Trust But Verify: Examining the Association Between Students' Sourcing Behaviors and Ratings of Text Trustworthiness

Alexandra List, Patricia A. Alexander & Lori A. Stephens

To cite this article: Alexandra List, Patricia A. Alexander & Lori A. Stephens (2016): Trust But Verify: Examining the Association Between Students' Sourcing Behaviors and Ratings of Text Trustworthiness, Discourse Processes, DOI: 10.1080/0163853X.2016.1174654

To link to this article: http://dx.doi.org/10.1080/0163853X.2016.1174654

Accepted author version posted online: 11 Apr 2016.
Published online: 11 Apr 2016.

Submit your article to this journal

Article views: 83

View related articles

View Crossmark data
Trust But Verify: Examining the Association Between Students’ Sourcing Behaviors and Ratings of Text Trustworthiness

Alexandra Lista, Patricia A. Alexander, and Lori A. Stephens

ABSTRACT

Three indicators of undergraduate students’ (n = 197) source evaluation were investigated as students completed an academic task requiring the use of multiple texts. The source evaluation metrics examined were students’ (1) accessing of document information, (2) trustworthiness ratings, and (3) citation in written responses. All three indicators of source evaluation were found to differ across source type. Based on a significant interaction between source type and accessing of document information in predicting trustworthiness ratings, an additive model of trustworthiness evaluation is proposed. Specifically, we suggest that students access document information to confirm a priori, heuristic judgments of trustworthiness. A model using all three source evaluation metrics was found to be significant in predicting response quality, with the number of citations included in students’ responses serving as the strongest predictor in the model.

Introduction

Scientists trace heat wave to massive star at center of solar system. The same headline attributed to the New York Times or the satirical newspaper, The Onion, would hold a dialectically different meaning for readers. Attention to the attribution of texts, or considering a text’s metadata, such as publisher or author (i.e., document information), has been referred to as sourcing. Wineburg’s (1991) seminal study identified sourcing as a key text use heuristic distinguishing the multiple text integration of expert historians and more novice advanced placement history students. Following Wineburg’s work in history (1991), sourcing has been examined across topics and domains (e.g., climate change: Bråten, Strømsø, & Britt, 2009; Bråten, Strømsø, & Salmerón, 2011; medical information: Mason, Junyent, & Tornatora, 2014; Sanchez, Wiley, & Goldman, 2006; Stadtler & Bromme, 2007) and associated with students’ performance on tasks requiring the use and integration of multiple texts (Britt, Wiemer-Hastings, Larson, & Perfetti, 2004; Rouet, Britt, Mason, & Perfetti, 1996; Strømsø, Bråten, & Britt, 2010).

Britt and colleagues introduced the Documents Model of Multiple Source Use framework to articulate the mechanisms whereby sourcing results in improved multiple text comprehension and integration (Britt, Perfetti, Sandak, & Rouet 1999; Britt, Rouet, & Braasch, 2013). The Documents Model framework suggests that multiple text comprehension involves representing or modeling information, found across texts, in two ways: through intertext and situations models of multiple texts (Bråten, Britt, Strømsø, & Rouet, 2011). In constructing an intertext model, students represent the relations among texts and the information within them. This involves the simultaneous development of (1) source–content links, to associate specific pieces of information with metadata about their sources of...
origin (e.g., author), (b) content–content links, to trace information found across texts, and (c) source–source links, to represent points of concurrence and discrepancy across texts (e.g., authors agreeing or disagreeing; Britt et al., 1999; Perfetti et al., 1999). When students construct a situations model, they represent multiple texts’ perspectives on a common topic or issue in a single, unified structure. Situations models are developed by integrating multiple texts or deciding which pieces of information, found across sources, to focus on and which to potentially dismiss in developing a consolidated understanding.

Integrating multiple texts (i.e., constructing both intertext and situations models) relies on students’ sourcing or connecting specific information in a text with metadata about the source it originates from (e.g., author, publisher; Britt et al., 2012). Specifically, sourcing underlies students’ judgments of text trustworthiness. These trustworthiness judgments, then, serve to determine which authors and information students trust and therefore privilege in constructing an integrated situations model and which sources students defer to when texts conflict. Sourcing supports the integration of information found to be congruent across texts, while allowing discrepant information to be qualified as representing disagreements between authors (Britt, Rouet, & Braasch, 2012).

Substantive empirical evidence is available to support the association between sourcing and students’ reconciliation of disagreements across texts (Bråten, Ferguson, Strømsø, & Anmarkrud, 2014; Stadtler & Bromme, 2014; Stadtler, Scharrer, Brummernhenrich, & Bromme, 2013). Indeed, recent investigations into the Content-Source Integration Model (Stadtler & Bromme, 2014) and the Discrepancy-Induced Source Comprehension assumption (Braasch, Rouet, Vibert, & Britt, 2012) have found sourcing to increase when students recognize conflicts or disagreements across texts. Conversely, when students do not engage in sourcing, it is more difficult for them to identify and reconcile discrepancies (Stadtler & Bromme, 2014). Further, ignoring the attribution of information may lead students to integrate inaccurate or biased information into their mental models or to develop incomplete mental representations of situations presented across multiple texts (Britt et al., 1999, 2013; Stadtler & Bromme, 2014).

A number of studies have examined both the association between sourcing and judgments of text trustworthiness (Anmarkrud, Bråten, & Strømsø, 2014; Bråten, Braasch, Strømsø, & Ferguson, 2015; Bråten et al., 2009) and sourcing and text integration, particularly in response to texts presenting conflicting information (Bråten, Ferguson, Strømsø, & Anmarkrud, 2014; Stadtler & Bromme, 2014; Stadtler, Scharrer, Brummernhenrich, & Bromme, 2013).

Across these studies, two sets of results have repeatedly emerged. First, students, even at the undergraduate level, have been found to engage in sourcing without prompting only to a limited extent and to require support in accessing and using document information (Britt & Aglinskas, 2002; Stadtler & Bromme, 2007; Strømsø, Bråten, Britt, & Ferguson, 2013). Second, students’ ratings of text trustworthiness have been found to differ across source types (Bråten et al., 2009, 2011; Rouet et al., 1996). However, more work is necessary to integrate these findings. Specifically, limited work has considered the extent to which students electing to access a particular text’s document information may be associated with source type. Additionally, the extent to which students’ accessing of document information, or choosing not to do so, may be associated with ratings of text trustworthiness requires further investigation.

In this study, we seek to further elucidate the relation between students’ sourcing and judgments of text trustworthiness and examine their association across source types. We do this by using a cued sourcing protocol, whereby students are provided with the option of accessing document information for each source they select within a digital library interface, but they must actively elect to do so. Using this methodology, we are able to determine the frequency with which students access document information across various source types and how students’ trustworthiness ratings of various source types differ according to whether or not texts’ document information is accessed. Further, although prior work has examined the sourcing of traditional source types (i.e., those with a print analog, like newspapers), the present study examines document information access for traditional and digital (i.e., source types unique to Internet consumption, like Twitter and Wikipedia) source types alike.
Defining sourcing

Wineburg (1991) first used the term *sourcing* to identify a heuristic that seemed to distinguish the multiple text engagement of expert historians and advanced placement high school students, comparatively more novice in the history domain. In Wineburg’s formulation sourcing referred to historians’ practice of looking first at a source or attribution of a document, before reading, and using information about attribution to predict, interpret, and evaluate document content. Beyond the domain of history, sourcing has been used more generally to refer to students’ association of information or arguments in text with their sources of origin (Stadtler & Bromme, 2007)—an important skill for learners to develop.

More operationally, Strømsø and Bråten (2014) define sourcing as having three components: attending to texts’ document information, using document information in interpreting texts’ content and judging text trustworthiness, and citing document information in written responses. Prior work has independently considered each of these indicators of sourcing, particularly focused on students accessing of document information and trustworthiness evaluations (e.g., Britt & Aglinskas, 2002; Britt et al., 2004; Bråten et al., 2009, 2011; Strømsø, Bråten, Britt, & Ferguson, 2013). Nonetheless, more research is needed to consider all three of these indicators of sourcing in interaction with one another and as jointly impacting students’ task performance.

Frequency of sourcing

Despite its theorized importance, a number of studies have raised concerns about the frequency with which students engage in sourcing when accessing texts (Britt & Aglinskas, 2002) and use document information in evaluating sources (Brem, Russell, & Weems, 2001; Kuiper, Volman, & Terwel, 2005). For instance, Britt and Aglinskas (2002) found that after using six documents, college students correctly sourced only between 16.2% (i.e., control condition) and 18.7% (i.e., prompted sourcing condition) of texts. Walraven, Brand-Gruwel, and Boshuizen (2009), looking at a sample of upper secondary students, found that they evaluated information presented in only 15% of webpages visited and evaluated sources for only 0.5% of sites accessed. More optimistically, Rouet et al. (1996) found that when students composed essays based on multiple historical documents, 52% of essays included at least one specific reference to a text, indicating sourcing. However, given these discrepant numbers, more work is needed to determine how commonly students engage in sourcing.

To explicitly examine the frequency with which students access document information, Gerjets, Kammerer, and Werner (2011) compared students’ sourcing across two conditions, when students were instructed to evaluate texts (i.e., instructed evaluation condition) or not (i.e., spontaneous evaluation condition). As might be predicted, students in the instructed evaluation condition gazed longer at websites’ document information (i.e., publisher information, source information, and user ratings) than did students in the spontaneous evaluation condition, according to eye-tracking data. Further, Gerjets et al. (2011) determined that in evaluating both search results and webpages, students in the instructed evaluation condition produced significantly more think-aloud utterances associated with text evaluation, particularly those addressing website credibility.

Generally, findings have commonly documented differences in the frequency of students’ sourcing between control conditions, where students’ sourcing is spontaneous, and experimental conditions, where students are explicitly trained or instructed to access texts’ document information (e.g., Britt & Aglinskas, 2002; Stadtler & Bromme, 2008). In the present study, a middle-ground between documenting students’ spontaneous sourcing and explicitly instructing students to evaluate texts (e.g., Britt & Aglinskas, 2002; Gerjets et al., 2011) was adopted. Specifically, students were cued to the option of accessing document information for each text visited; however, they received no explicit instruction to do so. Such a protocol has not been examined in prior research but is a potentially promising indicator of students’ document information use.
Evaluating text trustworthiness

Although learners may be infrequent or inconsistent in their sourcing (Britt & Aglinskas, 2002; Rouet et al., 1996), attending to document information has been considered to underlie students’ judgments of text trustworthiness (Bråten et al., 2009, 2011; Brem et al., 2001; Rouet et al., 1996). Research has looked at text trustworthiness as the evaluative dimension capturing students’ general assessments of source quality (Bråten et al., 2009, 2011; Rouet et al., 1996; Strømsø, Bråten, & Britt, 2011; Wiley et al., 2009). At the same time, defining trustworthiness has proven difficult. Brazilai and Zohar (2012) considered students’ trustworthiness ratings to be composed of judgments of text credibility, accuracy, reasonableness, and the degree of support offered for the arguments provided. In turn, Flanagin and Metzger (2007) suggest that credibility ratings may themselves be composed of students’ judgments of as many as 22 dimensions, including source believability, reliability, authoritativeness, accuracy, informativeness, professionalism, bias, and organization, among others. Stadtler and Bromme (2014) offer a simpler suggestion: that judgments of source trustworthiness reflect students’ determinations of author expertise, or ability to provide accurate information, and author benevolence, or determination to prove accurate and unbiased information to the reader. Determinations of author expertise and benevolence may be inferred when students access a text’s document information and can be confirmed as students process information in text (Stadtler & Bromme, 2014).

Assessing sourcing and trustworthiness evaluations

A variety of methods have been forwarded to assess students’ sourcing and resulting trustworthiness evaluations. Commonly, students have been asked to recall source information after their completion of a task requiring the use of multiple texts (Braasch, Rouet, Vibert, & Britt, 2012; Strømsø et al., 2010). Rouet et al. (1996) asked students to rank seven documents according to trustworthiness and to justify their rankings. They coded students’ justifications into four categories: those based on text content, their own opinions, author, or document type; with the later two categories reflecting sourcing information. Generally, content-based justifications were reported significantly more frequently than justifications based on author or document type. Using a similar methodology, Bråten et al. (2009) asked students to rate the trustworthiness of six texts on climate change and to report the extent to which they relied on five sourcing characteristics (i.e., author, title, date, publisher, source type) in determining text trustworthiness. Students most often justified trustworthiness ratings based on text content; further, justifications were more often based on considerations of document type and publisher than on author and date of publication.

Think-aloud studies have also been used to examine how students evaluate texts during, rather than after, source use (e.g., Gerjets et al., 2011; Mason, Ariasi, & Boldrin, 2011). Walraven et al. (2009) coded students’ think-aloud utterances as pertaining to information or source evaluation. Although emphasizing students’ limitations in spontaneous sourcing and text evaluation, results from think-alouds reinforce the association between students’ consideration of document information and text evaluations (Gerjets et al., 2011; Mason et al., 2011; Walraven et al., 2009).

In sum, prior work has adopted one of two approaches to assessing students’ sourcing in association with evaluations of text trustworthiness. The first approach has been to examine students’ spontaneous consideration of document information and text evaluation, as they think aloud (Britt & Aglinskas, 2002; Gerjets et al., 2011; Stadtler & Bromme, 2008). The second approach has been to ask students to retrospectively justify text trustworthiness ratings and to examine the extent to which justifications include references to document information or not (Bråten et al., 2009; Wiley et al., 2009).

Missing from these analyses has been an association between whether or not students access texts’ document information and resulting trustworthiness ratings. The cued document information access protocol adopted in this study allowed us to compare students’ text trustworthiness evaluations when document information was accessed or not.
Trustworthiness ratings across source types

Among the key findings to emerge from investigations of sourcing and text evaluation has been that students are able to distinguish between various source types (Rouet et al., 1996) and, more specifically, that students’ trustworthiness evaluations differ by source type. In the domain of history, students rate primary sources as more trustworthy than they do secondary sources (Rouet et al., 1996). Across domains, undergraduate students consider textbooks to be highly trustworthy sources, whereas more expert source users (e.g., graduate students, historians) do not (Bråten et al., 2009, 2015; Rouet et al., 1996; Wineburg, 1991). Further, the extent to which students are able to discriminate between high-reliability (e.g., NASA, PBS) and low-reliability (e.g., commercial sites) sources in their trustworthiness ratings has been associated with learning from online inquiry tasks (Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012; Sanchez et al., 2006; Wiley et al., 2009).

Sourcing and trustworthiness evaluation in digital contexts

The associations between sourcing and trustworthiness evaluations, rendered across source types, have most commonly been examined for print sources (e.g., textbooks, newspapers) presented either on paper or digitally (Bråten et al., 2009; Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991). Researchers have further begun to examine sourcing on the Internet. Although some have found students’ ratings of website quality to be consistent with students’ judgments of more traditional print texts (e.g., Keck, Kammerer, & Starauschek, 2015), it may be the case that sources on the Internet present students with unique evaluative challenges. When using online sources, document information may be more difficult to locate, traditional markers of credibility (e.g., publisher) may be less present, and the volume of available information may discourage students from engaging in deliberative and effortful sourcing behaviors (Britt & Gabrys, 2002; Coiro, 2003; Gerjets et al., 2011; Metzger & Flanagin, 2003). Such problems may further be exacerbated for new, digital source types or texts created online for online consumption, such as blogs. For instance, Wikipedia, a digital source, intentionally obfuscates the authorship of articles to cultivate a collective authoritative voice (Miller, 2005). At the same time, sourcing and evaluating information may be all the more important in digital contexts where low reliability sources exist alongside and can easily be made to look like high reliability sources (Coiro, 2003). Students have been found to place a high premium on relevance, in the context of Internet source use, to reduce the volume of information they may need to consider in responding to a query; sourcing and judgments of trustworthiness may be an additional mechanism students can use to cull the volume of information any Internet search necessarily generates (Kii, Laurinen, & Marttunen, 2015). Nonetheless, limited work has examined how sourcing and trustworthiness ratings may function when students are confronted with new, digital source types alongside more traditional print-based texts. In the present study, we examine students’ accessing of document information and trustworthiness evaluations of both traditional (e.g., analysis essay, newspaper article) and new, digital (e.g., blog, Wikipedia, Twitter) document types.

Present study

The present study had three primary goals. First, given that prior research has reported large variations in the frequency with which students access document information, we were interested in using a cued sourcing protocol to determine the frequency with which students accessed document information (Britt & Aglinskas, 2002; Kii, Laurinen, & Marttunen, 2015). Although studies have varied in either explicitly asking students to evaluate texts or examining students’ spontaneous sourcing (Gerjets et al., 2011; Stadtler & Bromme, 2008), in the present study students were provided with the option of accessing document information via a hyperlink when they accessed any text in the digital library. This served to cue, although not explicitly instruct, students to source and allowed us to determine how frequently and for which sources students accessed document information.
Second, students were asked to rate the trustworthiness of each source accessed. A cued response protocol allowed us to compare students’ trustworthiness ratings both across document types and according to whether or not students had elected to access a text’s document information. Prior methods have either prompted students to report the extent to which they considered document information in formulating trustworthiness ratings (Bråten et al., 2009, 2011) or counted the number of times students referenced document information in justifying trustworthiness ratings (e.g., Mason et al., 2011; Rouet et al., 1996). A cued sourcing protocol allowed for a direct behavioral measure of sourcing to be used in understanding students’ trustworthiness evaluations.

Third, as today’s students are increasingly called to evaluate and integrate texts that are both traditional sources, with print analogues, and digital sources, created online for online consumption, we were interested in comparing the frequency of students’ accessing not only across source types but also for digital versus traditional texts. Additionally, we were interested in examining the association between students’ accessing of document information or not and trustworthiness evaluations, across both traditional source types and across traditional vis-à-vis digital texts.

We had the following research questions and associated hypotheses. Our first research question was descriptive, examining the nature of students’ sourcing (i.e., accessing of document information, trustworthiness ratings, citation) when completing a multiple source use (MSU) task. Prior work has examined rates of document information access when students were provided with an intervention to promote sourcing or not (Britt & Aglinskas, 2002) or otherwise instructed to access document information (Gerjets et al., 2011). Likewise, students’ rates of spontaneous citation, without specific instruction, have begun to be examined (Strømsø et al., 2013). We generally expected students’ accessing of document information and citation to be comparable with what has been found in prior research, however, no specific rates of sourcing were hypothesized.

The second research question examined differences in students’ sourcing behaviors according to document type. Specifically, we were interested in examining the association between document type and students’ accessing of document information and citation behaviors. Given that prior work has found students to consider document type in rendering trustworthiness judgments (Bråten et al., 2009, 2011; Rouet et al., 1996), we expected document type-based differences to extend to students’ sourcing and citation behaviors. Further, given concerns about students’ sourcing in online contexts (Coiro, 2003; Metzger & Flanagin, 2003), we expected rates of document information access and citation to be lower for digital source types (e.g., blog, Wikipedia) than for document types with print analogs (e.g., analysis essay, newspaper article).

The third research question considered the associations among sourcing behaviors (i.e., accessing document information, trustworthiness ratings, citation) and the differences in these associations across document types. Although each of these sourcing measures has previously been associated with text-based task performance, the associations among these sourcing indicators have been examined to a more limited extent. For instance, Britt and Aglinskas (2002) investigated citation behaviors after students’ participation in an intervention to improve sourcing (i.e., Sourcer’s Apprentice). However, the direct association between accessing document information and citation was not examined. Although we generally expected all sourcing behaviors to be associated with one another, this research question was considered to be more exploratory in nature.

The fourth research question examined the association between various sourcing behaviors (i.e., accessing of document information, trustworthiness ratings, citations) and response quality. Based on prior research as well as the Documents Model framework (Britt et al., 1999; Perfetti et al., 1999) all three measures of sourcing (i.e., accessing document information, trustworthiness evaluations, citation) were expected to be associated with task performance. Considering the importance of source features (e.g., author) in multiple text integration, we were especially interested in the role of document information access and citation in response quality (Britt & Aglinskas, 2002; Stadtler & Bromme, 2007, 2008).

Although the association between sourcing behaviors and task performance is well established (Britt & Aglinskas, 2002; Britt et al., 1999; Perfetti et al., 1999; Stadtler & Bromme, 2007, 2008), limited prior
work has examined the associations among various indicators sourcing and how these may jointly impact response composition. Therefore, we believed a more exploratory study of these variables was warranted. Specifically, in this study our primary concern was in establishing the associations among students’ accessing of document information, trustworthiness ratings, citation behaviors, and performance, with further experimental work to follow.

Methods

Participants

Participants were 197 undergraduate students enrolled at a large Mid-Atlantic university (age: \(M = 20.47; SD = 2.08\)). The sample was mostly women (65.48%, \(n = 129\); men: 29.95%, \(n = 59\)) and racially and ethnically diverse. Specifically, 49.23% of participants were White (\(n = 97\)), 19.29% were Asian (\(n = 38\)), 16.24% identified as African American (\(n = 32\)), 3.55% were Latino (\(n = 7\)), and 6.09% self-identified as biracial or multiracial (\(n = 12\)). Nine participants declined to provide demographic information.

The sample included participants reflecting a range of class standings, including freshmen (21.32%, \(n = 42\)), sophomores (21.83%, \(n = 43\)), juniors (27.41%, \(n = 54\)), seniors (19.80%, \(n = 39\)), and advanced students, enrolled in courses beyond their senior year (\(n = 5.08%, n = 10\)). Most students were social science majors (56.06%, \(n = 111\)), with others majoring in the natural sciences (27.27%, \(n = 111\)), the humanities (5.05%, \(n = 10\)), and undecided in their majors (6.60%, \(n = 13\)). The sample reported a grade point average of 3.26 (\(SD = .53\)).

Measures

The study was carried out over two sessions. In Session 1, participants were asked to complete a variety of individual difference measures, including a prior knowledge assessment of the MSU topic, the Arab Spring in Egypt. In Session 2, participants completed an MSU task. This involved researching a prompt using a library of six texts (i.e., research phase) and generating a response to the target prompt based on texts accessed (i.e., response phase). Participants completed Session 1 online at a time and location of their choosing. Session 2 was completed online in a lab setting with the researcher present. Data were collected in spring 2015, as the Arab Spring in Egypt was an ongoing issue in the popular press.

Session 1

In Session 1 participants completed a term-identification prior knowledge measure and were briefly asked about their attitude with regard to the target topic, the Arab Spring in Egypt.

Prior knowledge. The prior knowledge measure was a seven-item short-answer, term-identification task. Specifically students were asked to please tell me about key terms (i.e., Arab Spring, Muslim Brotherhood, Tamarod), people (i.e., General el Sisi, Hosni Mubarak, Mohamed Morsi), and places (i.e., Tahrir Square) associated with the Arab Spring in Egypt. Each item was scored as correct or incorrect, based on whether or not the term was correctly and distinctly identified. Total scores on the prior knowledge measure ranged from zero to seven. The reliability for the measure was \(\alpha = .91\). The sample had limited prior knowledge, with mean knowledge scores of 2.23 (\(SD = 2.61\)) on a seven-point scale. The prior knowledge measure is presented descriptively to characterize the target sample and used as a control variable in predicting response quality.

Attitudes. Students’ attitudes toward the topic of the task were assessed via two items. The first item mirrored the performance task and asked students to select whether they support Morsi or el Sisi or to specify I don’t know. The second item asked students about their attitudinal valence. Specifically, students were asked to rate the item, how strong do you feel about events associated with the Arab Spring in Egypt? Students in our sample had limited attitudes with regard to the topic of the task. Specifically,
94.29% of students selected the *I don’t know* option when asked to designate who should hold power in Egypt, and students’ mean attitude valence was only a 2.05 (SD = 1.54) on a seven-point scale.

**Session 2**

In Session 2, students were asked to complete a MSU task including a *research phase* and a *response phase*.

**Research phase.** In the research phase, students were asked to use a library of six digital sources to research the following prompt: *Should the United States support General el-Sisi and the military regime or Mohamed Morsi and the Muslim Brotherhood?* Specifically, students were instructed as follows: *Please answer as you would if assigned to write a brief essay in response to this prompt for an academic class. In responding to the prompt you will be asked to take a position (i.e., in support of Mohamed Morsi or General el-Sisi or an alternative) as well as to provide specific evidence to support your position.* Additionally, students were told they could take notes during the research phase and would have access to their notes, although not the texts themselves, in composing their responses. There was no time limit for the task.

**Source library.** The digital library included six texts, ranging in document type (i.e., blog post, analysis essay, newspaper, public opinion poll, Twitter, and Wikipedia). Texts also varied in author expertise, the nature of evidence provided, and the position adopted, with some texts written in a neutral tone (i.e., newspaper, Wikipedia) and others advocating for Morsi (i.e., analysis essay, Twitter) or el Sisi (i.e., public opinion poll, blog).

The Arab Spring in Egypt was selected as the target topic, in part, because it allowed for document types traditionally considered to be low in reliability (e.g., blog) to be presented alongside more traditionally academic document types (e.g., analysis essay by a historian) as legitimate sources for students to draw on in composing their responses. Indeed, digital media (e.g., Twitter) was found to play a particularly important role in the Arab Spring in Egypt (Khondker, 2011; Lotan, Graeff, Ananny, Gaffney, & Pearce, 2011).

All texts were relevant to the target prompt. Texts were arranged as an array in the digital library and presented to students by document type (e.g., newspaper). Order of presentation was randomized for each participant.

**Sourcing.** During the research phase, students’ sourcing was captured in two ways. The first was a behavioral measure of document information access. When students accessed a text, only the content of the text was visible. At the top of each text was a hyperlinked button, “Click here to learn more about this source.” Clicking the button revealed document information about the text (i.e., title, author, publisher, date and location of publication, and URL). Whether or not students elected to access document information about each text accessed was recorded. Proportion scores were computed reflecting the ratio of the number of texts for which document information was accessed to the total number of texts students used.

Second, text-specific trustworthiness ratings were captured. After participants had completed using a text, indicated by their selecting to return to the library, they were asked to rate text trustworthiness on a scale from 0 to 100.

**Response phase.** During the response phase, students were asked to identify their position as in favor of Mohamed Morsi, General el Sisi, or an *Other* option as well as to justify their position.

**Scoring.** Students’ justifications were scored using two primary metrics. First, responses were scored according to the number of *arguments* (i.e., claims with justifications) included. Second, the number of *evidentiary justifications* provided to support each argument was used as a metric of response quality. Evidentiary justifications referred to the distinct number of reasons or evidence, examples, or explanations provided to support a particular claim. Table 1 includes sample response coding. In the
examples claims are bolded, justifications are numbered, and citations are underlined. Inter-rater reliability was determined based on scoring 18 participant responses (9.14%). Cronbach’s alpha reliability for the number of arguments in students’ responses was .73. Reliability for the number of elaborative justifications in students’ responses was $\alpha = .89$.

**Sourcing.** Finally, the total number of citations included in participants’ justifications was recorded. Citations were explicit references to texts in the digital library, either by source type (e.g., newspaper) or by document information (e.g., author). Descriptive information is provided for the percent of students citing each available source at least once in Table 2. However, in analyses each instance of citation was counted toward students’ total citations, regardless of whether or not students were citing the same source multiple times. Participants were not explicitly instructed to cite texts, so citation metrics represent *spontaneous* citation (Strømsø et al., 2013). All students’ responses ($n = 197$) were scored by the first and third author to identify instances of citation. Cohen’s kappa inter-rater reliability for the number of citations included in students’ responses was $\kappa = .77$, $p < .001$.

## Results

**MSU task overview**

In completing the MSU task, participants accessed an average of 4.67 sources ($SD = 1.54$) of the six texts available, with 46.19% of participants ($n = 100$) accessing all six texts. Students devoted an average of 15.46 minutes to source access ($SD = 9.22$) and used each source for an average of 3.31 minutes ($SD = 2.63$).
Research question 1: Sourcing during MSU

To answer the first research question, three indicators of sourcing were examined: students’ accessing of document information, trustworthiness ratings, and citation. Most subjects (i.e., 59.39%) accessed document information for at least one source accessed. On average, students accessed document information for 44.28% of sources used ($SD = 42.56\%$). Participants’ average rating of text trustworthiness was 53.40 ($SD = 14.26$) on a 100-point scale. Students cited an average of 1.62 sources ($SD = 1.94$) in their responses, ranging from 0 to 10 citations included. Overall, 55.33% of students cited at least one text in composing their responses. On average, 26.47% of texts accessed were cited in students’ responses ($SD = 29.51\%$).

Research question 2: Sourcing across source types

For the second research question, sourcing indicators (i.e., document information access, trustworthiness ratings, and citation) were examined across source type.

**Document information.** Pearson’s chi-square test of association was used to determine whether there was a correspondence between source type and accessing of document information. The chi-square was significant, $\chi^2(5, n = 921) = 11.54, p < .05$, Cramer’s $V = .11$, corresponding to a small effect size. Because post-hoc analyses could not be run for chi-square, adjusted standardized residuals were examined. These suggested that students were significantly less likely to access document information for Wikipedia ($se_{adj} = 2.05$). Further, students being more likely to access document information for the analysis essay ($se_{adj} = 1.79$) and being less likely to do so for Twitter ($se_{adj} = 1.87$) made substantial contributions to the significant association between source type and document information access. Table 2 includes descriptives of the percent of students accessing document information, across source types.

**Trustworthiness ratings.** One-way analysis of variance (ANOVA) was run to determine the extent to which students’ ratings of text trustworthiness differed across source types. The ANOVA was significant, $F(5, 915) = 90.19, p < .001$, $\eta^2 = .33$, corresponding to a large effect size. Tukey’s HSD post-hoc analyses determined that students rated the analysis essay ($M = 64.03, SD = 24.24$) as significantly more trustworthy than the blog post ($M = 30.17, SD = 22.63; p < .001$), Twitter ($M = 31.40, SD = 23.42; p < .001$), and Wikipedia ($M = 48.73, SD = 24.94; p < .001$). Students rated the newspaper article ($M = 71.20, SD = 20.37$) as significantly more trustworthy than the blog post ($p < .001$), Twitter ($p < .001$), and the analysis essay ($p < .05$). Likewise, the public opinion poll was scored as significantly more trustworthy than the blog post ($p < .001$), Twitter

<table>
<thead>
<tr>
<th>Source</th>
<th>Percent Accessing</th>
<th>Percent Access Document Information</th>
<th>Average Trustworthiness</th>
<th>Percent Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blog</td>
<td>61.42%</td>
<td>51.24%</td>
<td>$M = 30.17$</td>
<td>14.88%</td>
</tr>
<tr>
<td>(n = 121)</td>
<td></td>
<td>$SD = 22.63$</td>
<td>(n = 18)</td>
<td></td>
</tr>
<tr>
<td>Essay</td>
<td>79.19%</td>
<td>53.85%</td>
<td>$M = 64.03$</td>
<td>33.97%</td>
</tr>
<tr>
<td>(n = 156)</td>
<td></td>
<td>$SD = 24.24$</td>
<td>(n = 53)</td>
<td></td>
</tr>
<tr>
<td>Newspaper</td>
<td>94.42%</td>
<td>46.77%</td>
<td>$M = 71.20$</td>
<td>39.25%</td>
</tr>
<tr>
<td>(n = 186)</td>
<td></td>
<td>$SD = 20.37$</td>
<td>(n = 73)</td>
<td></td>
</tr>
<tr>
<td>Public opinion survey</td>
<td>79.19%</td>
<td>52.56%</td>
<td>$M = 64.53$</td>
<td>29.49%</td>
</tr>
<tr>
<td>(n = 156)</td>
<td></td>
<td>$SD = 22.29$</td>
<td>(n = 46)</td>
<td></td>
</tr>
<tr>
<td>Twitter</td>
<td>76.65%</td>
<td>40.40%</td>
<td>$M = 31.40$</td>
<td>17.22%</td>
</tr>
<tr>
<td>(n = 151)</td>
<td></td>
<td>$SD = 23.42$</td>
<td>(n = 26)</td>
<td></td>
</tr>
<tr>
<td>Wikipedia</td>
<td>76.65%</td>
<td>39.74%</td>
<td>$M = 48.73$</td>
<td>15.23%</td>
</tr>
<tr>
<td>(n = 151)</td>
<td></td>
<td>$SD = 24.94$</td>
<td>(n = 23)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>44.28%</td>
<td>$M = 53.26$</td>
<td>$SD = .42$</td>
<td>27.96%</td>
</tr>
</tbody>
</table>

Table 2. Descriptives of Source Evaluation.
(p < .001), and Wikipedia (p < .001). Finally, Wikipedia was rated as significantly more trustworthy than the blog post (p < .001) and Twitter (p < .001).

**Citation.** Pearson’s chi-square test of association was used to examine whether there was a relation between document type and source citation in students’ responses. The chi-square was overall significant, $\chi^2(5, n = 921) = 55.83, p < .001$, Cramer’s $V = .25$, corresponding to a small to medium effect size. Adjusted standardized residuals suggested that students were significantly less likely to cite the blog post ($se_{adj} = 3.04$), Twitter ($se_{adj} = 2.20$), and Wikipedia ($se_{adj} = 3.73$) in their responses, while being significantly more likely to cite the newspaper source ($se_{adj} = 6.10$).

**Research question 3: Sourcing in trustworthiness ratings and citation**

The third research question examined the relation between document information access and subsequent sourcing (i.e., trustworthiness evaluations and citations).

**Document information and trustworthiness ratings.** A two-way ANOVA was run to determine whether there were differences in trustworthiness ratings across document types (Main Effect 1) and according to whether or not a text’s document information had been accessed (Main Effect 2). The ANOVA determined a significant main effect for document information access [$F(1, 909) = 7.96, p < .01, \eta^2 = .33$] and for source type [$F(1, 909) = 90.76, p < .001, \eta^2 = .01$]. However, these main effects should be interpreted with caution as there was also a significant interaction between source type and document information access [$F(1, 909) = 4.63, p < .001, \eta^2 = .02$]. Tukey’s post hoc analyses determined that for the analysis essay (accessed: $M = 69.66, SD = 22.57$; did not access: $M = 57.46, SD = 24.63$; $p = .001$), the newspaper article (accessed: $M = 75.13, SD = 20.07$; did not access: $M = 67.74, SD = 20.11$; $p < .05$), and the public opinion poll (accessed: $M = 69.15, SD = 21.71$; did not access: $M = 59.41, SD = 21.95$; $p < .01$), accessing document information resulted in higher ratings of source trustworthiness. On the contrary, accessing document information for Wikipedia resulted in lower trustworthiness ratings (accessed: $M = 42.55, SD = 23.93$; did not access: $M = 52.81, SD = 24.87$; $p < .01$). There were no differences in trustworthiness ratings for the blog ($p = .66$) and for Twitter ($p = .20$) across students accessing document information and not.

**Document information and citation.** Pearson’s chi-square test of association was used to determine whether there was a significant association between document information access and citation. The overall chi-squared was significant, $\chi^2(1, n = 925) = 4.54, p < .05$, Cramer’s $V = .07$, and determined that students who accessed document were significantly more likely to cite sources in their responses. A three-way crosstab and chi-square tests of association was used to examine whether there was a correspondence between students accessing document information and citing sources, and source type. Although the overall association was significant, partial chi-square tests of association did not find within-source associations between document information access and citation (analysis essay: $p = .74$; blog: $p = .43$; newspaper: $p = .94$; public opinion: $p = .17$; Twitter: $p = .14$; Wikipedia: $p = .26$).

**Trustworthiness ratings and citation.** Two-way ANOVA was used to determine whether there were differences in students’ trustworthiness ratings across source types and according to whether or not texts were cited. There was a significant main effect for both source type [$F(1, 909) = 51.01, p < .001$] and citation [$F(1, 909) = 3.83, p = .05$] but no significant interaction effect ($p = .65$). Texts cited in students’ justifications were rated as significantly more trustworthy ($M = 54.88, SD = 1.93$) than texts not cited ($M = 50.75, SD = .85$).

**Research question 4: Source evaluation and response quality**

The final research question examined the association between source evaluation metrics (i.e., accessing of document information, trustworthiness evaluations, and citations) and response quality. Two metrics of response quality were considered. Specifically, evaluation metrics were used to predict the
number of distinct arguments students generated as well as the number of justifications or evidentiary reasons included in students’ responses.

**Number of arguments.** A hierarchical linear regression was used to predict the number of arguments included in students’ responses. After controlling for prior knowledge in Step 1, the number of texts for which document information was accessed was entered in Step 2. This number was dependent on both the total number of texts students accessed and the number of texts for which document information was accessed. In Step 3, students’ maximum trustworthiness rating of any source accessed was entered. Maximum trustworthiness scores were preferred over average trustworthiness scores as they better capture the maximum trustworthiness students ascribed to the information accessed. Finally, the total number of citations included in students’ responses was entered in Step 4.

The full model, including the number of texts for which document information was accessed, maximum trustworthiness ratings, and the total number of citations included, was significant, $F(4, 184) = 18.25, p < .001, R^2_{adj} = .27$, corresponding to a large effect. The number of texts for which document information was accessed, although a significant predictor when initially entered in Step 2 [$F(2, 186) = 3.75, p < .05$] was not a significant predictor in the full model, $p = .76$. Students’ maximum ratings of source trustworthiness was also not a significant predictor in the model [$B = .01, SE(B) = .01, \beta = .10, p = .14$]. However, the total number of citations included [$B = .36, SE(B) = .05, \beta = .50, p < .001$] was a significant predictors of the total number of arguments included in students’ responses. After controlling for prior knowledge, document information accessed, and students’ maximum trustworthiness ratings, the partial correlation between the total number of citations included and number of arguments generated was $r = .50$, indicating a large effect. **Table 3** shows a correlation among all variables in the model, and **Table 4** shows a summary of the full regression model.

**Number of elaborative justifications.** Hierarchical linear regression was used to predict the number of justifications included in students’ responses based on metrics of source evaluation. Specifically, the number texts for which document information was accessed was entered in Step 2, after prior knowledge was controlled for in Step 1. Step 3 included students’ maximum trustworthiness ratings, whereas Step 4 included the total number of citations included in students’ responses.

The full model was significant, $F(4, 184) = 18.92, p < .001, R^2_{adj} = .28$, indicating a large effect. The number of documents for which information was accessed was a significant predictor in the model, $B = .29, SE(B) = .14, \beta = .14, p < .05$. Controlling for other variables in the model, the partial correlation between document information and the number of elaborative justifications included in students’ responses was .15, corresponding to a small effect size. Students’ maximum ratings of text trustworthiness were neither significant when entered in Step 3, $p = .54$, or included in Step 4 of the model, $p = .58$. However, the number of citations included was a significant predictor in Step 4, $B = 1.14, SE(B) = .16, \beta = .45, p < .001$. Controlling for prior knowledge, accessing of document information, and maximum trustworthiness ratings, the partial correlation between the number of citations and the number of evidentiary justifications included in students’ responses was $r = .46$, corresponding to a large effect. **Table 3** includes correlations among key variables, whereas **Table 5** includes a summary of the regression model.

**Table 3. Correlation Among Key Variables.**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prior Knowledge</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Document information accessed</td>
<td>.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Maximum trustworthiness ratings</td>
<td>.07</td>
<td>.34***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. No. of sources cited</td>
<td>.11</td>
<td>.23**</td>
<td>.13</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. No. of arguments</td>
<td>.12</td>
<td>.16*</td>
<td>.16*</td>
<td>.51***</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>6. No. of justifications</td>
<td>.19**</td>
<td>.26***</td>
<td>.17*</td>
<td>.51***</td>
<td>.72***</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01, ***p < .001.*
Discussion

This study adopted a cued sourcing protocol to examine the frequency with which students accessed document information and to compare students’ ratings of text trustworthiness according to whether or not document information was accessed. Students’ accessing of document information and trustworthiness ratings were compared across document types, for both tradition (e.g., analysis essay, newspaper, public opinion poll) and digital (e.g., blog, Twitter, Wikipedia) source types. Additionally, students’ accessing of document information and trustworthiness ratings were associated with citation, an additional sourcing metric, and overall task performance.

Research question 1: Sourcing during MSU

The first research question examined the prevalence of sourcing, including the frequency with which students accessed document information for texts used and cited sources in generating written responses. Over half of participants were found to access document information for at least one text used (59.39%) and to cite at least one text in their written responses (53.30%).

Directly comparing these figures to what has been identified in the literature previously is a challenge. Prior work has focused on capturing sourcing either through think-alouds or by asking students to justify post-hoc evaluations of text trustworthiness (e.g., Bråten et al., 2009; Rouet et al., 1997; Stadtler & Bromme, 2008). As a result, rates of sourcing are typically reported as the percentage of think-aloud utterances or written justifications that make reference to document information. For example, in asking undergraduates to justify their rankings of text trustworthiness, Rouet et al. (1997) found 44.5% of justifications to reflect considerations of document information, whereas Perfetti, Britt, Rouet, Georgi, and Mason (1994) found 33% of trustworthiness justifications to cite considerations of author. In this study, because sourcing was assessed behaviorally (e.g., whether or not students elected to access document information), we examined, among other factors, the percentage of students engaging in particular sourcing behaviors or not.
A limited number of comparisons for such student-based, rather than response-based, metrics of sourcing are available in the literature. In a study of spontaneous sourcing, Strømsø et al. (2013) found 59.3% of participants to explicitly refer to document information during a think-aloud. However, most of these utterances reflected only attention to document information (67.2%), with only 27.3% of source-related utterances resulting in trustworthiness evaluations. Examining citation, Perfetti et al. (1994) found 39% of students to reference at least one text in composing written responses based on secondary sources; when primary sources were included among available documents, 65% of respondents referenced at least one text. Although further investigation is needed to track students’ spontaneous accessing of document information and citation, the rates of sourcing identified in this study are consistent with prior work (e.g., Perfetti et al., 1994; Strømsø et al., 2013).

The present study adopted behavior-based metrics to disambiguate rates of sourcing during text engagement (i.e., document information access) from sourcing during response composition (i.e., citation). More specifically, whereas students were cued to access document information, they received no explicit instruction to do so and were not directed to cite sources in their written responses. Although the measures used in this study reflect spontaneous sourcing, as demonstrated by Strømsø et al. (2013), simply accessing document information does not guarantee its use in determining text trustworthiness. Likewise, the mere inclusion of citations does not necessarily indicate that students evaluated texts’ trustworthiness or considered such evaluations in formulating written responses.

**Research question 2: Sourcing across document types**

The second research question examined whether students differed in their accessing of document information, trustworthiness ratings, and citation across document types. All three of these metrics were found to differ by document type, extending prior research focusing on differences in trustworthiness ratings across source types (e.g., Bråten et al., 2009, 2011; Britt et al., 2013; Rouet et al., 1996).

Participants rated the newspaper as the most trustworthy source of the six texts available in the digital library. This was an unexpected finding, as the newspaper article was included as a source moderate in reliability. When Bråten et al. (2009) presented students with newspaper articles, alongside other texts, students rated the newspapers as lower in trustworthiness than both a policy research report, considered to be a high reliability source, and a textbook. However, the newspaper articles used by Bråten et al. (2009) were from liberal and conservative dailies. We presented students with a newspaper article attributed to the Associated Press, recognized as an esteemed and objective media source. The newspaper article used in the present study included critiques of both Morsi and el Sisi; had it evidenced bias, students may have rated this sources lower in trustworthiness. Additionally, the disciplinary relevance of the newspaper article may have served to bolster students’ ratings of the text’s trustworthiness. In other words, students may have considered the newspaper article to be the most trustworthy source to use in responding to a prompt about a current international issue.

Digital source types (i.e., blog, Twitter, and Wikipedia) were all rated as significantly less trustworthy than the newspaper article, the analysis essay, and the public opinion poll. On the one hand, this may be expected given that all three of these were nontraditionally academic document types, limited in authoritativeness. At the same time, these were sources that presented first-hand accounts of events on-the-ground (e.g., blog post) and, in the case of Twitter, were attributed to official sources involved in the controversy (i.e., Muslim Brotherhood). Although these sources expressed clear bias, they also provided direct and current accounts of events in Egypt. Indeed, prior work has found students to consider primary sources, even when biased, to be high in trustworthiness (Perfetti et al., 1994; Rouet et al., 1996). The low trustworthiness ratings assigned to the blog post and Twitter may suggest that students were not viewing these sources as primary texts. This may have been because the task addressed a contemporary topic rather than a clearly historical issue. As suggested by Brem et al. (2001), the most likely explanation for students’ low trustworthiness ratings of digital source types may be that they were applying a heuristic approach to source evaluation. Specifically, students may have been relying on
document type as an easy indicator of trustworthiness rather than balancing competing considerations of direct experience and currency vis-à-vis bias and authority in determining trustworthiness ratings. More generally, context-specific nature of document types needs to be acknowledged. This study was conducted with undergraduates in the United States, who were expected to have certain perceptions of the document types included in the text library (e.g., Wikipedia: Ioannou & Artino, 2009; Lim, 2009). However, students in other learning contexts or cultures may have different perceptions and assumptions about document type. For instance, newspapers in the United States have been found to differ in character from their European counterparts (Esser, 2008; Esser & D’Angelo, 2006; Esser, Reinemann, & Fan, 2001). This may serve to explain some of the discrepancies in relative trustworthiness ratings found between appraisals of the newspaper article in this study and those found in prior research (Bråten et al., 2009, 2011). Likewise, students in the United States and particularly participants low in prior knowledge may not have fully appreciated that Twitter constituted an uncensored communication outlet and mobilization tool for detractors in autocratic countries (Passini, 2012; Russell, 2011). This type of insight required students to have an understanding of both document type and political context, related to the topic of the task. Additional work is needed to understand how perceptions of document type differ across cultural and social contexts.

In addition to being considered less trustworthy than more traditional source types, the blog post, Twitter, and Wikipedia were also sourced to a more limited extent. Students were significantly less likely to access document information when using Twitter or Wikipedia and significantly less likely to cite the blog post, Twitter, and Wikipedia in their responses. There are at least three possible explanations for this results pattern. First, it may be the case that digital source types, perhaps because they are not traditionally academic texts, simply receive less evaluative attention. This means that in addition to being rated lower in terms of trustworthiness, these texts are also less likely to be sourced (i.e., have document information accessed) and cited. A secondary explanation may be that these limitations in evaluative behaviors are causal. In other words, students may not feel the need to access document information for texts they already consider to be low in trustworthiness. Or, students may not want to cite texts they know to be low in trustworthiness and therefore may not feel the need to access such texts’ document information.

Third, it may the case that students are lacking in the skills needed to evaluate digital source types and incorporate such texts into response composition. Determining the authorship and authenticity of some digital text types (e.g., Twitter, Wikipedia) is a challenge (Jessen & Jørgensen, 2011; Layton, Watters, & Dazeley, 2010; Rowley & Johnson, 2013). Because of the unique difficulties that sourcing digital texts presents (Coiro, 2003), students may elect not to access such texts’ document information or not to cite such texts in their responses. It may be that instructors’ calls to limit students’ use of nontraditional academic sources (Colón-Aguirre & Fleming-May, 2012; Lin, Hoffman, & Borengasser, 2013; Wannemacher & Schulenburg, 2009) has resulted in a skills deficit, whereby students are limited in their abilities to evaluate and use such texts for academic purposes. Had students wanted to draw on information provided by Twitter in composing their response, they may not have had the skills necessary to contextualize Twitter information and incorporate it into an argument or to properly cite a Twitter account.

It is important to note that although students’ evaluative behaviors (i.e., document information access, citation) were more limited for digital texts, students accessed such sources at the same rates they did more traditional academic source types (e.g., analysis essay, public opinion poll) (Table 2). This may suggest that students are not abstaining from using digital texts, rather they are just more limited in their evaluation of such sources. Research Question 3 provides further insight into how source evaluation may differ across document types.

Research question 3: Sourcing in trustworthiness ratings and citation

From a theoretical perspective, judgments of text trustworthiness have been considered to be based on students’ accessing and consideration of texts’ document information (Bråten et al., 2009, 2011; Rouet
et al., 1996; Wiley et al., 2009). As a result, prior work has focused on examining the extent to which students’ trustworthiness ratings are variably justified by document features like author or publisher (Bråten et al., 2009, 2011; Stahl, Hynd, Britton, McNish, & Bosquet, 1996). What was unique to this study was the opportunity to examine students’ trustworthiness ratings when document information about particular sources was accessed or not. An interesting pattern emerged. Specifically, there was a significant interaction between accessing document information and source type and resultant trustworthiness ratings. For texts considered to be low in trustworthiness (e.g., blog post, Twitter), accessing document information or not did not result in differences in trustworthiness ratings. As a contrast, for texts considered to be high in trustworthiness (e.g., newspaper, analysis essay, public opinion poll) accessing document information increased ratings of text trustworthiness. This seems to suggest an additive model of trustworthiness evaluation. In other words, data seem to support the conclusion that students form general judgments of text trustworthiness based on heuristics such as source type (Brem et al., 2001). Then, in accessing document information, if students determine a text to be written by an authoritative source, they consider it to be all the more trustworthy, resulting in a significantly higher rating of text trustworthiness. At the same time, students do not seem to discount texts they already considered to be low in trustworthiness for having document information indicating bias (e.g., Twitter) or limited authoritativeness (e.g., blog post).

Alternately, it may be the case that after making an initial heuristic determination of text trustworthiness (see Brem et al., 2001), students seek out document information for sources they consider to be trustworthy to confirm their judgments. However, when sources are determined to be low in trustworthiness, students may not feel the need to engage in any additional evaluative behaviors, such as accessing document information (see Research Question 2). Additionally, whereas students may have experience evaluating sources typically used for academic tasks, they may be unaccustomed to considering document information for sources like Twitter or accessing author information for Wikipedia.

There was a strong association between students’ trustworthiness ratings and citation patterns. Students cited sources they considered to be trustworthy. Regardless of whether students were simply adhering to academic conventions or truly privileging information they considered to be trustworthy, this is an encouraging finding. At the same time, failing to cite sources like the blog post or Twitter may indicate that students have difficulties with appropriately contextualizing information from primary sources (Wineburg, 1991). Although it is preferable to always use information from highly reliable sources, students also need the skills to be able to draw relevant information from imperfect sources, while tempering their reliance on such texts. In this case, the blog post and Twitter could have provided relevant and meaningful information to inform students’ responses, despite being biased sources, low in authoritativeness. Incorporating information from such controversial texts into their response likely presents students with a greater evaluative, argumentative, and writing challenge than dismissing such texts outright. At the same time, the contextualization and flexible use of such imperfect text is a marker of expert source use (Wineburg, 1991). Further, it can be argued that sourcing, or critically examining texts’ document information, is all the more important when texts are of seemingly low, rather than high, trustworthiness. In the case of the analysis essay or the public opinion poll, document type may have served as a sufficient heuristic of text trustworthiness (Brem et al., 2001). However, as far as digital texts (e.g., blog, Twitter) are concerned, examining document information may provide crucial contextualization to aid in evaluating the trustworthiness of such sources. For instance, examining the editorial history on Wikipedia may be necessary when evaluating entries about controversial topic, as Wikipedia has been found to be edited in real time, as events unfold (Brandes & Lerner, 2008; Rosenzweig, 2006). In this study, examining the attribution of the Twitter account revealed it to be the English language Twitter of the Muslim Brotherhood, a potentially a critical primary source to consider in understanding the Arab Spring. Indeed, texts, like Twitter, were specifically included in this study to provide students with mixed cues with regard to their trustworthiness.

In addition to their association with ratings of text trustworthiness, there was a correspondence between document information access and citation. This may be explained in at least two ways. First,
students who were diligent in their source use likely both accessed document information about texts and cited sources in the responses. More functionally, it may be the case that accessing document information provided students with the information they needed to cite sources in their writing (e.g., author name). However, it should be noted that students did not need to access a text’s document information in order to be able to cite it in their responses. Citations referencing texts only by document type (e.g., references to the newspaper article), as texts were presented in the source library, were likewise counted toward the number of citations included in students’ responses.

Another explanation for the association between document information access and citation behaviors is that this association serves as an indicator of corroboration. Specifically, it may be the case that when accessing novel information, students consider a text’s document information and cite this information in composing their responses. However, when accessing sources, such as Wikipedia, only to corroborate information a different behavioral pattern may emerge. When using sources only for corroboration, students may see no need to access those texts’ document information or to cite those sources in their responses. As such, it may be the case that the expert text evaluation strategies identified by Wineburg (1991)—sourcing and corroboration—are used more iteratively by novice learners. Specifically, when there is no prior information along which to evaluate a text’s claims, students may access its document information; however, when a text confirms claims made in earlier sources, allowing for corroboration, students may see no further need for document information access or citation. For students, accessing document information, corroboration, and citation, are likely not default behaviors but rather contextually dependent standards along which information may be evaluated.

Research question 4: Sourcing and performance

The three metrics of source evaluation (i.e., accessing document information, trustworthiness ratings, citation) were used as predictors of response quality (i.e., number of arguments and evidentiary justifications included in students’ responses). Both full models, including all three source evaluation metrics, and partial models, including each variable entered in a stepwise fashion, were significant in predicting the number of arguments and justifications included in students’ responses. However, the number of citations included in students’ responses was by far the strongest predictor of response quality.

This may have been the case for at least four reasons. First, citations, although a source evaluation metric, directly originated in participants’ responses. Further, citation may be considered to be the culminating evaluation indicator, encapsulating both students’ accessing of document information and resulting trustworthiness ratings. Second, statistically speaking, as students were not specifically instructed to source, there was great variability in this measure. In fact, almost half of students elected not to cite any sources in their response, making citation an effective predictor of response quality. Third, the relation between citation and response quality may be explained in terms of individual factors. Specifically, those students who did include citation information in their responses may have been diligent source users—looking to document information for evidence of source quality and drawing on texts for evidence in responding to the prompt. Finally, more elaborative responses, including more arguments and justifications, may have allowed students more room to incorporate citations as well.

Unexpectedly, trustworthiness ratings had a limited effect on response quality. In part, this may be due to the limited variability in students’ maximum ratings of text trustworthiness. Indeed, a variety of alternate trustworthiness metrics could have been used, including ratings of individual sources, discrepancy in trustworthiness ratings (i.e., difference in highest and lowest trustworthiness ratings; Braasch, Bråten, Stromsø, & Anmarkrud, 2014), average ratings of text trustworthiness, or how trustworthy students rated the first source they accessed. However maximum ratings of text trustworthiness were used as the metric of choice because these represent, at the upmost, how trustworthy students considered the available information to be and therefore how likely they may have
been to incorporate such information into composing a written response. Accessing of document information was found to have a more pronounced effect on response quality, perhaps because of its strong association with citation.

**Study contributions**

This study contributes to the research literature in at least four ways. First, it is one of the few investigations to comprehensively examine and link the three components of sourcing identified by Strømsø and Bråten (2014), accessing of document information, trustworthiness evaluations, and citation. Second, this study uses a novel, cued sourcing protocol to assess students’ document information access. This protocol allowed for a behavioral measure of sourcing, not reliant on explicit instruction. Along with examining spontaneous citation, this protocol allowed for a more naturalistic examination of students’ sourcing while completing a multiple text task. In addition to capturing the frequency of students’ cued accessing of document information and spontaneous citation, we were able to examine the correspondence between these two sourcing behaviors. Perhaps more importantly, the cued sourcing protocol allowed for the examination of students’ trustworthiness ratings when document information was accessed and not, developing our understanding of how trustworthiness ratings may be formed.

Third, this study serves as an initial investigation of how students use and evaluate nontraditional academic source types (e.g., blog, Twitter) during task completion. Students’ engagement with these document types was comprehensively examined. We considered students’ accessing of document information when visiting these source types, their evaluation of these texts’ trustworthiness, as well as their propensity to cite these sources when composing written responses. We were further able to compare students’ sourcing behaviors when using digital texts to their sourcing of more traditional document types, with print counterparts. Finally, this study presents a model examining the relative contribution of various sourcing behaviors to overall task performance. Confirming insights from the Documents Model framework (Britt et al., 1999; Perfetti et al., 1999) and consistent with our hypothesis, we found students’ accessing of document information and citation behaviors to be significant predictors of response quality.

**Limitations**

Despite the merits of the study, a number of limitations should be acknowledged. First, this study examined a low knowledge sample. MSU and evaluation behaviors have been found to differ as a consequence of both domain and topic knowledge (e.g., Rouet et al., 1997; Wineburg, 1991). Further, work should compare the evaluation of novice vis-à-vis more experienced or expert samples. Second, as documented by Strømsø et al. (2013), simple attendance to, or in this case accessing of, document information does not mean that students use such information in formulating responses. Particularly for texts considered to be controversial, like Twitter or the blog post, more qualitatively examining students’ reasoning about the trustworthiness of such texts may be a fruitful line of investigation. In this study, only a single text was used to represent each document type (e.g., newspaper). In addition to differences in document type and other source information (e.g., author), there were a vast number of idiosyncratic difference between texts, including content, point of view, writing style, and the nature of evidence provided. Indeed, texts were intentionally selected to vary with regard to all of these dimensions. Although selecting single, idiosyncratic texts to represent document type is common in the literature (e.g., Bråten et al., 2009, 2011), it limits the conclusions that can be drawn about students’ perceptions of texts and trustworthiness ratings. Further experimental work is needed to decompose and systematically manipulate the text components that may contribute to students’ ratings of text trustworthiness.

Third, the temporal sequencing of evaluative behaviors should be further considered. Although, within the design of the study, document information was accessed before students rated text trustworthiness, students likely formed initial impressions of each source prior to accessing the texts’ document information. Likewise, the number of citations included in students’ responses was treated as
a predictor of task performance, as assessed by other metrics of response quality (i.e., number or arguments and elaborative justifications included). In reality, students’ citation occurred contemporaneously with their response composition. As such, citations may also be considered to be an outcome measure, alongside the number of arguments and elaborative justifications included in students’ responses. Beyond citation, it is difficult to capture text evaluation at each phases of the MSU process, where it may occur. For instance, in this study students were asked to rate text trustworthiness once they had completed using a source; however, we know trustworthiness evaluations to be formed iteratively and to be updated throughout source use (Brem et al., 2001; Fogg, 2003; Rieh, 2002). Further work, both theoretical and empirical, is needed to understand how students evaluate texts throughout the MSU process and how such evaluations may best be assessed.

On a related note, the present study was correlational rather than experimental in nature. Specifically, we were interested in exploring the associations among various sourcing indicators (i.e., accessing document information, trustworthiness ratings, citation), associations considered to a limited extent in prior research. Although we were successful in characterizing the nature of students’ sourcing when completing a multiple text task and associating measures of sourcing with one another, we cannot make causal claims. Future work should adopt an experimental approach to examining these associations. For instance, further experimentation can be used to determine whether actively accessing document information, as students did in this study, rather than being presented with document information, has an effect on citation behaviors and response quality. Adopting an experimental approach would further allow for multilevel modeling to examine the nested effects of students accessing particular document types as associated with trustworthiness ratings and citation behavior. Such an approach would be particularly warranted since a key limitation of this study is that the chi-squared tests of association examining the correspondence between document type and sourcing behaviors (i.e., accessing document information, citation) were non-independent. For instance, when considering the extent to which source type was associated with accessing of document information or not, document information access was likely further associated with differences among individual students. A multilevel approach, modeling both student-level and document-level factors would address this limitation. It would further help to examine students’ sourcing behaviors both aggregated across texts accessed (e.g., total number of citations) and specific to particular document types. Finally, a more qualitative examination of students’ responses is needed to examine the extent to which, for particular texts cited, students were agreeing or disagreeing with the information that text presented. This would also allow us to gain more insight into how trustworthiness ratings are iteratively developed based on considerations of document information and text content.

**Implications for future research**

In light of this study’s strengths and limitations, a number of implications for future research may be identified. First, more needs to be understood regarding students’ reasoning in considering document information, formulating trustworthiness evaluations, and citing sources in responses. Several models have been introduced to demonstrate how students may make determinations of text trustworthiness. Rieh (2002), adopting a procedural approach, examined experts’ search for information and found them to generate both predictive and evaluative judgments of text trustworthiness; although predictive judgments were made before a text was accessed, evaluative judgments were made while visiting webpages and were used to express a preference. As such, rendering trustworthiness evaluations may occur through an iterative process, where students reciprocally reference text features and update their judgments of text trustworthiness (Fogg, 2003; Rieh, 2002). Rather than examining the process whereby trustworthiness evaluations may be determined, Petty and Cacioppo (1986) consider the levels at which trustworthiness ratings may be rendered. Specifically, they suggest that students may make either heuristic determinations of text trustworthiness, based on superficial text features, or may engage in effortful, systematic processing to render a deliberative judgment of text trustworthiness. Yet both process-oriented (Rieh, 2002) and criteria-based (Petty & Cacioppo, 1996) understandings of evaluative
judgments fail to fully consider how other indicators of sourcing (i.e., document information access, citation) may be associated with students’ trustworthiness judgments. This is certainly an area for future work.

Second, the relation between students’ overall judgments of text trustworthiness and evaluations of specific arguments require further examination. As with much research on texts, it is difficult to separate students’ judgments of texts holistically, based on document type, and their judgments of specific arguments included in texts. Although prior research has devoted much attention to how students render ratings to text trustworthiness based on document type and other source features (Bråten et al., 2009, 2011; Brem et al., 2001; Rouet et al., 1996; von der Mühlen, Richter, Schmid, Schmidt, & Berthold, 2015), students no doubt judge texts based on content as well. Indeed, Lombardi, Seyranian, and Sinatra (2014) define information sources as characterize by both document-level (e.g., author trustworthiness) and content-level (e.g., argument complexity) features.

Evaluations of information sources at these two levels (i.e., document-level and content-level) may be sequential in nature. Students’ judgments of text trustworthiness, alongside their determinations of author expertise, have been found to contribute to their perceptions of content credibility (Hilligoss & Rich, 2008; Stadtler & Bromme, 2014; Tseng & Fogg, 1999) or argument plausibility (de Perera, Britt, Braasch, & Rouet, 2014; Lombardi et al., 2014). Britt, Richter, and Rouet (2014) suggest that in evaluating specific content in texts, students may potentially adopt one of three approaches including corroborating information in texts against their general world knowledge, domain-specific knowledge, and against other available documents. Similar dimensions may also be used to evaluate sources at the document level. For instance, students may use their knowledge of various publication outlets in evaluating source trustworthiness (Britt et al., 2014). More work is needed to disentangle how students evaluate both documents and the content within them and which of these evaluations may take precedence when conceptualizing texts. Jointly, these implications suggest that further work on source evaluation must examine students’ reasoning about texts as related to both task and context.

References


Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: spontaneous re-


