# EXAMINING EPISTEMIC AND NON-EPISTEMIC DIMENSIONS OF SOURCE EVALUATION

Alexandra List, Emily M. Grossnickle, and Patricia A. Alexander

University of Maryland

DRAFT

**EXAMINING EPISTEMIC AND NON-EPISTEMIC DIMENSIONS OF SOURCE**

**EVALUATION**

Evaluating sources on the Internet has been identified by researchers and education policymakers as a crucial, but particularly challenging, competency for today's 21st century learners (National Governor's Association, 2010; Wiley, Goldman, Graesser, Sanchez, Ash, & Hemmerich, 2009). The Internet context has been implicated in making source evaluation particularly difficult for students, as source information presented on a search results page often looks uniform (Gerjets, Kammerer, & Werner, 2011), and certain source markers, such as publisher and author, may be obscured in favor of relevancy indicators, such as key words (Goldman, 2011). Further, any website can be made to look professional and reliable with great ease, obfuscating the difference between reliable and unreliable sources (Alexander & Tate, 1999). Given the volume of information available on the Internet and the open nature of the online publishing culture, source evaluation has been found to be an essential skill for today's students (Brand-Gruwel & Stadtler, 2011). Thus, in this confounding and confusing digital landscape, it becomes vital to understand how students evaluate sources, and to understand how source evaluations differ across the different types of sources that students may frequently encounter.

As Bråten and Strømsø (2006) have argued, source evaluation is a skill students must often develop independently, as the Internet provides limited indicators of source reliability, and as credentialing on the Web exists only to a limited extent (Metzger, Flanagin, & Zwarun, 2003). However, empirical studies of students' source evaluation are limited. Further, the majority of those multiple source use studies have asked students to evaluate a set of sources along pre-determined dimensions (e.g., Bråten, Strømsø, & Britt, 2009), rather than considering the extent

to which students may spontaneously evaluate texts when completing tasks requiring multiple sources.

Moreover, even in cases where students' spontaneous source evaluations have been considered, studies have focused on source evaluations along epistemic dimensions (i.e., dimensions concerned with the quality or truthfulness of sources or information), such as the extent to which students considered sources to be biased or authors to be credible (e.g., Mason, Boldrin, & Ariasi, 2010b). However, it is probable that when interacting with texts while completing multiple source use or MSU tasks, learners evaluate texts not only based on epistemic dimensions, but consider non-epistemic dimensions of sources as well. For example, the extents to which sources are easy to use and relevant to the task are likely issues that enter into students' evaluations (e.g., Gerjets, Kammerer, & Werner, 2011).

In addition to understanding the nature of the evaluations that students provide, it is necessary to ascertain the role of the task and domain in MSU evaluations. Students' source evaluations have been found to vary based on a number of factors including the topic and task guiding multiple source use and the textual materials available (Bråten et al., 2009). Consistent with conceptions of MSU as a task-driven process (Anmarkrud, McCrudden, Bråten, & Strømsø, 2013), it is therefore crucial to conceptualize students' source evaluations under varying task parameters. The present investigations focused on students' source evaluation, along both epistemic and non-epistemic dimensions, with the goal of identifying how such dimensions play out in these evaluations (Study 1) and to consider how such evaluations may differ across domains and question types (Study 2).

## Theoretical Models of MSU

Students' multiple source use has been conceptualized as a procedural process, consisting of progressive steps that guide learners' interactions with multiple texts (Goldman, 2004). In one of the most prominent MSU models, the Multiple Documents Task-Based Relevance Assessment and Content Extraction Model (MD-TRACE; Britt, Perfetti, Sandak, & Rouet, 1999), this process is conceptualized as consisting of five steps, beginning with students' formulation of a task-model (Step 1); that is, their cognitive understanding of task demands and their plan to meet them. Step 2 in this process is the determination of an information need, which then serves as the motivator for MSU. Step 3, document processing, is said to consist of three sub-steps, namely (a) source selection, (b) source comprehension, and (c) information integration based on previously encountered texts. Step 4 involves the creation of a textual product, and the final step, Step 5, requires the matching of the task product to initial task demands. It is Step 3, document processing, and specifically source evaluation, that is the primary focus of the current studies. Specifically, these studies seek to illuminate how source evaluation unfolds given differing task conditions, varying by domain and question type, and potentially by document type.

### Source Evaluations

Source evaluation has been defined as judgments about a text or the information it contains based on "available or accessible information about the source" (Bråten et al., 2009, p. 6). These judgments may be based on information beyond the immediate text, including students' prior knowledge or experience, or other texts related or unrelated to the task. In particular, source evaluations may be epistemic or non-epistemic in nature. Epistemic source evaluations are text judgments related to information quality, source trustworthiness, credibility,

and reliability (Wiley et al., 2009). These evaluations may derive from considerations of document author/publisher or source type (e.g., newspaper, journal article; Bråten et al., 2009). By comparison, non-epistemic evaluations are source judgments not based on considerations of source trustworthiness, and rather capture the extent to which the relevance, usefulness, and accessibility of a source and the information within it are considered by the individual (Rouet & Britt, 2010).

**Dimensions of Source Evaluation**

Much of the literature on students' source evaluation has focused on epistemic dimensions (e.g., Mason, Ariasi, & Boldrin, 2011; Mason, Boldrin, & Ariasi, 2010a; Mason et al., 2010b; Wiley et al., 2009). For example, Bråten et al. (2009) asked participants to evaluate each of seven sources along the dimension of *trustworthiness*, and to determine the extent to which respondents based their trustworthiness rating on one of six source characteristics (i.e., author, publisher, document type, date of publication, source content, or their own opinion). Participants were asked to evaluate a range of document types (e.g., newspaper, magazine, research report) for how trustworthy they were perceived. All of the documents were presented to students on paper and no websites or digital texts were included.

However, evidence suggests that students evaluate the reliability of sources relatively infrequently when using information found online (Metzger et al., 2003; Wiley et al., 2009). This may be because students lack requisite skills for evaluating sources along epistemic dimensions (Flanagin & Metzger, 2000; Grimes & Boening, 2001). In online contexts, students have difficulty attending to source features, such as author or publisher (Britt & Aglinskas, 2002; Rouet, 2006). Moreover, when interacting with texts in digital contexts, students tend to focus on non-epistemic dimensions of sources, such as relevance or ease of use, rather than source

quality or reliability (e.g., Grimes & Boening, 2001). This focus on non-epistemic features in students' source evaluations is consistent with the ways sources are presented to students in online contexts (Kammerer, Werner, & Gerjets, 2008), with relevance emphasized by search engines and sourcing features often obscured (Metzger et al., 2003).

More recently, a limited number of studies have turned to examining students' source evaluations along both epistemic and non-epistemic dimensions. For example, Gerjets et al. (2011) identified five evaluative dimensions based on an analysis of participants' think-aloud data: two pertaining to *topicality* (i.e., topicality, scope) and three pertaining to *quality* (i.e., credibility, up-to-datedness, and site design). These five source-evaluation dimensions addressed primarily non-epistemic dimensions (i.e., topicality, scope, site design), but also reflected students' potential epistemic concerns (i.e., credibility). However, these evaluative dimensions concerned students' evaluations of source presentation of a search engine results page, rather than evaluations of the sources themselves. More work is needed to investigate how students' evaluate sources, rather than search information, along both epistemic and non-epistemic dimensions.

**Methods of Source Evaluation**

Examinations of students' source evaluations have typically provided students with a predetermined set of dimensions along which to evaluate sources (e.g., Bråten et al., 2009). However, more open-ended methods to elicit source evaluations have been introduced in the literature. For example, Wiley et al. (2009) presented students with seven digital texts to simulate a Google search results page. To assess source evaluations, students were first asked to rank the texts they were presented with based on how *reliable* they considered them to be. Participants were then asked to justify their ranking; a more open-ended methodology to assess

students' evaluations.  Based on participants' reports, eight categories of source evaluations were identified.  In a study by Mason et al. (2010a), students thought aloud during an Internet search task and these data were used to analyze source evaluations.  Students' epistemic cognitions corresponding to the *justifications for knowledge* encompassed evaluations of source content and epistemic cognitions corresponding to the *source of knowledge* concerned students' evaluations of the documents themselves (Mason et al., 2010a; Mason et al., 2011).

Although both the Wiley et al. (2009) and the Mason et al. (2010a; 2010b) studies demonstrated the importance of using less constrained methods to examine students' source evaluations, both of these studies focused solely on epistemic dimensions and did not address the non-epistemic reasons students offered for source selection or use.  Thus, the focus of Study 1 was to use an open-ended methodology (i.e., a retrospective guided interview) to uncover students' source evaluations, both epistemic and non-epistemic.  Rather than presenting students with a predetermined set of criteria along which to evaluate sources, the aim of Study 1 was to comprehensively catalog the criteria students' used in their evaluations of sources, be they epistemic or non-epistemic in nature.

Walraven, Brand-Gruwel, and Boshuizen (2009) conducted a study wherein they considered evaluations of search results, source information, and the source itself as students thought aloud while researching academic questions on the Internet.  Based on prior literature, the authors determined that students could have evaluated sources along dimensions of technicality, usability, verifiability, and reliability.  However, these researchers reported that, in reality, students' evaluations were primarily concerned with evaluating search results and that participants evaluated sources or the information within them very rarely throughout the course of the Internet search.

Gerjets et al.'s (2008) discussion of methodologies to assess students' source evaluation reached a similar conclusion. Participants' source evaluations were examined under two task conditions: prompted and spontaneous. In the prompted task condition, students were explicitly asked to engage in source evaluation while thinking-aloud, while in the spontaneous condition, students received think-aloud instructions. In the prompted task condition, participants evaluated sources significantly more frequently than those in the spontaneous task condition. As the aim of Study 1 was to maximally identify the dimensions along which students based their epistemic and non-epistemic source evaluations and to ensure that students accurately reported those dimensions, we decided to employ a retrospective guided interview. It was our determination that interviewing students after task completion, with sources available, would provide the best conditions for students to both articulate the evaluative dimensions they considered and to embed those dimensions in the task.

**Task and Source Evaluation**

In addition, we were interested in the extent to which students' evaluations may differ across task demands. Models of MSU conceptualize that process and the source evaluation process as task driven (e.g., Rouet & Britt, 2010). In other words, based on this literature, one would expect students' source evaluations to vary based on the type of task motivating their use of multiple sources (e.g., Wiley & Voss, 1999). Yet, few studies have systematically examined how students' source evaluations may vary epistemically (or non-epistemically) under differing task conditions, despite acknowledging this as an important area for further research.

In one such study, Gerjets et al. (2011) presented students with differing guidance for source evaluation (i.e., either explicit instructions to evaluate sources or not). However, in this study, the task—which was to write a report on what diet methods were preferable for weight

loss—remained constant.  What these researchers determined was that the number of students

recommending each diet plan differed significantly across the two prompts.  In another

investigation, Cerdán and Vidal-Abarca (2008) presented two types of tasks to students

pertaining to three texts on bacteria and biological resistance.  Specifically, students were asked

to either write an integrative report on the three texts or to answer intratextual questions about

each of the three texts.  Although students' source evaluations were not explicitly assessed, both

tasks were conceptualized as requiring students to make relevance judgments regarding which

information was most pertinent in each of the three texts.  Indeed, authors found that students in

the intertextual essay condition were better at distinguishing relevant from irrelevant information

(e.g., devoted more reading time to relevant pieces of text), pointing to the need to examine the

influence of task features on students' multiple source use behaviors, and source evaluations

more specifically.

        Given the theorized importance of task in driving students' multiple source use behaviors

(e.g., Britt et al., 1999), the present study examined students' source evaluations under differing

task conditions.  In both Study 1 and Study 2, students' source evaluations were considered in

response to two different question types, one discrete and one open-ended.  The two question

formats were thought to differ in the type of multiple source use interactions they would demand

of students and, thus, the type of responses they would elicit.  In particular, the discrete question

was thought to require a single precise answer, while the open-ended question was expected to

necessitate a more elaborated and extended response.  Further, while both questions could elicit

MSU, the purpose of sources was expected to differ by question type.  In essence, for the more

discrete question, we hypothesized that the multiple sources would provide corroboration for the

more precise answer requested, whereas the integration and consolidation of information from

multiple sources would seem warranted to provide a clear and comprehensive answer to the open-ended question.

**Domain-Specific Differences**

An additional parameter of students' source evaluations that has been under-examined in prior research is potential domain-specific differences. Bråten et al. (2009) noted that the majority of studies examining students' source evaluation have been nested in the domain of history. Bråten et al. suggested that there may be disciplinary differences in the ways students interact with multiple texts; for example, in the domain of history, multiple source use may involve putting together various accounts of events to form a coherent narrative, while climate change, the domain chosen by authors, represents a complex topic that requires participants to evaluate potentially conflicting scientific evidence. However, only a limited number of studies have examined differences in multiple source use across domains.

Walraven et al. (2009) examined differences in students' evaluations when researching information problems in the domains of geography, physics, and language and culture. Although these researchers identified significant differences in students' evaluations of results on search engine pages (SERPs) across domains, the limited number of evaluations of sources themselves students reported in the study did not allow for extensive analysis (Walraven et al., 2009). As differences in source evaluation by domain may emerge based on differential prior knowledge, understandings of author/source credentials, and varying source types across domains (e.g., Rouet, 2006), there is a need for studies to examine how evaluations of sources may differ across domains. This was our goal in Study 2.

**Source Type**

In addition to considering the parameters of the task, domain, and question type as factors impacting students' source evaluations, there was a need to consider how the sources themselves affected students' evaluations.  Source type has been identified as an important factor differentiating sources (e.g., Wiley & Voss, 1999), with variations in source type corresponding to differences in author/publisher credibility, format and style, and the type, quality, and scope of information presented in each sources.  In the multiple source use literature, differences in document type have been used primarily to differentiate sources along epistemic dimensions of trustworthiness or reliability.  For example, Mason et al. (2010b) presented students with four sources they considered to be authoritative and four they considered non-authoritative.

Similarly, Bråten et al. (2009) presented students with a set of seven different source types to investigate students' differential trustworthiness ratings of these sources.  In the text-set used by Bråten and colleagues, the reports presented by the government agency and the university-affiliated Cicero research center were considered to be high in trustworthiness; the two newspaper articles, as well as the for-profit corporate source, were intended to be low in trustworthiness, whereas the textbook and the science magazine article were intended to be of ambiguous reputability (e.g., Bråten et al., 2011; Bråten et al., 2008).  Given the importance of source type as contributing to students' source evaluations, we were interested in building on this literature to examine how students differentially evaluated sources of different types not only along epistemic dimensions, but along non-epistemic dimensions as well.

**Research Questions**

Given the prior research in MSU and the limited areas of exploration that were identified, we posed several research questions for investigation. Specifically, for Study 1 and Study 2 we sought to address the following questions

**Study 1**

1. What is the nature of differences in the number of epistemic and non-epistemic source evaluations that students report when evaluating sources in response to discrete and open-ended questions?

2. To what extent do the average number of epistemic and non-epistemic source evaluations differ when students are responding to discrete versus open-ended questions?

3. What specific types of epistemic and non-epistemic evaluations do students report when evaluating sources in response to discrete and open-ended questions, and do the types of evaluations differ across question type?

**Study 2**

1. How do students' evaluations of source accessibility, usefulness, trustworthiness, and familiarity differ when responding to discrete versus open-ended questions?

2. How do students' evaluations of source accessibility, usefulness, trustworthiness, and familiarity differ when responding to questions in the domain of psychology versus astrophysics?

3. How do students' evaluations of source accessibility, usefulness, trustworthiness, and familiarity differ by the type of sources students accessed?

**Study 1**

**Method**

Participants. Participants for Study 1 were 31 undergraduate students from a large mid-Atlantic university. The sample skewed female (87.10%) and the average age was 22.16 years old (*SD*= 0.34). The sample was predominantly White (58.06%) and Asian (32.26%), with one student reporting African-American ethnicity, and two students selecting the "other" category (6.45%). In the sample, 16.13% of students reported being non-native English speakers. Students in the study represented a wide variety of social-science majors including education (22.58%), psychology (19.35%), and public health (12.90%). Students' average grade point average on a four-point scale was 3.28 (SD=0.42).

Students were recruited from two human development courses and were offered extra-credit for participation. Human development students were selected as the sample for this study as the topics of the search task (i.e., fertility rates and government-sponsored childcare) were considered to be relevant and interesting to participants. Moreover, it was expected that participants would have some, but not extensive, familiarity with the topics that comprised the task. Thirty-five students initially participated in the study. However, four participants were excluded: one had previously participated in a similar study, two declined to be audio-recorded during the interview, and the fourth was not recorded due to a computer malfunction. Thus, a final sample of 31 participants was retained.

Study components and procedure. The present study consisted of three parts: (a) a graphics task asking students to depict and justify their understandings of the relations among the constructs of *knowledge*, *information*, and *truth*, (b) a search task, in which students responded to two questions using a library of eight sources, and (c) a guided retrospective interview, asking

participants to narrate how they went about answering two academic questions. Only the results

from the search task and the subsequent retrospective interviews pertain to the question of

students' justifications for source selection and will be discussed herein.

For all parts of Study 1, researchers met individually with students who conducted the

online task in a designated room. Researchers were positioned in such a way as to discretely

observe the participants' actions and to record via field notes essential behaviors, including the

sources students accessed and participants' scrolling through texts. These data were to be used

during the structured interviews that followed task performance. Prior to receiving the questions

and interacting with sources, students were instructed to answer two questions "using a list of

sources, as if you were doing so for a social science class." Students were also instructed to only

use the library of sources provided and that the sources in the library were arranged in

alphabetical order. Participants were told to use all the time they needed for the task, and that

after clicking the "submit" button for the first question, they would be presented with the second.

Pressing the "Start" button below the instructions took students to the first question and library

of eight sources.

**Materials: Search questions.** The two questions included one discrete question (i.e., In

the USA, what is the replacement fertility rate?) and one open-ended question (i.e., What is the

role of government sponsored childcare in high fertility rates?). Questions were presented to

participants in counterbalanced order. Each question was presented, along with a textbox and a

"submit" button at the top of the page, above the source library. This question and textbox

display remained visible while students accessed sources from the library, and students were not

restricted as to the length of their response.

**Materials: Source library.** The source library, designed to resemble a Google hits page, included the title and brief description of eight sources that could be used to answer both questions. Students were informed that these sources were presented alphabetically. Included were three websites, two PDF book chapters, one newspaper article, one magazine article, and one PDF journal article. Each source description had three parts: (a) author/publisher and title, (b) source type (i.e., website, pdf book chapter, magazine) and link, and (c) description of the source. Descriptions were identical to those obtained from a Google search for each source.

The first source, *Global Politician*, was a conservative website that provided summaries of various blog and article postings. Source two was a chapter titled "Reasons for Fertility Decline," from a book published by the National Research Council. The third source was a chapter, "What Determines Fertility," from the book *Six Billion Plus: World Population in the 21st Century*. A Newsweek article titled "The End of Motherhood," was the fourth source and the Wikipedia entry for "Fertility Rate" the fifth. The sixth source was a website from the Population Research Bureau titled, "Tracking Trends in Low Fertility Countries, " and the seventh source was a journal article from the Japanese Journal of Population titled, "Very Low Fertility: Consequences, Causes, and Policy Approaches." The eighth source was a Washington Post newspaper article, "With Each French Birth, A Dividend From the State." The sources were chosen to represent a variety of source types and to have differing degrees of credibility. A screenshot of the source list is included as Appendix A.

Each of the sources provided information of varying quality relevant to the two questions. The sources provided complementary, and at times conflicting information, representing a variety of views and perspectives on the same issues. For example, in the case of the open-ended question, there was a source providing anecdotal evidence of the benefits of government

sponsored childcare for women's careers (magazine article), as well as descriptions of pro-natal policies in foreign countries (blog, book chapter), and sociological analyses of the relation between government supported child-care, traditional cultural practices, and economic growth (book chapter, journal article). Sources were intended to reflect the range of information students would encounter through the course of researching the topics of the question.

The Flesch-Kincaid Reading Ease and Flesch-Kincaid Grade Level measures were used as readability quotients to assess the appropriateness of the sources for a college-age sample. On average, the eight sources had a Flesch-Kincaid Reading Ease score of 35.39 (range: 18.2-45.6) and a Flesch-Kincaid Grade Level of 11.7 (range: 10.5-12.0), suggesting that all eight sources were appropriate for use with an undergraduate sample.

**Measures: Interview protocol.** Following completion of the search task, the first or second author interviewed the students about their process for answering each question. A guided retrospective interview was used. For each question participants were first asked, "To start off, I want you to tell me a little bit about how you went about answering the first/second question." Once students overviewed their response process, participants recreated their source navigation for the interviewer, re-clicking and responding to questions asking them to reflect on each source they accessed. If students forgot or misremembered the order of their source use, their recall was prompted based on the observational data interviewers had recorded in the researchers' field notes.

During the guided interview, participants were asked questions pertaining to their source use, source evaluation response formulation, overall reflections on the search process, and response formation. For this study, questions pertaining to students' source evaluation, and specifically students' judgments of each selected source, were analyzed. For example, for each

source that participants accessed, we asked, "What did you think of this source?" and "Did this source meet your expectations?" to elicit students' source evaluations. The complete interview protocol is included as Appendix B. The time it took to complete the interviews ranged from 8.20 minutes to 39.08 minutes, averaging approximately 15.83 minutes.

**Measures: Interview coding.** Students' source evaluations were coded using a four-step process. First, students' interviews were coded into idea units, with each unit corresponding roughly to a single thought. Next, each idea unit was coded as either an evaluation or a non-evaluation. Evaluative comments were those that made judgments or appraisals of any of the sources access or the information within them (e.g., "It got kind of repetitive, not really repetitive, but I felt like I wasn't learning anything new). In contrast, non-evaluative statements were those that pertained to any other aspect of students' multiple source use process.

Evaluations were then coded as either epistemic or non-epistemic in nature. Epistemic evaluations were those that concerned the quality of knowledge and the truthfulness of a source or the information within it. As such, epistemic source evaluations were those that judged sources based on their *reputability* or *author credentials*. In contrast, non-epistemic evaluations were those that did not evaluate sources based on dimensions of knowledge quality of truthfulness but rather on the basis of such characteristics as source *relevance* or *ease of use*.

In the final coding step, epistemic and non-epistemic evaluations were characterized into sub-categories. There were a total of seven subcategories along which students evaluated a source's epistemic properties, and a total of nine subcategories were used to characterize source evaluations based on non-epistemic features. Sub-categories of source evaluation were independently derived and confirmed by two researchers based on an exhaustive reading of participant interviews.

***Epistemic dimensions of source evaluation.***  Epistemic evaluations were classified as belonging to one of seven categories: (a) *reputability*, (b) *author*, (c) *source type*, (d) *source bias*, (e) *evidentiary quality,* (f) *corroboration,* and the (g) *nature of information* found in the source. Source evaluations consistent with the *reputability* category included fairly general appraisals of how reputable or reliable a source was considered to be.  For instance, one student reported, "I know that like if the .org websites than it's more reliable than Newsweek."  Evaluations identified as addressing the dimension of *authorship* were those that discussed either the author's or publisher's credentials or the peer-review/editorship process associated with the composition of a particular source.  For example, a source evaluation based on, "Just the fact that it's clear that the author did a lot of background research," was classified into the *authorship* category.

Source evaluations based on *source type*, judged sources according to the type of source it was.  This category was used most often by participants to express their trust in books.  As one participant reported, "They're all scholarly research papers as opposed to like a Newsweek article…like I know I can trust it more if its from a website."   Some source evaluations considered the potential bias of a source.  The *source bias* category included evaluations that considered author and source motivations in the information presented.  For example, one student considered the potential for source bias based on the topic of the source, "I thought maybe it was biased because abortion can be a touchy subject for some people."

When source evaluations were coded as belonging to the *evidentiary quality* category, students reported evaluating these texts either because they contained data or research findings or because the sources included a lot of citations, suggesting evidence for the information presented.  One student evaluated a source as one, "that has citations at the end…it's a scientific study, with research."   This type of evaluation was coded as consistent with the evidentiary

category.  Source evaluations corresponding to the *corroboration* category were those that

evaluated sources relative to the information presented in other texts.  One student explained, "I

think I found another, something else confirmed what I had found from Wikipedia."  Finally,

epistemic source evaluations coded as within the *nature of information* category evaluated

sources based on the type of information presented.  Usually these were evaluations of the

easiness, quickness, or comprehensiveness of the information in a source.  This final category

was most consistent with Schommer's (1990) quick learning dimension of epistemic beliefs.

    ***Non-epistemic dimensions of source evaluation.***  Considering students' source

evaluations along non-epistemic dimensions, nine evaluative categories were identified: (a)

*content,* (b) *information use*, (c) *relevance*, (d) *time*, (e) *accessibility*, (f) *answer*, (g) *familiarity*,

(h) *informativeness,* and (i) *interest*.  Source evaluations corresponding to the *content category*

judged texts based on the specific information found within the source.  However, such

evaluations did not consider the quality or reliability of this information.  Those students whose

source evaluations were coded as within the *information use* category judged texts based on what

they were able to do with the information found in each source.  For example, students whose

evaluations were coded as such discussed sources as enabling them to infer or deduce a response

or to make "an educated guess" based on the information found in the source.  As one student

explained, "I kind of pieced it together with that and what I read and not a shot in the dark but I

tried to make an educated guess."

    Evaluations coded as based on *relevance* were those that judged a source based on its

pertinence to the question.  For example, as one student stated, "It didn't [meet my expectations]

because I didn't find anything really relating."  Some evaluations of sources were concerned

with *time*, this category included source judgments specifically based on how much time the

source seemed to required for its use or how much time students had spent using the source.

Source judgments coded into the *accessibility* category, were source evaluations concerned with

a source's ease of use. This included factors such as source length, complexity, and text

formatting, including font size, use of headings, or presence of graphs in the text. For example,

one student described a source this way: "It was very, very long…it was too, too much." When

evaluations were placed into the *answer* category, these evaluations concerned whether students

found the exact response to the question they were looking for in the text. Those students

responding as such seemed to want to be able to copy an exact answer from texts into their

response. One student reported, "Like when I was trying to write the answer…I wish I could

have copied and pasted."

Evaluations coded as belonging to the *familiarity* category were those that judged a

source based on how familiar it was to students. The majority of evaluative comments in this

category included positive reviews of Wikipedia, because it was a standardized source with

which the students had a great deal of experience. Some evaluations were based on whether the

sources were *informative*—a broad classification that seemed to involve determinations of the

volume of information a source provided or the extent to which a source offered participants

"facts." As one student simply put it: "They give good information." Finally, some students

reported evaluating sources based on their *interestingness*, or how engaged students were with a

text. For example, one student reported, "I was also looking for if there are any other things I

hadn't thought about it, why there are low birth rates, it was interesting…"

## Results: Study 1

Study 1 examined the proportion of epistemic and non-epistemic source evaluations

students provided when responding to discrete and open-ended questions, and sought to

characterize the different types of epistemic and non-epistemic source evaluations and examine

differences by question type.  Across both the discrete and open-ended questions, participants

reported a total of 632 independent utterances pertaining to source evaluation, with 149 (23.59%)

of those classified as epistemic in nature and 483 (76.42%) of utterances classified as non-

epistemic.  Students produced an average of 10.19 source evaluations for each source selected

($SD = 8.93$), with an average of 2.40 of these evaluations classified as epistemic ($SD = 3.06$), and

an average of 7.79 classified as non-epistemic ($SD = 7.31$).

    We were first interested in examining whether students differed in the amount of

epistemic versus non-epistemic source evaluations offered in response to the discrete and open-

ended questions.  When evaluating sources in response to both the discrete and the open-ended

questions, paired sample $t$-tests revealed that students produced significantly more non-epistemic

source evaluations than epistemic source evaluations [discrete: $t(30) = -4.15$, $p<.001$; open-

ended: $t(30) = -4,69$, $p<.001$].  In the case of the discrete question, students produced an average

of 2.16 ($SD = 2.89$) epistemic source evaluation utterances and an average of 7.71 ($SD = 7.79$)

non-epistemic source evaluation utterances.  In responding to the open-ended question, students

produced an average of 2.65 ($SD=3.33$) epistemic source evaluation utterances and an average of

7.87 ($SD=6.83$) non-epistemic source evaluation utterances.  Three paired-sample t-tests were

then conducted to ascertain whether the total number of evaluations, number of epistemic

evaluations, and number of non-epistemic source evaluations differed across the discrete verses

open-ended questions.  No significant differences by question type were found.

    In this study, we were also interested in exhaustively cataloging the basis along which

students evaluated sources.  These sub-categories of source evaluations, along both epistemic

and non-epistemic dimensions, are reported on descriptively (see Table 1), as the limited sample

size and the relatively small number of evaluations in each category precluded parametric

comparisons. Percentages were calculated within categories. The results that follow are based

on the percentage of total number of epistemic evaluations (n=149) belonging to each

subcategory and the total number of non-epistemic evaluations (n=483) belonging to each

subcategory.

      **Epistemic categories of source evaluation.** In evaluating sources along the epistemic

dimension, students' responses most often fell within the *reputability* sub-category. In fact,

30.20% of students' epistemic evaluations (n=45) were in this category. Further, 16.11% of

students based their source judgments on the extent to which sources *corroborated* information

found in other texts (n=24), and 15.44% of students based their source evaluations on *source*

*features*, such as author (n=23) and source-type (n=23). In responding to the discrete versus

open-ended questions, these evaluative categories were consistent. For example, in evaluating

sources for the discrete question along epistemic dimensions students' evaluations were most

often in the *reputability* category (34.29%, n=24), followed by the *corroboration* category

(18.57%, n=13). Similarly, students responding to the open-ended dimensions most often based

their source evaluations on source reputability (26.58%, n=21). However, in responding to the

open-ended question, students also frequently based their evaluations on author credentials

(17.72%, n=14) and on source type (18.99%, n=15).

      When examining patterns across the two questions, we found that 34% of students'

evaluations considered source reputability when they were responding to the discrete question, as

compared to 27% for the open-ended question. Conversely, when responding to the open-ended

question, 18.99% of source evaluations were based on source type, as compared with 11.43% of

source evaluations in response to the discrete question.

**Non-epistemic categories of source evaluation.** In evaluating sources along non-epistemic dimensions, students most often referenced the accessibility of the source (22.57%, n=109), basing that judgment on factors such as length and layout.  Further, students also based their rating on how informative they perceived the source to be (20.50%, n=99), source content (15.73%, n=76), and whether the source provided them with an answer to their question (15.53%, n=75).  Interestingly, although prior studies suggest that students often select sources primarily based on relevance (e.g., Kammerer & Gerjets, 2012), students referenced relevance only to a relatively limited extent in this study (13.25%, n=64).  Additionally, very few students reported evaluating sources based on interestingness or engagement, (1.86%, n=9).

Certain discernible differences in types of non-epistemic judgments were identified across question type.  First, *informativeness* served as a basis for source evaluation in 26.43% of evaluations responding to the open-ended question (n=74), as compared with only 12.32% of source evaluations in response to the discrete question (n=25).  Further, 19.64% of non-epistemic source evaluations were based on *source content* in the case of the open-ended question (n=55), however in responding to the discrete question only 10.34% of non-epistemic source evaluations cited this reason (n=21).  Conversely, in responding to the discrete question, students based their judgments more often on *source relevance* (18.23%, n=37) and whether the source offered an answer to the question (20.20%, n=41), than they did when evaluating sources used in response to the open-ended question (relevance: 9.64%, n=27; answer: 12.14%, n=34).  *Source accessibility* was frequently reported as a justification across both question types, with 25.62% and 20.36% for the discrete and open-ended questions, respectively.  A summary of the descriptive statistics on source evaluations is displayed in Table 2.

## Discussion: Study 1

In responding to both discrete and open-ended questions, students offered significantly more source evaluations based on non-epistemic dimensions, as compared with epistemic dimensions. This was not only a difference in quantity. In effect, students seemed to offer more varied non-epistemic reasons for source-evaluations, than epistemic ones. When students forwarded epistemic source evaluations, the majority of those evaluations were based on the somewhat vague characteristics of *reputability*. When providing non-epistemic source evaluations, students seemed to draw on a diverse range of factors, including source content, relevance, accessibility, and how informative they considered the source. Findings from the study confirmed the multidimensional nature of students' source evaluations. That is, students tended to provide a large number of different types of source evaluations for each source accessed, both epistemic and non-epistemic, suggesting the need to examine the extent to which these dimensions may function in conjunction with one another in the process of multiple source use.

A second aim of this study was to understand differences in source evaluations by question type. Overall, the number of epistemic and non-epistemic source evaluations did not differ by question type. However, there was some evidence that the specific types of epistemic and non-epistemic source evaluations students provided did differ across question types, suggesting that although students may be providing a similar number of epistemic and non-epistemic evaluations for each of the questions, the nature of those evaluations may depend on the question that students are answering. For instance, when responding to the discrete versus open-ended question, students were twice as likely to provide source evaluations based on the source providing an answer to the question and the relevance of the source; concern for these

particular dimensions seems to be aligned with conceptions of discrete questions as having a single, precise answer.  Conversely, in response to the open-ended versus discrete question, students were more than twice as likely to evaluate sources based on either the content or informativeness of those sources.  Concern for these dimensions perhaps reflects an understanding of open-ended questions as requiring a more elaborated response.  This difference in the nature of source evaluations seems to suggest that students have sensitivity to the demands of the task when evaluating sources, at least when evaluating sources along non-epistemic dimensions.

To the extent that we were interested in cataloging the span of justifications students used in evaluating sources, this study was successful in identifying dimensions of source evaluation that occurred somewhat frequently in student interviews but were relatively underrepresented in the literature.  This included a concern with the accessibility of a source and the extent to which the source provided "the answer" to the question.  Further, some differences were identified in dimensions along which students are frequently asked to evaluate sources in the literature, and the extent to which students in our sample referenced these dimensions.  For instance, while prior research has suggested that relevance might be one of the most importance factors in students' evaluations of sources (Kammerer & Gerjets, 2012), this type of evaluation was not the most popular in our sample of students, particularly when they were engaged in responding to the open-ended question.

Although Study 1 was effective in capturing the range of dimensions students use to base their source evaluation, due to the categorical nature of data collected, we were not able to examine differences in students' source evaluations across tasks.  Thus, in Study 2 we selected a subset of evaluative dimensions identified in Study 1 and occurring in the literature (i.e.,

trustworthiness, utility, accessibility, and familiarity), and examined how these dimensions functioned across variations in task, namely across question type and domain. In addition, a limitation of Study 1 was the relatively limited extent to which we were able to investigate how students' source evaluations varied according to source type. Specifically, we were able to account for source type only by presenting students with a range of texts differing in reliability. However, by limiting the number of dimensions to examine in Study 2, we were able to consider the extent to which these evaluated dimensions differed not only by question type but also by domain and by source type. Additionally, although Study 1 examined different types of questions within a single domain, Study 2 aimed to better understand domain differences suggested in prior research (Bråten et al., 2009; Rouet, 2006).

## Methods: Study 2

### Participants

Study 2 included 367 participants (256 female; 69.75%). Participants were recruited from a variety of human development courses, including educational psychology, research methods, and lifespan development and in some cases students were offered extra credit for participation. Participants had an average age of 20.90 years old ($SD = 2.22$). The sample consisted primarily of upperclassmen: 1.09% (n=4) of students were freshmen, 17.71% (n=65) were sophomores, 34.06% were juniors (n=125), 42.78% were seniors (n=157), and 4.35% of students (n=16) were taking the courses for post-graduate credit. Students represented a wide range of ethnicities. Although the sample was majority White (59.95%, n=220), it also included 12.53% (n=46) Asian students, 11.72% (n=43) African American students, and 7.90% (n=29) students who identified as Hispanic/Latino. Additionally, 3.54% (n=13) of participants self-identified as Multiracial, while 2.45% (n=9) responded "Other" or did not specify their ethnicity.

The majority of students were majors in the social sciences (n=310; 84.70%). However,

the sample also included students majoring in the natural sciences (n=41; 11.20%) and in the

humanities (n=15; 4.10%). Those students majoring in the social sciences were primarily

studying psychology, kinesiology, and family science. Participants had an average GPA of 3.20

on a four-point scale ($SD = 0.43$). Also, 9.23% of the sample (n=33) was foreign born, and

11.86% of participants (n=42) reported that English was not their native language.

**Materials, Measures, and Procedures**

Similar to Study 1, Study 2 was part of a larger study that examined students' beliefs

about knowledge, information, and truth, and how these beliefs may manifest when students

interact with information online to complete academic tasks. The part of the study that is

reported here was a search task similar to the search task used in Study 1. In this task,

participants responded to two different questions using a digital library of sources.

The search task featured four questions: a discrete and an open-ended question in the

domain of developmental psychology and a discrete and an open-ended question in the domain

of astrophysics. Question presentation was counterbalanced across domain and question type.

For example, if students were first presented with the discrete developmental psychology

question (i.e., "How many words are in a normally developing 24-month-old's vocabulary"?)

they would then receive the open-ended question from astrophysics (i.e., "Which planetary

features may promote or hinder habitability? Please explain.") Conversely, if students

responded first to the open-ended question in the domain of developmental psychology (i.e.,

"Which factors may promote or hinder linguistic development in young children? Please

explain."), they then responded to a discrete question in the domain of astrophysics (i.e., "How

many stars are in the Milky Way galaxy.")  Students were instructed to respond to the questions "as they would for an academic course."

The two domains of developmental psychology and astrophysics, one social science and one natural science, were chosen to represent the range of fields to which students may be exposed during their undergraduate education.  Further, as the majority of our sample were recruited from social science courses, we expected these two domains to be ones in which students would exhibit differing prior knowledge and interest.

In responding to each of the four questions, students were presented with a library of digital sources varying in type and reliability.  There were a total of seven sources included for each question: a blog, newspaper article, a science news site, journal article, government report, Wikipedia entry, and a website.  Across the four questions, we matched sources by source type (e.g., newspaper) and publisher (e.g., New York Times), word-length, and readability.  Sources were all excerpts from actual text found on the Internet.  We identified these texts by first conducting a Google search for the target question, and then selecting sources that would be appropriate for an undergraduate audience and would represent sufficient diversity in terms of source type and reliability.  Given prior research (Mason et al., 2010b), the journal article and government report were thought to represent highly reliable sources; the Wikipedia entry, science news site, and newspaper articles were thought to represent sources of moderate reliability; and the blog and website were thought to represent sources low in reliability.  Sources were validated by experts in educational psychology and deemed appropriate for use with a general undergraduate population.  In addition, we piloted the materials and the search task with undergraduate students to ensure that sources tended to function effectively and elicited variable responses.

All of the sources included information relevant to answering the target question, however the information varied in quality and reliability. Each of the seven sources presented a differing perspective on the issue. For example, in the sources pertaining to young children's linguistic development, we included a government report on assessment of linguistic development; a science news site on the role of ambient noise, like television, in young children's linguistic development; a newspaper article on the role of reading in young children's language development; a Wikipedia entry including milestones in language development; and a website introducing parenting techniques, such as gesturing, to promote linguistic development. While some information was unique to each sources, sources also included elements repeated in other texts. The seven sources presented both complimentary and partially conflicting information in response to each of the questions. The range of sources presented was thought to represent the range of information students would encounter through the course of researching on the Internet.

The sources were presented to students as a part of a digital library with seven tabs, one for each source. Each tab was labeled with source type (e.g., newspaper) and no additional information. This was done to concentrate students' source selection on source type rather than other features such as relevance. The sources were arranged in alphabetical order by source type. After students had completed using a source, indicated by them either closing the source and selecting another in the digital library, or hitting the "submit" button for their response, a box popped up asking them to rate the source on each of four dimensions: accessibility ("How accessible was this source"), trustworthiness ("How trustworthy was this source"), familiarity ("How familiar was this source"), and utility ("How useful was this source"). Participants

evaluated each of these dimensions by clicking on a 100mm line ranging from *not at all* to *very much so*.

The aforementioned dimensions were selected based on prior research and, specifically, the findings of Study 1, and were thought to correspond to both epistemic (i.e., trustworthiness) and non-epistemic dimensions (i.e., accessibility, familiarity, utility). Further, the dimensions required students to make general evaluations about the source (e.g., trustworthy, accessible) as well as to evaluate the source in reference to the question (e.g., useful). The *familiarity* dimension was included to account for the fact that undergraduates were likely to have had more experience with sources like Wikipedia and the New York Times than with other sources such as the science news site, Science Daily.

Participants were allowed to access the sources in any order, and to access none, some, or all of the sources in the digital library. Students were also allowed to revisit sources as needed, and students were asked to evaluate each source after use. The participating students were able to refer to sources while composing their response. Once students had formulated their response to the first question and hit the "submit" button, they were presented with the second question.

## Results: Study 2

We first briefly overview students' search and source use process in responding to questions differing in domain and question type, before turning to an analysis of students' source evaluations more specifically.

### Overall Source Use

Across the four questions, students used an average of 4.36 sources of the possible seven, with some students using all of the sources and some students using none. Further, students spent an average of 132.80 seconds, or a little over two minutes, examining each source. Across

all four questions, participants accessed a total of 953 sources. Of all sources accessed, across all four questions, the journal article was the most popular source, constituting 29.4% (n=280) of source accessed, followed by the government report (n=211, 22.1%) and Wikipedia entries, (n=140, 14.7%).

In responding to the discrete question in developmental psychology (i.e., "How many words are in a normally developing 24-month-old's vocabulary"?), students favored the government report, the journal article and the Wikipedia entry when responding to the question. Specifically, of all sources accessed in responding to this question, 31.6% (n=92) were the journal article, 20.6% were the government report (n=60), and 14.1% (n=41) were the Wikipedia entry. In the case of the discrete question in astrophysics (i.e., "How many stars are in the Milky Way galaxy"?) students similarly favored the journal article (n=65, 35%), followed by the government report (n=41, 22.5%), and the newspaper article (n=24, 13.2%).

Similarly, in the case of the open-ended question in the domain of developmental psychology (i.e., "Which factors may promote or hinder linguistic development in young children? Please explain."), the government report (n=41, 24.8%) and the journal article, (n=39, 23.6%) were favored sources. However, the most popular source in responding to this question was the Wikipedia entry (n=48, 29.1%). Finally, in the case of the open-ended question in the domain of astrophysics (i.e., "Which planetary features may promote or hinder habitability? Please explain."), the government report (n=69, 21.9% ), the journal article (n=84, 26.7%), and the newspaper article (n=49, 15.6%) were the most frequently used sources.

In the descriptions of sources accessed, the results were presented in terms of the total number of sources accessed, rather than the number of participants who may have accessed a particular source. This was because the students were able to re-visit sources as many times as

they wanted. However, in subsequent analyses of source evaluations only students' initial source evaluations were analyzed. We now turn to an examination of whether students' evaluations of various source types, along dimensions of utility, accessibility, familiarity, and trustworthiness, differed across question-types (RQ1), domains (RQ2), and source types (RQ3).

**Question Type**

A one-way analysis of variance was used to examine differences in students' source evaluations along dimensions of utility, accessibility, familiarity, and trustworthiness across the two question types. In responding to discrete verses open-ended questions, students significantly differed in their source evaluations only along the familiarity dimension, $F(1, 951) = 7.55, p<.01$, but not for the dimensions of accessibility, $F(1,951) = 0.13, p=0.71$, utility, $F(1, 951) = 0.50$, $p=0.48$ or trustworthiness, $F(1, 951)=1.04, p= 0.31$. In evaluating the familiarity dimension, students found sources to be more familiar when responding to the discrete ($M = 54.82, SD = 29.14$) versus the open-ended question ($M = 49.36, SD = 32.09$).

**Domain**

A one-way analysis of variance was used to examine differences in students' source evaluations along dimensions of utility, accessibility, familiarity, and trustworthiness across the two domains, developmental psychology and astrophysics. In responding to questions across the two domains, students significantly differed in their source evaluations only for the familiarity dimension, $F(1, 951) = 5.10, p<.05$, but not for the dimensions of accessibility, $F(1,951) = 1.47$, $p=0.23$, utility, $F(1, 951) = 2.80, p=0.09$, and trustworthiness, $F(1, 951)=0.13, p= 0.72$. In evaluating the familiarity dimension, students found sources to be more familiar when responding to the question in developmental psychology ($M = 54.41, SD = 30.69$) versus astrophysics ($M = 49.92, SD = 30.71$).

**Source Type**

Finally, a series of one-way analysis of variances were conducted to examine students'

differences in source evaluations along dimensions of utility, accessibility, familiarity, and

trustworthiness across the seven different source types (e.g., blog, journal article, government

report).

**Familiarity.** Significant differences were found in students' evaluations of familiarity

across the seven source types: blog, Wikipedia entry, general website, government report, journal

article, newspaper, and science news site, $F(6, 946) = 15.33$, $p<.001$. Tukey's post hoc analyses

were used to examine which source types students were found to be more familiar. The

Wikipedia entry ($M=71.39$, $SD=27.93$) was significantly more familiar than all of the other six

sources: the government report ($M=47.37$, $SD=29.37$), journal article ($M=45.19$, $SD=31.06$),

newspaper article ($M=59.39$, $SD=29.84$), science news site ($M=48.57$, $SD=27.63$), blog

($M=44.00$, $SD=30.42$), or website ($M=51.72$, $SD=27.43$). Further, students were significantly

less familiar with the government report ($M=47.37$, $SD=29.37$) and the journal article ($M=45.19$,

$SD=31.06$), than they were with the newspaper article ($M=59.39$, $SD=29.84$).

**Accessibility.** There were significant differences found in students' evaluations of source

accessibility and ease of use across the seven source types, $F(6, 946) = 2.74$, $p <.05$. Post-hoc

analyses using Tukey's post-hoc procedure found that participants rated the Wikipedia entry

($M=72.94$, $SD=26.25$) more accessible than both the journal article ($M=62.31$, $SD=28.77$) and

the government report ($M=63.67$, $SD=26.47$).

**Utility.** There were significant differences found in students' evaluations of source utility

across the seven source types, $F(6, 946) = 2.95$, $p <.01$. Post-hoc analyses using Tukey's post-

hoc procedure determined that students rated the Wikipedia entry ($M$=65.03, $SD$=30.30) as being more useful than the journal article ($M$=54.19, $SD$=31.33).

   **Trustworthiness.** A one-way ANOVA determined significant differences in students' trustworthiness ratings across the seven source types, $F(6, 946) = 15.08$, $p<.001$.  Post-hoc analyses using Tukey's procedure found that the students considered the journal article ($M$=67.24, $SD$=25.90) and the government report ($M$=68.72, $SD$=25.32) to be significantly more trustworthy than the newspaper ($M$=58.40, $SD$=26.17), the Wikipedia entry ($M$=49.24, $SD$=9.23), and the blog-post ($M$=39.76, $SD$=29.81).  Further, the newspaper article ($M$=58.40, $SD$=26.17), the news site ($M$=62.24, $SD$=22.96), and the general website ($M$=59.10, $SD$=25.42) were both rated by students as significantly more trustworthy than the blog post ($M$=39.76, $SD$=29.81), and the news site ($M$=62.24, $SD$=22.96) was rated significantly more trustworthy than the Wikipedia entry ($M$=49.24, $SD$=9.23).

## Discussion: Study 2

  In considering differences in source evaluations along dimensions of familiarity, accessibility, utility, and trustworthiness, only students' source evaluations along the dimension of familiarity were found to differ across both question type and domain.  However, students differed in their source evaluations along all four of these dimensions when source type was examined.  When evaluating sources along the trustworthiness dimension, students were able to distinguish not only between highly trustworthy sources, such as the journal article and government report, and sources low in trustworthiness, such as the blog post, but also seemed to be able to identify nuances in the trustworthiness for the different source types.  For example, students rated the science news site as slightly more reliable than Wikipedia.

Looking at source evaluation data across these dimensions, we found that the students found Wikipedia to be more useful and more accessible than other sources. Nonetheless, these students were sensitive to the fact that Wikipedia was only moderately trustworthiness as an online source. Similarly, students were also significantly more familiarity with Wikipedia and newspaper articles than other potentially more trustworthy sources. Although not surprising, we find this pattern interesting, given that the upper-level undergraduate students in our study might be expected to have used journal articles in their course of study (Fister, 1992).

Participants evaluated sources differentially in response to variations in task only to a limited extent, differing only along the familiarity dimension. This may be because the four dimensions we selected to investigate in Study 2 did not fully capture the scope of justifications along which students base their source evaluations. It may also be the case that when not prompted to engage in deliberative source evaluation, as participants did in Study 1, students are less able to tailor their source evaluations to specific task demands; a more sophisticated approach to source evaluation. Given the previously established theoretical importance of considering task in studies of multiple source use (e.g., Rouet & Britt, 2010) and the differences in source evaluations across question type determined in Study 1, there is a need to consider the interaction between source type, domain, and question type in subsequent examinations.

### General Discussion and Conclusion

Together, Study 1 and Study 2 were undertaken in order to examine the epistemic and non-epistemic dimensions of students' source evaluations when responding to different question types (Study 1) and to examine how source evaluations may vary when students respond to tasks differing in domain and question type (Study 2). In terms of students' competency with regard to source evaluation, there appears to be a complicated story. On the one hand, Study 1 provides

evidence of students' more frequent evaluations of sources along non-epistemic, rather than

epistemic, dimensions. On the other hand, students proved able to differentiate the

trustworthiness of different sources (Study 2), even if not to differentially evaluate the texts in

response to differing task demands. At the very least, students proved capable of differentiated

source evaluations in response to the task (Study 1) as well as in response to characteristics of

the texts themselves, such as source type (Study 2).

All of the source evaluation dimensions considered in Study 2 were significantly

correlated with one another. The presence of the interrelations among these varying dimensions,

point to the need to consider source evaluations not in isolation, but rather in the ways in which

students evaluate each individual text multidimensionally. This was evidenced in both Study 1

and Study 2, with students producing a number of evaluations, both epistemic and non-epistemic,

for each source accessed in Study 1. It was also evidenced in the manner in which students

evaluated sources of the same type variably across dimension in Study 2. Further, across the two

studies, source type proved to be a potent heuristic that students applied in their evaluations,

serving as a justification for their epistemic source evaluations in Study 1. By comparison, this

same heuristics was brought to bear in Study 2 as students' demonstrating significant differences

in source evaluations across source types on all for dimensions examined.

At the same time, participants in both studies did not evidence ease with source use or

source evaluations. In Study 1, students' source evaluations were often based on parameters

associated with simplifying the demands placed on them by the text. In other words, students

often evaluated source based on their length or ease of reading (accessibility) and the extent to

which sources delivered a ready-answer (answerability). In Study 2, the sources students

considered most useful were Wikipedia and the newspaper article, sources notable for their easy

layout and clear presentation of information. Although considerations of accessibility are not a problem per se, it may suggest deficits in students' multiple source use capabilities. Elsewhere in the literature it has been suggested that multiple source use is a complex problem and may be significantly more challenging for students than use of a single text (Bråten, Strømsø, & Samuelstuen, 2008). This two-study investigation provides support for that claim. Further, students' over-concern with ease of use may come at the cost of selecting more trustworthy and reliable sources, such as journal articles, even when these are required to meet task demands.

Strengths of these two studies include their examination of both non-epistemic and epistemic dimensions of source evaluation in conjunction with one another and under varying task conditions. Together these two studies were able to capture the range of features to which students attend in evaluating sources, and to examine how students appraise particular dimensions of source evaluation under varying task conditions. In conjunction, these studies offer insights into potential methodologies useful for examining multiple source use. Specifically, in Study 1, we examined students' source evaluations based on retrospective guided interviews, thereby capturing the range of dimensions along which students reported judging sources. In particular, the volume of source evaluations produced by students in Study 1 suggests retrospective guided interviews may indeed be a potent methodology for eliciting students' source evaluations. For Study 2, we presented students with pre-determined dimensions along which to evaluate multiple sources. In this instance, we determined that ratings of this sort were able to capture the variety of ways students conceptualized the same type of source, relative to different evaluative dimensions.

While these studies present potent information regarding students' source evaluations, it remains to examine the relation between source evaluations and the extent to which these

evaluations manifest in students' understanding of the sources or the response they construct. A remaining question remains the extent to which students rejected sources they deemed to be unreliable when formulating responses. Data from interviews from Study 1 suggest that often times concerns regarding source reliability were compromised in favor source accessibility. As one participant explained, "I know that it's not like very reliable, but that's where I get most of my answers."

Another area for further consideration is the extent to which source evaluations refer to a specific text, or are themselves multi-textual. In other words, to what extent do students evaluate sources relative to one another? That is to say, do students' judgments of one source as highly reliable lead them to then see subsequent sources as less reputable? An assumption in the analyses featured in this study is that sources evaluations were independent of one another. In other words, evaluations of each source were compared as independent observations. However, the extent to which this is the case remains an empirical question.

In addition, although these findings present initial analyses of students' source evaluations, we are interested in examining the extent to which student students' domain and topic prior knowledge and interest are related to students' source evaluations. Further, we seek to understand the relation between students' source evaluations and task performance. The next step in this program of research is to consider the interactions between all of the dimensions identified in the study (i.e., namely domain, question type, and source type) for a more comprehensive view of the complex process of students' source evaluation.

## References

Alexander, J. E., & Tate, M. A. (1999). *Web wisdom: How to evaluate and create web page quality on the web.* Mahwah, NJ: Lawrence Erlbaum Associates.

Anmarkrud, Ø., McCrudden, M. T., Bråten, I., & Strømsø, H. I. (2013). Task-oriented reading of multiple documents: Online comprehension processes and offline products. *Instructional Science*, 1573-1952. DOI: 10.1007/s11251-013-9263-8.

Brand-Gruwel, S., & Stadtler, M. (2011). Solving information-based problems: Evaluating sources and information. *Learning and Instruction, 21*(2), 175-179.

Bråten, I., Strømsø, H. (2006). Effects of personal epistemology on understanding of multiple texts. *Reading Psychology, 27*(5), 457-484. DOI: 10.1080/02702710600848031

Bråten, I., Strømsø, H., Britt, M. (2009). Trust matters: examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*, 6-28. DOI: 10.1598/RRQ.44.1.1

Britt, M., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485-522. DOI: 10.1207/S1532690XCI2004_2

Britt, M. A., Perfetti, C. A., Sandak, R., Rouet, J. F., (1999). Content integration and source separation in learning from multiple texts. In S. R.Goldman, A.C.Graesser, , & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence* (pp. 73-90). Mahwah, NJ: Lawrence Erlbaum Associates.

Cerdán, R., Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology, 100*(1), 209-222. DOI: 10.1037/0022-0663.100.1.209

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, *77*(3), 515-540.

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

Goldman, S. R., (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Faris, & D. Bloome (Eds.), *Uses of intertexutality in classroom and educational research* (pp. 317-347).  Greenwich, CT: Information Age.

Goldman, S. R., (2011). Choosing and using multiple information sources: Some new findings and emergent issues. *Learning and Instruction, 21*(2), 238-242.

Grimes, D. J., & Boening, C. H., (2001). Worries with the web: A look at student use of web resources. *College & Research Libraries*, *62*(1), 11-23.

Kammerer, Y., Werner, B., & Gerjets, P. (2008). *What evaluation processes are applied during web search? The impact of prompts to evaluate*. Paper presented at the 29th International Congress of Psychology, Berlin, Germany.

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

Mason, L., Boldrin, A., Ariasi, N. (2010a). Epistemic metacognition in context: evaluating and learning online information. *Metacognition and Learning, 5*(1), 67-90. DOI: 10.1007/s11409-009-9048-2

Mason, L., Boldrin, A., Ariasi, N. (2010b). Searching the Web to learn about a controversial topic: are students epistemically active? *Instructional Science, 38*(6), 607-633. DOI: 10.1007/s11251-008-9089-y

Metzger, M.J., Flanagin, A.J., & Zwarun, L. (2003). Student Internet use perceptions of information credibility, and verification behavior. *Computers & Education*, 41, 271-290.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common Core State Standards in English Language Arts*. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington, DC.

Rouet J. F., (2006). *The skills of document use: From text comprehension to web-based learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Rouet, J. F., & Britt, M. A. (in press). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, G. Schraw, (Eds.), *Text relevance and learning from text* (pp. 1-27). Greenwich, CT: Information Age.

Walraven, A., Brand-Gruwel, S., Boshuizen, H. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education, 52*(1), 234-246. DOI: 10.1016/j.compedu.2008.08.003

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301-311.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks.

*American Educational Research Journal, 46*(4), 1060-1106. DOI:

10.3102/0002831209333183

Table 1

*Epistemic Dimensions of Source Evaluation*

|  | Discrete | | Open-ended | | Total | |
|---|---|---|---|---|---|---|
|  | N | % | N | % | N | % |
| Reputability | 24 | 34.29% | 21 | 26.58% | 45 | 30.20% |
| Author | 9 | 12.86% | 14 | 17.72% | 23 | 15.44% |
| Source Type | 8 | 11.43% | 15 | 18.99% | 23 | 15.44% |
| Evidentiary | 9 | 12.86% | 11 | 13.92% | 20 | 13.42% |
| Bias | 3 | 4.29% | 4 | 5.06% | 7 | 4.70% |
| Corroboration | 13 | 18.57% | 11 | 13.92% | 24 | 16.11% |
| Nature if Information | 4 | 5.71% | 3 | 3.80% | 7 | 4.70% |

Table 2

*Non-Epistemic Dimensions of Source Evaluation*

|  | Discrete | | Open-ended | | Total | |
|---|---|---|---|---|---|---|
|  | N | % | N | % | N | % |
| Content | 21 | 10.34% | 55 | 19.64% | 76 | 15.73% |
| Information Use | 11 | 5.42% | 20 | 7.14% | 31 | 6.42% |
| Relevance | 37 | 18.23% | 27 | 9.64% | 64 | 13.25% |
| Time | 8 | 3.94% | 4 | 1.43% | 12 | 2.48% |
| Accessibility | 52 | 25.62% | 57 | 20.36% | 109 | 22.57% |
| Answer | 41 | 20.20% | 34 | 12.14% | 75 | 15.53% |
| Familiarity | 6 | 2.96% | 2 | 0.71% | 8 | 1.66% |
| Informative | 25 | 12.32% | 74 | 26.43% | 99 | 20.50% |
| Interest | 2 | 0.99% | 7 | 2.50% | 9 | 1.86% |
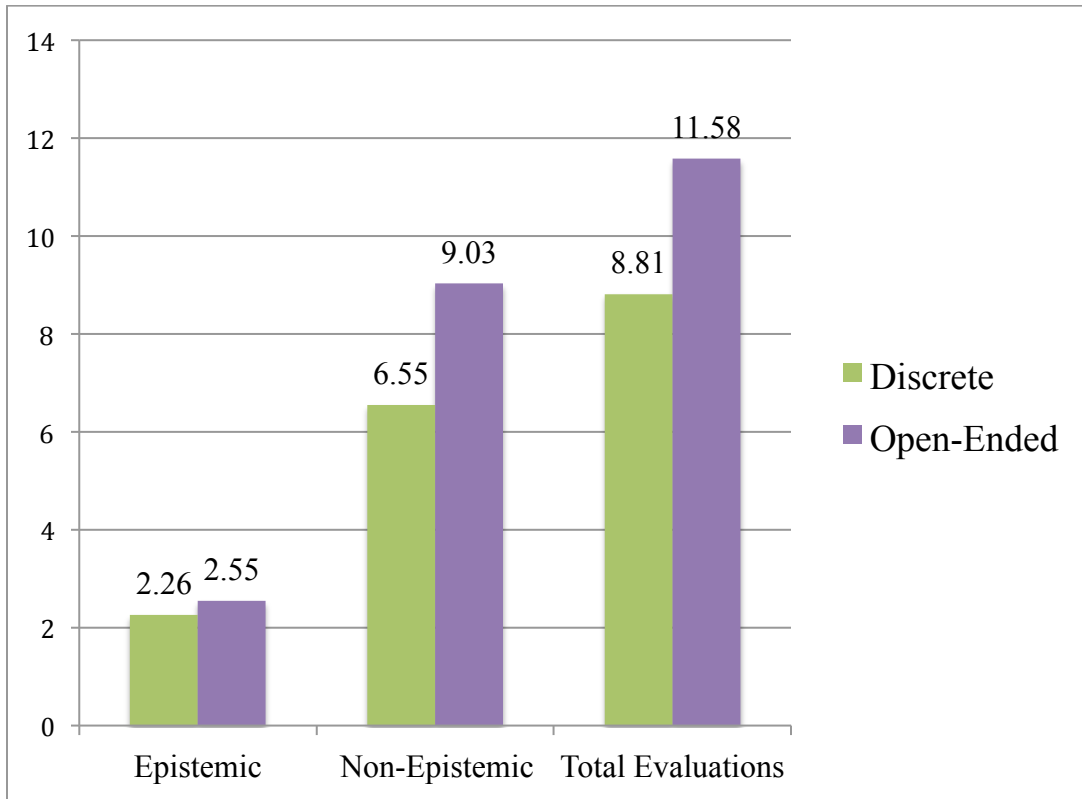
*Figure 1*. Epistemic and non-epistemic source evaluations across question type.
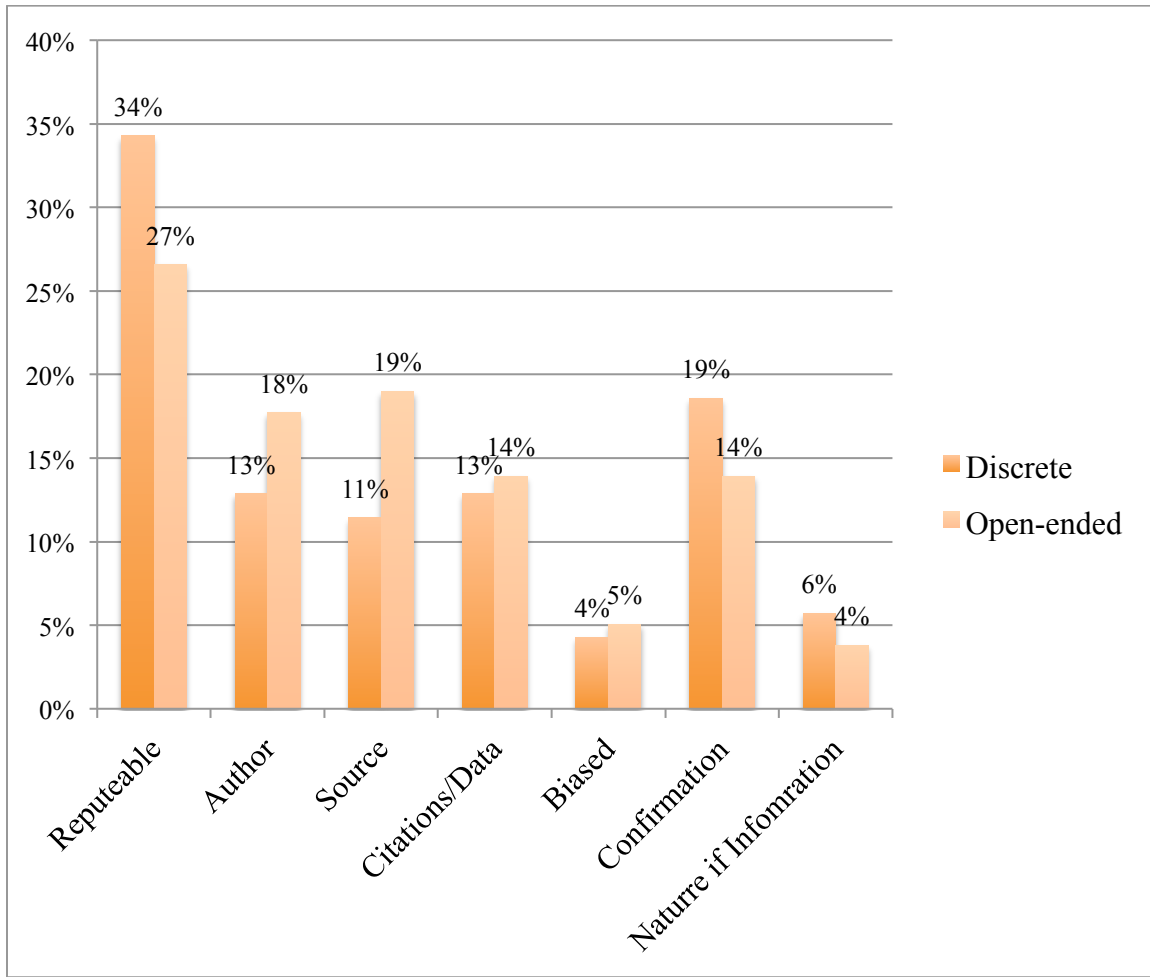
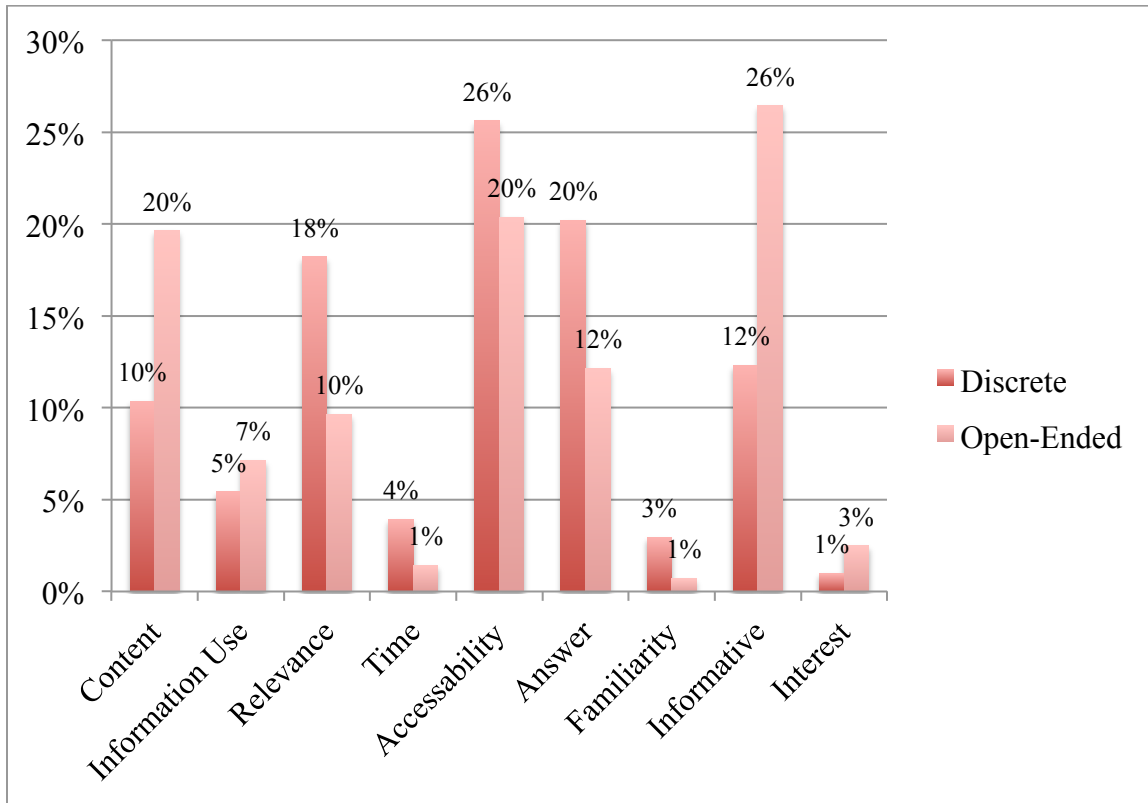*Figure 2.* Epistemic sub-categories of source evaluation by question type.

*Figure 3*. Non-epistemic sub-categories of source evaluation by question type.

## Appendix A
*Screen shot of search task*

**Appendix B**
*Retrospective Interview Questions*

After completion of the search task, students will be asked to reflect on the processes they used to respond to each question. The below questions sequenced participants through each of the websites they accessed, first for question 1, then for question 2.  Questions about sources were followed in the order of log-file data.

Question 1

*Source Use*
      a. Why did you choose X source to answer this question?
      b. What were you expecting from source X?
      c. What did you do when source X loaded?
      d. What did you think of source X/Did source X meet your expectations?
      e. How did you know to stop using this source?
      f. Why did you choose source Y next?
      g. Why did you decide to use only a single source/X sources to answer Q1?
      h. Why did you elect not to use any additional sources to answer Q1?

*Question Answer*
      a. Why did you begin writing your answer before/after/during this source?
      b. How did you know you were done answering Q1?
      c. Were you able to find what you needed to answer Q1?
      d. How confident were you with your response to Q1?

*Overview*
      a. How was the process you used to find information in Q1/Q2 similar or different to the process you usually use to find information on the Internet?
      b. How was the process you used to find information for Q1 similar or different to the process you used to find information for Q2?

Protocol repeated for question 2.

*Concluding Questions:*
      a. What was your experience of search task overall?
      b. Did you find this task interesting or not so interesting?
      c. Did you find this task easy or difficult?
      d. Are there any additional comments you would like to make at this time?