaceted approach to developing and assessing source evaluation.

Disclosure statement

No potential conflict of interest was reported by the authors.

References


Kammerer, Y., D. G. Amann, and P. Gerjets. 2015. “When Adults Without University Education Search the Internet for Health Information: The Roles of Internet-Specific


