Reading Psychology

ISSN: 0270-2711 (Print) 1521-0685 (Online) Journal homepage: http://www.tandfonline.com/loi/urpy20

Profiling Students' Multiple Source Use by Question Type

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

To cite this article: Alexandra List, Emily M. Grossnickle & Patricia A. Alexander (2016) Profiling Students' Multiple Source Use by Question Type, Reading Psychology, 37:5, 753-797, DOI: 10.1080/02702711.2015.1111962

To link to this article: https://doi.org/10.1080/02702711.2015.1111962

Published online: 09 Dec 2015.

Reading Psychology, 37:753–797, 2016

ISSN: 0270-2711 print / 1521-0685 online
DOI: 10.1080/02702711.2015.1111962

Routledge
Taylor & Francis Group

PROFILING STUDENTS' MULTIPLE SOURCE USE BY QUESTION TYPE

ALEXANDRA LIST, EMILY M. GROSSNICKLE, and PATRICIA A. ALEXANDER
Department of Human Development and Quantitative Methodology, University of Maryland, College Park, Maryland

The present study examined undergraduate students' multiple source use in response to two different types of academic questions, one discrete and one open-ended. Participants ($N = 240$) responded to two questions using a library of eight digital sources, varying in source type (e.g., newspaper article) and reliability (e.g., authors' credentials). Log-data captured students' source use process. Differences were found in the total number and types of sources students accessed when responding to the two questions. Across the questions, cluster analysis identified five consistent patterns of source use; however, the percentage of student in each cluster differed across the two questions.

In a recent survey from the Pew Internet and American Life Project, 70% of Internet users agreed with the statement, "The amount of news and information available from different sources today is overwhelming" (Raine & Purcell, 2010). Considering the constant presence of the Internet in our everyday lives and the volume of websites and blogs constantly updated to provide us with the most in-demand instantaneous information, this sentiment is far from surprising. As never before, there is a need to understand how students navigate the online world as they identify and make sense of the mass of information available from a multitude of sources to answer questions and complete academic tasks (Brand-Gruwel & Stadtler, 2011; Coiro, 2003; Goldman, 2011).

This sense-making process, termed multiple source use (MSU), has been defined as encompassing the identification, use,

Alexandra List is now at Ball State University. Emily M. Grossnickle is now at James Madison University.

Address correspondence to Alexandra List, 3240 Benjamin Building, Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20740. E-mail: alist@umd.edu

comprehension, and evaluation of diverse sources of information to satisfy task goals (Gil, Bråten, Vidal-Abarca, & Strømsø, 2010; Rouet, 2006). Indeed, MSU has been identified as a crucial competency for today's learners (Common Core State Standards Initiative, 2010; Goldman & Scardamalia, 2013) and, as a result, has been integrated into curricula to an ever-increasing extent (Coiro & Kennedy, 2011; Metzger, Flanagin, & Zwarun, 2003). Yet, despite the frequency with which students are asked to engage in MSU to complete academic tasks (Purcell et al., 2012), MSU has been shown to pose a challenge for learners (e.g., Metzger et al., 2003; Wang & Artero, 2005). Students' difficulties with MSU may stem from challenges in selecting from among the volume of available sources (Coiro & Dobler, 2007; Salmerón & Kammerer, 2012) or from the complexity associated with using and evaluating sources differing in type (e.g., newspaper, book) and reliability (Braasch et al., 2009; Bråten, Strømsø, & Britt, 2009).

MSU has been described as task-driven or pragmatic (e.g., Rouet & Britt, 2011). As such, students' processing and use of texts has been shown to vary according to the task, goal, or purpose motivating source use (Anmarkrud, McCrudden, Bråten, & Strømsø, 2013; Brand-Gruwel & Stadtler, 2011). As a task-oriented process, students' interactions with multiple texts are thought to be guided by a desire to satisfy particular task goals (McCrudden & Schraw, 2009) and by the knowledge that a performance measure or task will need to be completed as a consequence of source use (Vidal-Abarca, Mañá, & Gil, 2010). In particular, the task parameter of question type has been emphasized as a factor leading to differences in students' performance on multiple source use tasks (e.g., Le Bigot & Rouet, 2007; Wiley & Voss, 1999). Variations in the types of questions motivating source use have been found to impact students' performance on measures of multiple source use, such as essay writing (Le Bigot & Rouet, 2007; Wiley & Voss, 1999) and comprehension tests (Rouet, 2006; Rouet, Britt, Mason, & Perfetti, 1996).

More recently, research has turned to examining how variations in task impact not only students' performance but also their process of source use (Anmarkrud et al., 2013; Cerdán & Vidal-Abarca, 2008; Cerdán, Vidal-Abarca, Martinez, Gilabert, & Gil, 2009; Rouet, Vidal-Abarca, Erboul, & Millogo, 2001). Common measures of source processing include the total and relative

frequency with which students access sources (e.g., Stahl, Hynd, Britton, McNish, & Bosquet, 1996), the time participants devote to text use (Cerdán et al., 2009; Payne & Reader, 2006), and order of source access, including which sources are opened first (Britt, Rouet, & Perfetti, 1996). Findings suggest that question type affects students' source processing. For example, in responding to questions classified as low- versus high-level, participants varied in the amount of time they devoted to reading texts and composing responses (Cerdán et al., 2009). However, more research is needed to systematically examine the impact of varying question type on students' processing of multiple texts (Rouet & Vidal-Abarca, 2002) and, more specifically, on source selections.

The present study examines how MSU unfolds when students respond to two different types of questions, one discrete (i.e. requiring a single answer located explicitly in text) and one open-ended (i.e. requiring a more elaborated response, integrating various sources of information). In particular, this study examines how students' source access manifests across the two questions. Moreover, as much of the literature has examined indicators of source use independently, such as whether or not students select individual texts (e.g., Stahl et al., 1996), rather than identifying patterns and combinations of source use indicators, we were interested in gaining a more comprehensive insight into students' source use. Therefore, the analyses from the present study used cluster analysis to identify patterns in students' source use by jointly considering the total number of sources students accessed as well as the types and reliability of sources they selected.

Theoretical Frame

The Multiple Documents Task-Based Relevance Assessment and Content Extraction (MD-TRACE) Model, a comprehensive model of MSU, conceptualizes the process of using multiple texts to respond to a query as consisting of five core steps (Rouet, 2006; Rouet & Britt, 2011). In Step 1, students are thought to formulate a *task model*, or cognitive representation of task demands and a plan for how these demands may be satisfied. In Step 2, learners decide what information they need, beyond what they already know, to respond to a given task; this determination of an *information need* serves as the motivator for source use. Step 3 includes

students' source use behaviors, namely their selection, processing, and integration of multiple texts (Rouet, 2006; Rouet & Britt, 2011). In Step 4, students formulate a textual product, and in the fifth and final step, they make the determination of whether task demands have been satisfied (Rouet & Britt, 2011).

Rouet (2006) referred to the MD-TRACE model as "relevance based" (p. 105), in that students' interactions with sources, outlined in Step 3, are based on their perceptions of task demands. Specifically, the MD-TRACE model establishes task as instrumental in both students' multiple source use (e.g., text selection, processing, and integration) and response formulation (Rouet, 2006; Rouet & Britt, 2011). Given that both source use behaviors (Step 3) and products (Steps 4 and 5) are conceptualized and evaluated by learners in reference to task demands, it is necessary to examine how variations in task features (e.g., question type) impact students' interactions with multiple texts. Guided by the MD-TRACE model, the present study will examine how students' source selections and process of source use differ when students respond to two questions, one discrete and one open ended.

Question Type

Research examining task effects on MSU, or students' task-oriented source use (Anmarkrud et al., 2013), has classified the tasks or questions guiding students' interactions with texts along a number of dimensions (see Andre [1979] and Graesser, Rus, and Cai [2008] for a review). Questions motivating students' source use have been contrasted along three primary dimensions: (a) *location*, (b) *integration*, and (c) *response form*. Location refers to where in the text students may be expected to locate information and how many pieces of information are needed to produce a response. Integration refers to the level of abstraction that questions demand and the extent to which students' responses are expected to include inferences beyond what is explicitly stated in texts. Finally, response form refers to the type of answer students are expected to select or generate; response forms may range from questions asking students to choose from several multiple choice options to prompts asking students to compose elaborated essays. Indeed, the literature is replete with varied terms and definitions

distinguishing the types of questions or tasks that may be assigned to students. Questions have been dialectically referred to as general or specific (Rouet et al., 2001), high-level or low-level (Rouet et al., 2001), inference based/global or text-based/explicit (Vidal-Abarca & Sanjose, 1998). See Table 1 for a partial listing of terms that have been used in the literature to describe the questions motivating students' source use and their definitions.

Tasks have been differentiated based on whether they ask students to locate a specific piece of information in text (e.g., specific, Rouet et al., 2001; text-based/explicit, Vidal-Abarca et al., 1996) or to consider multiple pieces of information across texts (e.g., general, Rouet et al., 2001; inference based/global, Vidal-Abarca & Sanjose, 1998; feature match, Guthrie, 1988). Tasks have also been distinguished by the level of processing they elicit from students, with low-level (Cerdán et al., 2009) or text explicit (Wixson, 1983) questions thought to require learners to engage only in low-level or superficial processing (e.g., finding a specific answer in texts) and high-level (Cerdán et al., 2009) or text implicit/inference questions (Wixson, 1983) requiring students to engage in higher-level processing, such as making inferences or drawing conclusions (e.g., integrating information across multiple texts).

Importantly, questions distinguished in these ways result in differences in students' text processing. Two studies examining the effects of high-level versus low-level questions on students' source use (Cerdán et al., 2009; Rouet et al., 2001) found that students confronted with low-level questions quickly scanned several paragraphs, whereas students responding to high-level question paused longer on more paragraphs. Similar findings have been reported elsewhere (e.g., Cerdán & Vidal-Abarca, 2008; Rouet, 2003). Likewise, in studies by Wiley and colleagues (Voss & Wiley, 1997; Wiley & Voss, 1999), students assigned various task goals (i.e. writing a narrative versus argumentative essay) demonstrated differences in text understanding when assigned to write descriptive versus argumentative essays. Those students in the descriptive condition devoted more time to visiting unreliable source pages (Wiley et al., 2009).

In this study, we elected to ask students to respond to two questions, one discrete and one open-ended. The discrete question was designed to require students to identify a specific piece of information explicitly found in text, whereas the open-

TABLE 1 Types of Questions Considered in MSU Studies and Definitions

Study	Question Types and Definition		
Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Rouet, Vidal-Abarca, Erboul, & Millogo, 2001 from Rothkopf & Bisbicos, 1967	*High Level:* deep comprehension; integration of text and inferences; broader set of concepts; higher-level processing		*Low Level:* single concept, superficial processing
Dreher & Guthrie, 1990	*Simple Questions*: answered in one step		*Complex Questions:* answered in multiple steps
Hofman & van Oostendorp, 1999	*Text-base:* recognition		*Situation model*: inference questions
Rouet et al., 2001	*General:* several pieces of text		*Specific:* single piece of text
Vidal-Abarca, Mengual, Sanjose, & Rouet, 1996	*Inference based/Global:* refer to several pieces of text and ask to draw conclusions		*Text based/Explicit:* single section of text; response explicitly stated in question
Wixson, 1983	*Text explicit*: identification questions	*Text Implicit*: inference questions	*Scriptally Implicit:* schema based; knowledge based

ended question was designed to require students to reconcile and integrate information across texts to develop a more elaborated response. In effect, these types of questions were chosen because they were both expected to require students to engage in the process of multiple source use but differently. In particular, while the discrete question required the consultation of multiple sources for response verification or corroboration, the open-ended question required multiple source use for students to be able to formulate complete and comprehensive responses that integrated information from across multiple texts.

Source Navigation

A variety of approaches have been adopted to investigate students' MSU, including think-aloud (e.g., Bråten & Strømsø, 2003; Mason, Boldrin, & Ariasi, 2010) and eye-tracking (e.g., Kammerer & Gerjets, 2012) methodologies. The limitations associated with these methods, including their time consuming nature, often result in samples limited in size (Eveland & Dunwoody, 2000; Vidal-Abarca et al., 2011). As such, work has turned to examining students' processing of multiple texts using log-data when students engage in MSU within a digital- or web-based interface (e.g., Cerdán et al., 2009). When considering log-based measures of MSU, summary indicators of source access such as the total number of texts selected (Ford & Chen, 2000) and the total and unique number of pages visited (Wallace, Kupperman, Krajcik, & Soloway, 2000) have most frequently been used. For example, Pieschl, Stahl, and Bromme (2008) examined the average time students devoted to using a particular node (i.e. text) and the percentage of nodes students accessed within a hyperlinked system.

The majority of MSU indicators have been behavioral in nature, focusing on what students do when accessing multiple texts (e.g., Brand-Gruwel, Wopereis, & Vermetten, 2005; Wallace et al., 2000). Limited studies have examined quality indicators of source use, such as what kinds of sources students access and use when completing MSU tasks. One of few studies that addressed quality indicators, Cerdán et al. (2009), examined the percentage of total source use time students spent reading paragraphs classified as relevant. Yet, more research is needed to understand the nature of the sources students select. While studies examining

characteristics of students' selected sources have extensively considered indicators of source relevance (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011; Wilkinson & Payne, 2006), there is a need to further analyze students' source selections along dimensions related to source quality and reliability (Gerjets, Kammerer, & Werner, 2011).

Sources

A variety of source characteristics have been considered to be pertinent to contextualizing and evaluating a particular text, including source type (e.g., textbook), author, publisher, and date of publication (Bråten et al., 2009; Britt, Rouet, & Braasch, 2013). Consideration of source information has been found to be necessary for students' processing, comprehension, and evaluation of texts both individually and within multiple source use contexts (Bråten et al., 2009; Britt, et al., 2013), allowing students to compare, contrast, and integrate information from various texts (Bråten, Britt, Strømsø, & Rouet, 2011). These document characteristics (e.g., author, source type) have been theorized to convey information, beyond the specific contents of a text, by helping students interpret the information in a source and relate a given source to other texts (Britt et al., 2013; Strømsø et al., 2010). Thus, in evaluating students' source use, examining not only how many but which sources students select has the potential to be particularly informative in understanding students' MSU. In the present study, we were interested in examining not only the total number of sources students accessed but also the type and reliability of the sources they selected. As in other work, source characteristics will be considered collectively (Britt et al., 2013), as all of these characteristics, particularly source-type (Bråten et al., 2009), have been found to contribute to determinations of relative source reliability. The reliability of sources was determined based on authors' credentials as well as whether or not data were presented. These two source characteristics (i.e. authoritativeness and data inclusion) have been established as reflective of overall source reliability in prior research (Mason et al., 2010).

Patterns of navigation

Studies have typically investigated the use of sources independently, for instance, examining whether specific sources are

selected for use in a particular task (Stahl et al., 1996). Few have addressed students' source navigation to identify patterns or profiles of source use. Guided by the MD-TRACE model, which posits MSU as a process involving the integrated use of multiple sources, we were interested in developing profiles of students' MSU, built on source indicators of both quantity (i.e. total number of sources) and quality (i.e. sources accessed). This is consistent with recent efforts in the MSU literature to adopt a more person-centered approach to understanding MSU (Ferguson & Bråten, 2013).

While few studies have approached source use in this way, certain exceptions provided us with a framework to approach our investigation. For instance, in a descriptive study, Wallace et al. (2000) characterized sixth-grade students' approach to MSU globally, as guided by desires to "find the perfect page" or "get the right number of hits," referring to students' desires to identify a large volume of information or to "find the answer" (pp. 84–85). Additionally, Reader and Payne (2007) identified two patterns of access when undergraduate students used multiple texts: sampling and satisficing. When employing a sampling strategy in their MSU, participants visit each of a number of texts, not to learn content but to learn about each of the texts themselves, before choosing a preferred source. When adopting a satisficing approach to MSU, students focused on gaining information about a particular topic, and as long as a particular text provides relevant information, they will continue to use it. While in sampling sources, students demonstrate a concern with identifying the best source along some dimension; in a satisficing approach to source use, students are limited in their source evaluations and are only concerned with a particular source meeting basic standards of utility (Reader & Payne, 2007). This identified differentiation in students' patterns of MSU further points to the need to consider not only summative measures of source use (e.g., total number of source accessed) but also qualitative characteristics of the sources students elect to access.

The particular orientations toward MSU that students adopt are partially dictated by the characteristics of available texts and the task directions. For example, the quality of texts in a set impacts students' adoption of a sampling approach to MSU. Likewise, the number of texts in a set that are good enough

informs the extent to which students' satisficing needs may be met (Reader & Payne, 2007; Wilkinson, Reader, & Payne, 2012). Further, Reader and Payne (2007) found that students' source use patterns were differentially instantiated depending on differences in task features, for instance, whether or not students were told that following source use they would be asked to complete performance measures. In the present study, we were interested in considering how students' patterns of MSU differed when learners were asked to respond to discrete versus open-ended questions.

Present Study

In the current study, we investigated differences in source use when students responded to two questions, one discrete and one open-ended. This study expanded on the prior literature by investigating not only how variations in task affect individual indicators of students' multiple source use (i.e. total number of sources accessed, first-source accessed) but also whether students exhibit differential patterns of source use when responding to the two questions. Cluster analysis was used to identify patterns in students' overall source use by classifying students into profiles based on the number and type(s) sources they accessed when responding to the two questions. Consistent with prior literature, sources type and reliability were varied across the texts provided to students in a digital library (Braasch et al., 2009; Bråten et al., 2009; Mason et al., 2010) and the sources students accessed were used in cluster analyses as indicators of text trustworthiness (Bråten et al., 2009) and authoritativeness (Mason et al., 2010). With these goals in mind, we sought to address the following research questions:

1. How does students' MSU (number of sources used, types of sources accessed) differ when responding to a discrete versus open-ended question?
2. In response to the discrete question, which profiles of source use emerge based on the number of sources used and types of sources accessed by students?
3. In response to the open-ended question, which profiles of source use emerge based on the number of sources used and types of sources accessed by students?

4. How does source-use cluster membership relate to response accuracy, in the case of the discrete question, and response quality, in the case of the open-ended question?

Methods

Participants

Two-hundred-forty undergraduate students from a large university in the United States participated in the study. Participants were recruited from human development and educational psychology courses over several semesters, and in some cases, instructors awarded extra credit for study participation. The sample had a mean age of 20.85 years ($SD = 2.32$) and was 82.5% female ($N = 198$; 17.5% male, $N = 42$), representative of the gender make-up of the classes from which participants were recruited. The sample consisted of 61.67% of participants identifying as White ($N = 148$), 15.83% as Asian ($N = 38$), and 13.33% as Black ($N = 32$); 4.58% Hispanic/Latino ($N = 11$) and 4.58% as other. Participants were primarily social science majors (82.08%, $N = 197$) with a portion of students majoring in the natural sciences (11.25%, $N = 27$) and the humanities (6.25%, $N = 15$). One student did not report a major. Students reported an average GPA of 3.17 ($SD = 0.48$) on a four-point scale, and reported having earned an average of 77.73 credits ($SD = 23.77$), with 12 credit hours per semester being the requirement for full-time enrollment. Four students did not report GPA and one student did not report the number of completed credit hours.

Measures

Search task

Participants were asked to respond to two academic questions, one discrete and one open-ended, on the topic of fertility rates, using a library of eight digital sources. We expected the topic of the task, fertility rates, to be one of interest and relevance to our sample, composed primarily of social science majors recruited from courses in the department of human development.

The discrete question was, "In the USA, what is the replacement fertility rate?" and the open-ended question was, "What

is the role of government-sponsored childcare in high fertility rates?" These questions were designed to be typical of those assigned in a university course. Questions were counter-balanced, with students randomly assigned to answer either the discrete ($N = 122$) or the open-ended ($N = 118$) question first. Given that both questions addressed the same topic and utilized the same library sources, in answering the first question, students may have encountered information pertinent to the second question. Therefore, only data from the first question with which students were presented (i.e. discrete or open-ended) were analyzed in order to compare source use across question types. Participants were instructed to answer the questions as if they were doing so for a social science class and to respond to the questions using the library of sources provided. They were instructed that sources were presented in alphabetical order and that no time limit was placed on the task. Participants had to submit their responses to the first question before they were able to proceed to the second.

The study was conducted online. Participants could access the study website via a link and could complete the study at a place and time convenient to them. Students were asked to set aside the required time to complete the study in one sitting and were instructed not to leave the site (i.e. not access any additional webpages) until the study was completed.

Source library

The source library included a list and brief descriptions of eight sources that could be used to answer either question and was designed to resemble a Google search engine results page. A description of each source was included in the library, in addition to source information, including the title and source type (e.g., website, PDF book chapter), the author or publisher (e.g., National Research Council), and date of publication. The URL for each source was also provided to students. All texts were naturally occurring, and the descriptions for each source were determined based on a Google search for that text.

The order of the sources was alphabetical and remained constant for all participants. Participants could access the sources by clicking on the hyperlinked title for each. When viewing any source, there was a button that allowed participants to return to the source library. The question students were responding to and

a textbox into which they could type their answer were located at the top of the screen so that students could see the question and compose their answer while engaged in source use. A sample screenshot of the study interface is included as Appendix A.

The sources consisted of three websites—a conservative blog, a Wikipedia entry, and a report from a policy institute—two PDF book chapters, one article from a magazine, one newspaper article, and one PDF journal article. These texts represented the range of information and sources students may be expected to encounter during the course of researching on the Internet. The Flesch-Kincaid Reading Ease and Flesch-Kincaid Grade Level measures were used as readability quotients. On average the eight sources had a Flesch-Kincaid Reading Ease score of 35.39 out of 100, with the *New York Times* article receiving a reading ease score of 39 and lower scores indicating grater text difficulty (Flesch, 1979). Flesch-Kincaid Grade Level ranged from 10.5–12.0, suggesting that sources were appropriate for use with an undergraduate sample (Kincaid, Fishburne, Rogers, & Chissom, 1975).

Sources

The first source, *Global Politician*, was a conservative website that provided blog and article postings and described itself as an, "independent online international political news" source. Source 2, *Beyond Six Billion,* was a chapter titled "Reasons for Fertility Decline," from a book published by the National Research Council through the National Academic Press. The third source was a chapter "What Determines Fertility," from the book *Six Billion Plus: World Population in the 21st Century.* A *Newsweek* magazine article titled, "The End of Motherhood" was the fourth source, and the Wikipedia entry for "Total Fertility Rate" the fifth. The sixth source was a report from the Population Reference Bureau (PRB), titled "Tracking Trends in Low Fertility Countries." PRB is a non-profit organization that provides policy-relevant data. 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 newspaper article from the Washington Post, "With Each French Birth, A Dividend From the State." Table 2 presents a summary of sources included in the library.

TABLE 2 Texts Included in Source Library

Title	Source	Source Type	Classification	Flesch Reading Ease	Flesch-Kincaid Grade Level
What is the Cause of Low Birth Rates	Global Politician	Blog	Low Reliability	45.6	11.3
Reasons for Fertility Decline	Book: *Beyond Six Billion,* National Research Council	Book Chapter	High Reliability	23.6	12.0
What Determines Fertility?	Book: *Six Billion Plus: World Population in the 21st Century*	Book Chapter	High Reliability	18.2	12.0
The End of Motherhood	*Newsweek*	Magazine Article	Popular Press	40.2	10.5
Total Fertility Rate	Wikipedia	Online Encyclopedia	High accessibility; moderate reliability	28.4	12.0
Tracking Trends in Low Fertility Countries	Population Reference Bureau	Research Report	High accessibility; moderate reliability	44.6	11.6
Very Low Fertility: Consequence, Causes, and Policy Approaches	Japanese Journal of Population	Journal Article	High Reliability	42.7	11.9
With Each French Birth, A Dividend From the State	*Washington Post*	Newspaper Article	Popular Press	39.8	12.0

The sources were chosen to: (a) represent a variety of source types, (b) differ in reliability (i.e. based on authors' credentials and data inclusion), and (c) offer a variety of perspectives and information pertinent to the two questions. Each of the sources was relevant to both of the questions (i.e. included pertinent keywords) and provided an answer to each of the questions; however, these answers were of varied quality. Multiple sources needed to be accessed to corroborate information, in the case of the discrete question, or to compose a comprehensive response, in the case of the open-ended question. The sources provided complimentary, and at times conflicting, information pertinent to the two questions and focused on varied aspects of the issue. For example, while the book chapter, *Beyond Six Billion*, discussed fertility in terms of changing demographic patterns, the journal article discussed fertility as related to economic development, and the *Washington Post* newspaper article discussed how government policies impact women's work-life balance.

Based on the content and style of the source, the Global Politician blog was considered to be the least reliable source in the set in part because it was written in first person and provided no evidence or citations to support authors' claims. The journal article and the two book chapters were selected to be highly authoritative sources. These sources were written by experts, presented data to support arguments, included citations, and were written using academic language. *The Washington Post* and *Newsweek* articles were selected as popular press sources that were more engaging for students to read (e.g., included compelling anecdotal stories) but were non-scientific in nature. Finally, the Wikipedia entry and the report from the Population Reference Bureau were selected as sources of moderate reliability. These sources included data, but the data originally came from external sources. Neither the Wikipedia entry nor the PRB article provided any author listing, and the texts were written in deliberately neutral and objective language. These sources also presented information in highly accessible and clearly formatted ways. Sources selected were piloted with undergraduate students to assure their readability and relevance in responding to target questions.

NAVIGATION DATA

The sources students accessed and the order of source access (i.e. navigation) were recorded in a database. Time devoted to source use was intended to also be collected, but due to site malfunction, time data were not recorded. A number of derived indices of source use were computed based on access data. First, a variable *used high reliability sources* was computed to capture whether or not students accessed any of the sources identified as having high reliability (i.e. the journal article and the two book chapters). Next, *used popular press sources* captured whether students selected the magazine or the newspaper article, and *used accessible sources* reflected whether or not students used either the Wikipedia article or the report from the PRB.

Many students accessed the Global Politician source, which was presented first in the library and did not appear, from its description, to be low in reliability. Therefore, rather than examining whether or not students selected this source, we were interested in whether or not students accessed this unreliable source and then did not access any additional sources to further substantiate or corroborate the information found in the Global Politician. *Used unreliable source* captured whether students accessed only the Global Politician with no subsequent verification.

Each of these four derived indices (e.g., used high reliability sources) was dichotomously coded, rather than ordinal. This was done for two reasons. First, while some of the indices could have been ordinal in nature (e.g., students could have accessed up to three sources high in reliability), other indices were based on whether or not students accessed only a single source (e.g., used unreliable source). Second, there was a concern that including ordinal measures of the types of sources accessed would lead to redundancy with the continuous predictor, number of sources accessed, particularly for students who accessed a relatively limited number of texts.

Response Coding

The discrete question was dichotomously coded as correct or incorrect based on a replacement fertility rate value of 2.1 reported in the sources. Responses to the open-ended question were coded along five dimensions: (a) word count (i.e. number of words that

students included in their responses), (b) the number of arguments (i.e. unique claims and warrants), (c) elaborations (i.e. arguments expanded with additional evidence, examples, or explanations), (d) citations (i.e. direct references to sources either through in-text citations or specific mention of the source name or author), and (e) response scores coded using the SOLO taxonomy (Biggs & Collins, 1982). These indices were based on previous coding schemes of students' open-ended responses to multiple source use tasks (e.g., Gil et al., 2010). It should be noted that students were not explicitly instructed to cite sources, thus the number of citations included in open-ended responses represents spontaneous sourcing, which has been identified as an important indicator of quality of multiple source use and text integration (Strømsø, Bråten, Britt, & Ferguson, 2013).

The SOLO taxonomy (Biggs & Collins, 1982) is a coding scheme that ranks students' written responses according to the number of pieces of relevant information they included, as well as whether the information is integrated and reasonable conclusions are drawn. The scale ranges from 0 to 4, including half-points, with 4 as the highest score. A score of 1 on the SOLO taxonomy corresponds to students providing a single relevant piece of information, and a score of 2 corresponds to students providing multiple pieces of relevant information. A score of 3 corresponds to responses that include multiple pieces of relevant information that are also integrated and internally consistent, and a 4 is reserved for students' responses that include multiple pieces of linked and integrated information, from which reasonable conclusions or generalizations are drawn. Half points were assigned if students were aiming for a certain level of response quality but did not achieve this desired level. For example, if students were trying to provide multiple pieces of relevant information but did not clearly articulate this information, a score of 1.5 would be assigned. The complete scoring rubric is included as Appendix B.

The first and second author coded 20% of the SOLO responses, with an exact agreement of 70%. The mean discrepancy was 0.30, with a standard deviation of 0.42, indicating that the discrepancies were less than a half-point apart on average. Any disagreements were resolved by discussion, and the first author then coded the remainder of the answers to the open-ended question. Together, these scoring dimensions capture attributes related to

both quantitative and qualitative aspects of students' open-ended responses and considered both the quality of the information that students included in their responses and the composition of the responses themselves.

Analysis

In addition to examining differences in students' source use in response to discrete versus open-ended questions using traditional inferential statistical methods, a more exploratory technique, cluster analysis, was used to uncover profiles of students' process of multiple source use. Cluster analysis is a procedure that is used to find groupings (i.e. clusters) of elements (e.g., students) in multivariate data sets with the purpose of classifying elements into groups such that there is homogeneity within groups and heterogeneity between groups (Milligan & Hirtle, 2013). Unlike classification techniques (e.g., discriminant function analysis) that seek to predict known category membership, cluster analysis is a technique to identify previously unknown groupings or categories in the data (Rokach & Maimon, 2005). As such, cluster analysis was considered to be an appropriate exploratory technique for this initial investigation of students' MSU in response to discrete versus open-ended questions.

A hierarchical clustering method, two-step cluster analysis, was selected because it offered a number of key advantages. First, it automatically determined a number of clusters to be retained, without the need for researcher specification, as is needed for other clustering methods (e.g., *k* means cluster analysis). Second, it is a method adept at creating clusters based on a combination of discrete and continuous variables (Bacher, Wenzig, & Volger, 2004; Norušis, 2005). In addition to the continuous clustering variable, total number of source used, four dichotomous clustering variables were included in analyses (e.g., used high reliability sources, coded as 0 or 1). A final advantage to adopting a two-step algorithm is that data are provided about the relative importance of each clustering variable in determining cluster membership to ensure that extraneous variables are not entered into the cluster analysis (Mooi & Sarstedt, 2011).

Because cluster analysis is an exploratory technique, it is important to consider the theoretical and practical validity of

identified clusters through external and internal cluster validation (Milligan, 1996). To validate clusters internally, it is recommended to consider indices of similarity and dissimilarity between the clusters and to determine whether created clusters significantly differ from one another along each of the key clustering variables (Milligan, 1996). In the present study, internal validation was conducted by (a) considering the silhouette measure of cluster similarity and dissimilarity and by (b) determining whether clusters significantly differed from one-another along target clustering variables (e.g., number of sources used).

For external validation, clustering may be replicated on a secondary data set and clusters may be examined to determine whether they significantly differ from one another with respect to exogenous variables not used in the original clustering. While it is typically recommended to validate clusters on either two data sets or a single partitioned dataset (Pastor, 2010), this was not done in the present study; rather, clustering was verified across question type. It was thought that creating additional holdout datasets would be overly taxing on sample size, as our sample was already partitioned across the two questions, limiting sample size for initial clustering. Sample size was further a concern as students in our sample were restricted in their source use, constraining the variance of the clustering variables, and therefore necessitating as large a sample as possible to successfully identify clusters. Instead, we examined the consistency of source use clusters emerging for students responding to the discrete versus the open-ended question to be a viable replication and validation method. Comparing clustering across the two question types allowed us to examine cluster consistency across different task conditions, an acceptable form of validation (Pastor, 2010). In other words, our research questions regarding the consistency of clustering across the two questions also served a validating purpose. Further, clusters were externally validated by examining the extent to which identified clusters differed in their performance when responding to the two questions. For the discrete question, we examined the extent to which clusters differed in their rates of response accuracy; for the open-ended question, we considered the extent to which clusters differed in the quality of responses produced. In addition to these statistical determinations of cluster validity, clusters were theoretically validated.

TABLE 3 Descriptive Statistics of Sources Accessed

	Total		Discrete		Open	
	Percent	N	Percent	N	Percent	N
Unreliable Source						
Global Politician (Blog)	46.67	112	48.36	59	44.92	53
Reliable Sources						
Beyond Six Billion (Book)*	22.50	54	28.69	35	16.10	19
Six Billion Plus (Book)	22.08	53	21.31	26	22.88	27
Very Low Fertility (Journal)	13.75	33	16.39	20	11.02	13
Popular Press Articles						
End of Motherhood (Magazine article)	17.50	42	20.49	25	14.41	17
With Each French Birth (Newspaper article)	8.33	20	5.74	7	11.02	13
Accessible Sources						
Total Fertility Rate* (Wikipedia)	37.92	91	55.74	68	19.49	23
Total Fertility Rate Trends (Report)	10.42	25	11.48	14	9.32	11

Note: * indicates significant different across question conditions at $\alpha < 0.05$.

Results and Discussion

Overview of Multiple Source Use

Across the two questions, students used an average of 1.79 sources ($SD = 1.55$), ranging from zero to eight sources. Of the eight sources included in the library, the most popular source was the blog, the Global Politician, accessed at least once by 46.67% of all participants ($N = 112$). See Table 3 for a breakdown of students' source access across question types.

In the case of the discrete question, 74.59% of students answered the question correctly ($N = 91$). For the open-ended question, students' responses were an average of 53.92 words in length ($SD = 49.71$), ranging from 4 to 368 words. Students included an

average of 1.90 arguments in their responses ($SD = 1.53$), ranging from 0 to 9 arguments. On average, 0.68 of these arguments were elaborated ($SD = 0.90$), ranging from 0 to 4 elaborations. On the SOLO taxonomy, students earned an average score of 1.76 ($SD = 1.00$), indicating that students were attempting to provide multiple pieces of relevant information in their responses but were not doing so effectively. Finally, students' responses included an average of 0.20 ($SD = 0.80$) citations, ranging from zero to seven sources cited.

Source Use Across Question Type

An independent sample *t*-test with adjusted degrees of freedom to compensate for violations of the homogeneity of variance assumption was conducted to examine whether students responding to the discrete versus open-ended question differed in the total number of sources they used. Students used significantly more sources in responding to the discrete ($M = 2.08$, $SD = 1.73$) compared to the open-ended question ($M = 1.49$, $SD = 1.28$), $t(222.97) = 3.02$, $p < 0.01$. Cohen's $d = 0.39$, corresponding to a medium effect size.

A series of chi-squared analyses were conducted to determine whether there were differences in the percentage of students who used (a) reliable sources, (b) popular press sources, (c) accessible sources, or (d) unreliable sources with no subsequent verification, when responding to discrete versus open-ended questions. Significantly more students selected reliable sources when responding to the discrete question (28.69%, $N = 35$) rather than the open-ended question (16.10%, $N = 19$), $\chi^2(1) = 5.45$, $p < 0.05$. Likewise, more students selected accessible sources when responding to the discrete question (55.74%, $N = 68$) than to the open-ended question (19.49%, $N = 23$), $\chi^2(1) = 33.48$, $p < 0.001$. Cramer's V was used as a measure of effect size. Difference in accessing reliable sources corresponded to a small effect ($V = 0.15$), while differences in selecting accessible sources corresponded to a medium effect ($V = 0.37$). Differences across question type were not identified for accessing popular press sources, $\chi^2(1) = 1.54$, $p = 0.22$, or accessing the unreliable source with no subsequent verification, $\chi^2(1) = 0.35$, $p = 0.55$.

Cluster Analysis

Cluster analyses were run to characterize students' source use in response to the discrete question and the open-ended question. Five clustering variables were used. These included one continuous predictor (i.e. total number of sources used), and four binary predictors capturing use of: (a) reliable sources, (b) popular press sources, (c) accessible sources, and (d) an unreliable source with no subsequent verification. Descriptive statistics for each clustering variable are presented in Table 4.

Discrete clustering

Based on students' source use in response to the discrete question, six clusters were identified. Broadly speaking, clusters were distinguished along two dimensions: whether or not students chose sources considered to be reliable and whether or not students had a high level of quantitative engagement in source use (i.e. accessed a comparatively large number of sources).

Cluster 1, the *accessibility* cluster, included 32.78% of students ($N = 40$) who primarily selected sources characterized by their accessibility (100% of students in this cluster accessed such sources). Overall, these students were modest in the number of sources they accessed (cluster center: 1.32 sources). Students in this cluster seemed concerned with efficiently finding an answer and selected sources in which they could most easily find an appropriate response. These sources were sufficient but not very reliable, and these students did not seem to extensively engage in source use or verification. Students in Cluster 2 (16.39%, $N = 20$) were termed *disengaged* source users. In this cluster were students who did not access any of the various types of sources and whose total source use was lowest of all of the clusters (cluster center: 0.80 sources), suggesting that this cluster consisted of students not engaged in source use and in the task. In the remaining five clusters, students seemed to be participating in the source use process, and the identified clusters differentiated the nature of this participation.

Students in the *engaged* source use cluster, Cluster 3 (8.20%, $N = 10$), were those who accessed popular press articles (100% of students accessed such sources) and used a fairly high number of sources (cluster center: 2.70 sources), relative to the sample. Students in this cluster were potentially driven by the engaging

TABLE 4 Descriptive Statistics for Clustering Variables

	Continuous Variable					
	Total		**Discrete**		**Open**	
Mean (*SD*) number of sources used	1.79 (1.55)		2.08 (1.73)		1.49 (1.28)	
Range of total sources used	0–8		0–8		0–6	
	Binary Variables					
	Total N = 240		**Discrete N** = 122		**Open N** = 118	
	Percent accessed	N	Percent accessed	N	Percent accessed	N
Highly reliable sources used	22.50	54	28.69	35	16.10	19
Accessible sources used	37.92	91	55.74	68	19.49	23
Popular press sources used	17.50	42	20.49	25	14.40	17
Unreliable sources used	15.83	38	17.21	21	14.40	17

nature of the popular press articles to also read a high number of sources. Students in Cluster 4 (13.93%, $N = 17$) were termed *non-critical* source users. These were students who accessed the unreliable blog source (100% of students accessed this source) and did not access any additional sources to verify the information found in the Global Politician (cluster center: 1.18). Conversely, Cluster 5 (13.93%, $N = 17$) included students who were *critical analytic* source users. These were students who not only selected highly reliable sources (100% of students accessed such sources) but also used a high number of sources (cluster center: 3.29 sources) to verify their responses.

In the final cluster, Cluster 6 (14.75%, $N = 18$), students seemed to be a combination of two previously identified clusters, using both reliable and accessible sources (100% of students accessed both reliable and accessible sources) and accessing the largest number of texts (cluster center 4.56 sources). These students were termed *comprehensive* source users because they seemed to be driven both by the desire to use highly reliable sources and corroborate information across texts and to find information with ease. Table 5 includes cluster descriptions and the relative importance of each clustering variable in determining students' cluster membership.

Discrete clustering internal validation

The silhouette measure of cohesion and separation (ranging from 0 to 1 with 1 representing the most distinct clusters) was 0.7 for this cluster solution, considered to be good. A series of χ^2 tests were run to determine whether the identified clusters differed significantly along each of the indicators of source use. Clusters were significantly different in the percentage of their constituent students who accessed reliable sources, $\chi^2(5) = 122.0$, $p < 0.001$; popular press sources, $\chi^2(5) = 69.45$, $p < 0.001$; accessible sources, $\chi^2(5) = 97.56$, $p < 0.001$; and unreliable sources without subsequent verification, $\chi^2(5) = 98.39$, $p < 0.001$. Further, an ANOVA determined clusters differed significantly in the total number of sources that students in each cluster accessed, $F(5, 116) = 31.74$, $p < 0.001$. However, these results should be interpreted with caution, as clusters were developed to differ along these indicators.

TABLE 5 Discrete Cluster Membership

Cluster	1	2	6	4	5	3
Label	Accessible Source Users	Disengaged	Comprehensive Source Users	Non-Critical Source Users	Critical Analytic Source Users	Engaged Source Users
Description	Accessible sources used; Low number of sources accessed	Very limited source engagement.	Reliable and accessible sources used; high number of sources accessed.	Unreliable source used; Low number of sources accessed	Reliable source used; High number of sources accessed.	Popular press source used; moderate number of texts accessed
Size	32.79% (N = 40)	16.39% (N = 20)	14.75% (N = 18)	13.93% (N = 17)	13.93% (N = 17)	8.20% (N = 10)
	Inputs (Predictor Importance) Cluster Centroid (Percentage Students at Centroid)					
Reliable Sources (1.0)	0.00(100%)	0.00(100%)	0.00(100%)	0.00(100%)	1.00(100%)	0.00(100%)
Source Total (0.8)	1.32	0.80	4.56	1.18	3.29	2.70
No Verification (0.8)	0.00(100%)	0.00(100%)	0.00(100%)	1.00(100%)	0.00(88.2%)	0.00(80.0%)
Accessibility (0.8)	1.00(100%)	0.00(100%)	1.00(100%)	0.00(70.6%)	0.00(100%)	1.00(50%)
Popular Press (0.6)	0.00(100%)	0.00(100%)	0.00(55.6%)	0.00(100%)	0.00(58.8%)	1.00(100%)

Note: Clusters ordered by size.

Open-ended clustering

Based on students' source use in response to the open-ended question, a five-cluster solution was identified. The largest cluster, Cluster 1, (47.5%, $N = 56$) included participants who were *disengaged* in source use and in the task. These were students who did not access any of the different types of sources and who accessed the lowest number of sources (cluster center: 0.71 sources). In the remaining clusters, students did engaged with sources and the task but differed in the nature of their engagement.

Participants in Cluster 2 were termed *non-critical* source users (13.56%, $N = 16$). These were students who used the unreliable source (100% of students accessed this source) without further verification and, in general, used a limited number of sources, suggesting limited corroboration of information (cluster center: 1.44). The smallest cluster, Cluster 3, included 11.86% of students ($N = 14$). Students in this cluster, termed *critical analytic* source users, both selected highly reliable sources (100% of students accessed such sources) and used a large number of sources (cluster center: 2.71), indicating verification of information.

Cluster 4 (14.41%, $N = 17$), *engaged* source users, included students who primarily accessed popular press articles (100% of students accessed such sources). Perhaps because the sources these students accessed were engaging, participants in this cluster accessed the most sources (cluster center: 2.82 sources). Cluster 5, the *accessibility* cluster, into which 12.71% of students were classified ($N = 15$), included participants who selected sources of moderate quality that were defined by their accessible and easy to use (100% of students accessed such sources). These students also used a moderate number of sources (cluster center: 1.80 sources). Table 6 includes cluster descriptions and the relative importance of each clustering variable in determining open-ended cluster membership.

Open-ended internal clustering validation

The silhouette measure of cohesion and separation was 0.7 for this five-cluster solution, identified as good. A series of χ^2 tests were run to determine whether the identified clusters differed significantly along each of the binary indicators of source use. Clusters were significantly different in the percentage of their constituent students who accessed reliable sources, $\chi^2(4) = 88.42$,

TABLE 6 Open-Ended Cluster Membership

Cluster	1	4	2	5	3
Label	Disengaged	Engaged Source Users	Non-Critical Source Users	Accessible Source Users	Critical Analytic Source Users
Description	Very limited source engagement	Popular press source used; high number of texts accessed	Unreliable source used; low number of sources accessed	Accessible sources used; moderate number of sources accessed	Reliable source used; high number of sources accessed
Size	47.46% (N = 56)	14.41% (N = 17)	13.56% (N = 16)	12.71% (N = 15)	11.86% (N = 14)
	Inputs (Predictor Importance) Cluster Centroid (Percentage Students at Centroid)				
Popular Press (1.0)	0.00(100%)	1.00(100%)	0.00(100%)	0.00(100%)	0.00(100%)
No Verification (0.8)	0.00(100%)	0.00(94.1%)	1.00(100%)	0.00(100%)	0.00(100%)
Reliable Sources (0.6)	0.00(100%)	0.00(76.5%)	0.00(93.8%)	0.00(100%)	1.00(100%)
Accessibility (0.6)	0.00(100%)	0.00(94.1%)	0.00(56.2%)	1.00(100%)	0.00(100%)
Source Total (0.4)	0.71	2.82	1.44	1.80	2.71

Note: Clusters ordered by size.

$p < 0.001$; popular press sources, $\chi^2(4) = 118.0$, $p < 0.00$; accessible sources, $\chi^2(4) = 86.91$, $p < 0.001$; and the unreliable source without subsequent verification, $\chi^2(4) = 110.37$, $p < 0.001$. Further, an ANOVA determined that clusters differed significantly in the total number of sources that students accessed, $F(4, 113) = 23.22$, $p < 0.001$. Again, these results should be interpreted with caution, as clusters were developed to be most different from one another.

EXTERNAL CLUSTER VALIDATION

Clusters solutions for both the discrete and open-ended questions were externally validated. First, we examined the consistency of cluster solutions across two independent samples: those students responding to the discrete question and those responding to the open-ended question. Further validation was established by examining whether clusters categorizing source use in response to the discrete question corresponded to response accuracy, and whether clusters characterizing source use in response to the open-ended question corresponded to measures of response quality.

Although the discrete question resulted in a six-cluster solution and the open-ended question resulted in only a five-cluster solution, there was a great deal of consistency in the clusters that emerged. Specifically, the five source-use clusters identified when students responded to the open-ended question were also those that emerged from a cluster analysis of students' source use behaviors in response to the discrete question. Indeed, the five identified profiles of source use appeared in two independent samples of students responding to two different types of tasks. In the case of the discrete question, an additional cluster emerged: comprehensive source users. This cluster was not identified in the clustering of students responding to the open-ended question, thus the potential stability of this cluster may require further investigation. However, based on the two independent samples, there is reason for confidence in the five identified clusters that emerged as characterizing students' multiple source use. Table 7 presents the number of students classified into each cluster by question type.

Finally, clusters were validated by examining whether discrete cluster membership was associated with response accuracy and

TABLE 7 Percent Cluster Membership in Response to Discrete versus Open-Ended Questions

	Disengaged	Accessible Source Users	Engaged Source Use	Non-Critical Analytic Source Users	Critical Analytic Source Users	Comprehensive Source Users
Discrete	16.39%	32.79%	8.20%	13.93%	13.93%	14.75%
	N = 20	N = 40	N = 10	N = 17	N = 17	N = 18
Open-Ended	47.46%	12.71%	14.41%	13.56%	11.86%	N/A
	N = 56	N = 15	N = 17	N = 16	N = 14	

whether open-ended cluster membership was associated with response quality. A logistic regression was run to determine whether cluster membership predicted response accuracy in the case of the discrete question, however, the model was not significant; Wald (1) = 0.48, $p = 0.49$, Exp(B) = 1.08, $\beta = 0.79$ ($SE = 0.8$). A one-way ANOVA was used to determine whether established clusters differed in the number of words students included in their open-ended responses, however the ANOVA was not significant, $F(4, 113) = 1.17$, $p = 0.33$.

The Kruskal-Wallis test was run to determine whether open-ended cluster membership was associated with any additional measure of response quality: the number of arguments, elaboration, or citations included in students' open-ended responses or with students' SOLO scores. The Kruskal-Wallis is a non-parametric test, parallel to the one-way ANOVA, to examine the relation between a categorical independent variable (i.e. cluster membership) and data that are at least ordinal, as were our indicators of source use quality. Due to the generally limited quality of students' responses, all of the measures of source quality had a restricted range and were treated as non-continuous in nature.

Students in differing clusters varied in the number of arguments they included in their open-ended responses, $H(4) = 10.77$, $p < 0.05$. Based on an examination of the descriptive statistics, specifically the mean rank of students in each cluster on the number of arguments they produced, students in the uncritical source use cluster ranked lower (mean rank = 36.06) than the disengaged (mean rank = 60.21), critical analytic (mean rank = 67.39), engaged (mean rank = 67.74), and accessible (mean rank = 65.13) clusters. Students' cluster membership did not differentiate their SOLO scores, $H(4) = 5.84$, $p = 0.21$, or the number of elaborations, $H(4) = 8.03$, $p = 0.09$, or citations that students included in their responses, $H(4) = 3.81$.

To check if these differences were significant, four paired comparisons using the Mann-Whitney U test were run. The Mann-Whitney U is a non-parametric test like the Kruskal-Wallis but is used when the independent variable has only two categories (i.e. to compare two clusters). A significance level was set, according to Bonferroni's adjustment for multiple comparisons, at $\alpha = 0.0125$, to maintain a family-wise error rate of $\alpha = 0.05$. Students in the remaining four clusters differed significantly from those students

in the *uncritical source use* cluster, Mann-Whitney U: 52.00–262.00, *p*s: 0.005–0.009.

Discussion and Conclusions

The purpose of this study was: (a) to examine whether particular indicators of MSU differed significantly across question type, (b) to consider these indicators jointly and to identify profiles of students' multiple source, (c) to determine the consistencies of these MSU profiles when students respond to discrete versus open-ended questions, and (d) to identify whether differences in source use cluster membership corresponded to differences in performance.

Research Goal 1: Multiple Source Use Across Question Type

The findings associated with the first study goal appear counterintuitive. In responding to the two questions, it was expected that students would access more sources to answer the open-ended question, as this was the question that required a more elaborated response. However, the opposite effect was found. Students used significantly more sources when answering the discrete question than they did when answering the open-ended question. Further, in responding to the discrete question, significantly more students used reliable sources and accessible sources than did students responding to the open-ended question. Students may have used more sources in responding to the discrete question in an attempt to find "*the answer*," a source use goal that has been found to be pronounced in prior studies of multiple source use (Wallace et al., 2000). Given that students may have perceived the open-ended question as not having an "answer" or a singular response, learners may have been more easily satisfied by any source providing seemingly relevant information. Reader and Payne (2007) have termed students' satisfaction with texts meeting a basic threshold of acceptability as reflecting *satisficing* approach to MSU. What appear to be counterintuitive findings, therefore, become understandable when we surmise that students treated the open-ended question as an occasion to forward their opinion, bypass-

ing the need for extended source use or source-based response justification.

Research Goals 2 and 3: Identify Profiles of Multiple Source Use Across Question Types

In meeting Study Goals 2 and 3, we identified the use of clustering as an approach to analyzing students' multiple source use. Distinct profiles of students' MSU were determined based not only on quantitative measures of source use but also on qualitative aspects of sources accessed. Based on internal and external validations, the five source-use clusters that emerged were consistent for students responding to both the discrete and open-ended questions. Clusters were formed based on how many sources students used but also the kinds of sources that students accessed. This method revealed students to be sensitive to difference between sources and to be able to differentiate texts along a variety of dimensions, including their reliability, accessibility, and engagingness. While prior work has examined source type as a factor differentiating students' source evaluations (e.g., Bråten et al., 2011), in the present study, it served to separate students' source engagement in responding to task demands. In jointly considering a number of source use indicators using cluster analysis, we were able to generate interpretable profiles of students' MSU and to better understand how students' source use differed by question type. Although the behavioral clusters were consistent across question type, they differed in the relative percentages of students adopting each approach to source use when responding to the two different questions.

A substantial portion of our sample was classified into the disengaged cluster of MSU. These were students who did not access any particular types of sources and used a low number of sources overall. On its own, the presence of a disengaged cluster suggests that students may have had limited motivation for study participation or task completion. However, the relative number of students classified into this cluster across the two questions varied vastly (see Table 7). In responding to the discrete question, 16.39% of students were classified as disengaged, however, for the open-ended question, 47.46% of students were classified in this way. This discrepancy in cluster membership was thought to reflect the

higher level of difficulty the open-ended question presented for students. When given a task requiring source corroboration and integration to craft a quality response, rather than engaging in the demanding process of MSU, students may have disengaged.

Yet, these profiles may again support the notion that many students conceive of the open-ended question as an "opinion" question. In effect, students' framing or re-framing of the open-ended question eliciting text integration to one allowing for any self-generated response may explain why so many students in the open-ended task were classified into the disengaged source use profile. Such an outcome may arise because the participants did not grasp the need to substantiate or justify their response with content from one or more sources or because they saw the generation of an unsubstantiated opinion as an expedient way to respond.

This interpretation is further supported by the finding that students accessed more sources and reliable sources when responding to the discrete rather than open-ended question. Given the relative ease of the discrete question, students could engage in the process of MSU when responding without over-complexifying the task. For the open-ended question, the MSU process would have been comparatively much more demanding for students. Rather than engaging with texts, students may have elected to opt-out of source use, to use reliable sources to a lesser extent, and to treat the question as allowing for any response. In terms of the MD-TRACE model (Rouet, 2006), such source use patterns may have been manifestations of differences in task models construction. Such variation in task model construction may have arisen from students' understanding and reinterpretations of task demands.

For those students who did engage in the source use process, there is evidence both to support source use profiles previously identified in the literature (i.e. sampling, satisficing; Reader & Payne, 2007) and to introduce additional profiles of MSU. The non-critical source users are a concerning profile of students. These are students who, upon finding a source low in reliability, did not seek out additional sources to verify or corroborate the information in the untrustworthy sources. This type of approach to source use is a challenge for researchers and teachers inter-

ested in promoting students' critical source use and evaluation (Braasch et al., 2013; Britt & Aglinskas, 2002). Students in this cluster seemed to be demonstrating the satisficing strategy described by Reader and Payne (2007). These students elected to simply use the first source they accessed, as it provided relevant, though unreliable, information, rather than expending the effort to identify additional, better sources. In contrast, the critical analytic cluster seemed to include students who engaged in the "best" source use process; however only a limited number of students populated this cluster. These were students who selected highly reliable sources and accessed a number of texts, potentially in order to verify and corroborate information. These source users seem to be demonstrating the sampling strategy identified by Reader and Payne (2007), accessing a number of sources to evaluate them and then selecting sources of high quality. As in the present study, such high-level source users were not frequently identified in prior research (Wilkinson, Reader, & Payne, 2012).

The task-specificity of students' approaches to source use was demonstrated in the present study to a limited extent. Although similar clusters were produced in response to both the discrete and open-ended questions, responding to the discrete question seemed to encourage students to select accessible sources. The accessibility profile was substantially more populated when students were responding to the discrete question (32.79%), rather than the open-ended question (12.71%). As the accessible sources were characterized as providing clear and easy access to concrete pieces of information, as students needed to do when responding to the discrete question, the accessibility profile may indeed have been most adaptive for this question.

The cluster that uniquely emerged when students responded to the discrete question, comprehensive searchers, was the only profile that seemed to favor a mixed approach to text access. Students in this cluster accessed not only highly reliable sources but accessible sources as well. These were learners who seemed motivated to both use quality sources and to respond to tasks efficiently (e.g., access easy to use sources). Alternatively, given the large number of sources accessed, classification into the comprehensive source use profile may indicate that these students may have been challenged in finding appropriate sources to satisfy task demands. More research is needed to examine differences in MSU

behaviors in response to differing text demands, as the majority of studies examining students' source navigation have focused on students responding to discrete questions (i.e. those having a single correct answer; Cerdán & Vidal-Abarca, 2008; Cerdán et al., 2009; Reader & Payne, 2007; Wilkinson et al., 2012) rather than more open-ended tasks.

Research Goal 4: Multiple Source Use Clusters and Task Performance

With regard to the fourth research goal, to examine the relation between cluster membership and task performance, cluster membership in response to the open-ended question was related to the number of arguments that students included in their responses. However, cluster membership was not significantly related to word count, the number of elaborations or citations included, or students' SOLO scores. Similarly, for the discrete question, there was no significant relation between cluster membership and whether students responded correctly. The limited relation between cluster membership and students' task performance may have stemmed from measurement limitations. The majority of students responded to the discrete question correctly, making a logistic regression model that classifies all students into the largest group (i.e. correct response) difficult to improve upon. As evidenced by the large percentage of students responding to the discrete question correctly, locating isolated pieces of information in text may be a fairly habituated task for students. Answering discrete questions based on multiple texts may be automatic to such an extent, or the task may afford such ease, that it may be difficult to identify differential results in product, even when the process of source use is quite dissimilar.

There was a large discrepancy in performance between the easier discrete question, which most students answered correctly, and students' responses to the open-ended question, which included only a limited number of arguments with minimal elaboration or integration. In the literature, composing integrated responses based on multiple texts has been identified as a challenge for students (e.g., List & Alexander, 2015; Britt & Sommer, 2004). Indeed, for students in the present study, their low level of performance may reflect a motivational issue or may be an issue of students' feeling challenged by translating the information

in the sources they access into quality responses. Moreover, our measures of MSU quality had limited variance. As a result, the relation between source use cluster membership and indices of response quality may have been attenuated. Nonetheless, membership in the uncritical source use cluster was associated with students providing significantly fewer arguments in their open-ended responses than with students classified into the other clusters. This finding provides some evidence of the relation between source use cluster membership and performance on an academic task. More work is needed to develop tasks that provide greater score variability. Further, there is a need to determine the baseline level of skills needed in order for open-ended questions to not pose an undue challenge for students.

Implications, Limitations, and Future Directions

This study offers at least three novel contributions to the literature. First, it employs clustering, a novel statistical method, to make sense of log data of multiple source use. Second, it evidences the presence of variability and discernable patterns in students' multiple source use. Whereas the MD-TRACE model posits Step 3, Source Processing, unfolding as three steps, the current investigation suggests variation in how these steps are enacted by learners. Third, it provides insights into how task demands are reinterpreted by students to frame their MSU process, particularly in cases where task demands may appear to be demanding.

Yet, many questions about the nature of the emerging clusters persist and point the way toward further investigations. First, there is a question of the stability of source use clusters: Do students engage in MSU in consistent ways across task and domains, or do students elect to engage in varied approaches to source use depending on task demands? The extent to which students are cognizant of their source use profile and purposeful in their source use is a question for further research. Salmerón, Kintsch, and Cañas (2007) have raised similar questions in profiling students' approaches to source selection. In addition to examining source use clusters in relation to students' task performance, the relation between learner characteristics, such as prior knowledge and epistemic beliefs, and students' pattern of source use has yet to be examined. Insights from such investigations may help to

understand the extent to which students' source use processes changes as students develop and advance within a domain.

There are certain limitations and delimitations associated with this study that further give rise to avenues for future research. First, the present study did not consider the relation among individual characteristics and MSU. As the focus of the present study was to identify patterns of source use across types of questions, individual differences beyond students' source use were not specifically addressed. As models of MSU indicate that variables such as knowledge and motivation have the potential to play a role in students' selection and use of sources (Rouet, 2006), future research should consider the extent that these factors influence source use. Moreover, building on the present study, identifying whether profiles can be distinguished by factors such as knowledge and motivation is a fruitful avenue for future research. Further, the inclusion of time data would certainly enrich our understandings of MSU and provide an additional indicator of students' source use process and engagement in the task. Introducing time spent on sources has the potential to further explicate, and perhaps differentiate, different types of source users.

In the present study, clustering was demonstrated to be an effective mechanism for examining students' MSU and consistent profiles of students' source engagement. Although cluster analysis provides interesting and important insights into profiles of students, it has certain limitations as a methodological approach. Cluster analysis is an exploratory technique, and while aligned with the goals present study, future research seeking to confirm the number and type of profiles may turn to other analytic techniques. As cluster analysis is dependent on decisions such as the choice of clustering variables, other clustering variables may be included in future analyses, such as variables related to students' source re-visits and the sources students first accessed. Further, while a two-step clustering method was selected to support the combination of discrete and continuous indicators, other procedures exist and may provide different results (Milligan, 1996). Given the insights clustering can provide to support a person-centered understanding of MSU, it is important to replicate clusters across other samples and to pursue additional qualitative data, such as interviews, may further clarify students' process of MSU.

Ultimately, while clustering was demonstrated to be an effective mechanism for examining students' MSU and consistent profiles of students' source engagement were identified, a number of questions remain for future research. One question is the extent to which similar clusters manifest when additional task variations are introduced (e.g., varied domain) or when texts are varied along dimensions besides source type and reliability (e.g., reading difficulty). In addition, students' navigation path, or order of source access, not only the sources that were selected, is an area for further investigation and may provide rich information to aid in further identifying patterns in students' source use. In terms of further validating clusters, it may be interesting to examine the extent to which these identified source use profiles differ in their source evaluations. Such person-centered analyses may also be examined by considering whether students' prior knowledge or interest predict or differentiate cluster membership. Given the prominence of MSU in the lives of today's learners, more work is needed to understand which approaches to source use may be more adaptive for students and how these may be fostered to help learners in today's digital age.

References

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research, 49*(2), 280–318.

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, 41*(5), 1–22.

Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS two-step cluster: A first evaluation.* Nürnberg, Germany: Lehrstuhl Soziol. Retrieved from: http://www.statisticalinnovations.com/products/twostep.pdf

Biggs, J. B., & Collins, K. F. (1982). *Evaluating the quality of learning: The Solo Taxonomy: Structure of the Observed Learning Outcome.* New York: Academic Press.

Braasch, J. L., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology, 38*(3), 180–195.

Braasch, J. L., Lawless, K. A., Goldman, S. R., Manning, F. H., Gomez, K. W., & MacLeod, S. M. (2009). Evaluating search results: An empirical analysis of middle school students' use of source attributes to select useful sources. *Journal of Educational Computing Research, 41*(1), 63–82.

Brand-Gruwel, S., & Stadtler, M. (2011). Solving information-based problems: Evaluating sources and information. *Learning and Instruction*, *21*(2), 175–179. doi:10.1016/j.learninstruc.2010.02.008

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508.

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*(1), 48–70.

Bråten, I., & Strømsø, H. I. (2003). A longitudinal think-aloud study of spontaneous strategic processing during the reading of multiple expository texts. *Reading and Writing*, *16*(3), 195–218.

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

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

Britt, M. A., Rouet, J. F., & Braasch, J. L. G. (2013). Documents as entities: Extending the situation model theory of comprehension. In M. A. Britt, S. R. Goldman, & J. F. Rouet (Eds.), *Reading: From words to multiple texts* (pp. 160–180). New York: Routledge, Taylor & Francis Group.

Britt, M. A., Rouet, J. F., & Perfetti, C. A. (1996). Using hypertext to study and reason about historical evidence. In J. F. Rouet, J. J. Levonen, A. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition* (pp. 43–72). Hillsdale, NJ: Lawrence Erlbaum Associates.

Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macrostructure focusing tasks. *Reading Psychology*, *25*(4), 313–339.

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*(1), 209–222.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13–27.

Coiro, J. (2003). Exploring literacy on the Internet: reading comprehension on the Internet: expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher*, *56*(5), 458–464.

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, *42*(2), 214–257.

Coiro, J., & Kennedy, C. (2011). *The Online Reading Comprehension Assessment (ORCA) project: Preparing students for Common Core Standards and 21st century literacies*. Retrieved from http://www.orca.uconn.edu/orca/assets/File/Research%20Reports/CCSS%20ORCA%20Alignment%20June%202011.pdf

Common Core State Standards Initiative. (2010). *Common Core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Council of Chief State School Officers and the National Governors Association Center for Best Practices.

Dreher, M. J., & Guthrie, J. T. (1990). Cognitive processes in textbook chapter search tasks. *Reading Research Quarterly, 25*(4), 323–339.

Eveland, W. P., Jr., & Dunwoody, S. (2000). Examining information processing on the World Wide Web using think aloud protocols. *Media Psychology, 2*(3), 219–244.

Ferguson, L. E., & Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs: Changes and relations to multiple-text comprehension. *Learning and Instruction, 25*, 49–61.

Flesch, R. (1979). *How to write in plain English: A book for lawyers and consumers.* New York: Harper.

Ford, N., & Chen, S. Y. (2000). Individual differences, hypermedia navigation, and learning: An empirical study. *Journal of Educational Multimedia and Hypermedia, 9*(4), 281–311.

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.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology, 31*(1), 30–68. doi:10.1080/02702710902733600

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

Goldman, S. R., & Scardamalia, M. (2013). Preface for the special issue multiple document comprehension. *Cognition and Instruction, 31*(2), 121–121.

Graesser, A., Rus, V., & Cai, Z. (2008). Question classification schemes. Proceedings from WS on the QGSTEC: *1st Workshop on Intelligent Tutoring Systems*. Retrieved from http://qg.cs.memphis.edu/papers/QG2008/PDFs/16-GraesserEtAl-QG08.pdf

Guthrie, J. T. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly, 23*(2), 178–199.

Hofman, R., & Van Oostendorp, H. (1999). Cognitive effects of a structural overview in a hypertext. *British Journal of Educational Technology, 30*(2), 129–140.

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: an eye-tracking study. *Behavior and Information Technology, 31*(1), 83–97.

Kincaid, J. P., Fishburne, R., Rogers, R. L, & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy enlisted personnel* (Branch Report 8-75). Millington, TN: Chief of Naval Training.

Le Bigot, L., & Rouet, J. F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39*(4), 445–470.

List, A., & Alexander, P. A. (2015). Examining response confidence in multiple text tasks. *Metacognition and Learning, 10*(3) 407–436.

Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science, 38*(6), 607–633.

McCrudden, M. T., & Schraw, G. (2009). The effects of relevance instructions and verbal ability on text processing. *The Journal of Experimental Education, 78*(1), 96–117.

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*(3), 271–290.

Milligan, G. W. (1996). Clustering validation: Results and implications for applied analyses. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and classification* (pp. 341–376). River Edge, NJ: World Scientific Publishing.

Milligan, G. W., & Hirtle, S. C. (2013). Clustering and classification methods. In I. B. Weiner, J. A. Schinka, & W. F. Vellicer (Eds.), *Handbook of psychology, Volume 2: Research methods in psychology* (pp. 189–210). Hoboken, NJ: John Wiley & Sons.

Mooi, E. A., & Sarstedt, M. (2011). Cluster analysis. In E. A. Mooi & M. Sarstedt (Eds.), *A concise guide to market research: The process, data, and methods using IBM SPSS statistics* (pp. 237–84). New York: Springer-Verlag.

Norušis, M. J. (2005). *SPSS 13.0 guide to data analysis*. Englewood Cliffs, NJ: Prentice Hall.

Pastor, D. A. (2010). Cluster analysis. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 41–54). New York: Taylor & Francis.

Payne, S. J., & Reader, W. R. (2006). Constructing structure maps of multiple on-line texts. *International Journal of Human-Computer Studies, 64*(5), 461–474.

Pieschl, S., Stahl, E., & Bromme, R. (2008). Epistemological beliefs and self-regulated learning with hypertext. *Metacognition and Learning, 3*(1), 17–37.

Purcell, K., Rainie, L., Heaps, A., Buchanan, J., Friedrich, L., Jacklin, A.,. . .Zickuhr, K. (2012). How teens do research in the digital world. *Pew Internet & American Life Project*. Retrieved from http://www.pewinternet.org/∼/media//Files/Reports/2012/PIP_TeacherSurveyReportWithMethodology110112.pdf

Raine, L., & Purcell, K. (2010). The economics of online news. *Pew Internet & American Life Project*. Retrieved from http://pewinternet.org/Reports/2010/5–The-economics-of-online-news/Media.aspx

Reader, W. R., & Payne, S. J. (2007). Allocating time across multiple texts: Sampling and satisficing. *Human Computer Interaction, 22*(3), 263–298.

Rokach, L., & Maimon, O. (2005) Clustering methods. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 321–352). New York: Springer.

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology, 58*(1), 56–61.

Rouet, J. F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Interacting with Computers, 15*(3), 409–428.

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

Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*(3), 478–493.

Rouet, J. F., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction, 21*(2), 205–219.

Rouet, J. F., & Vidal-Abarca, E. (2002). Mining for meaning: Cognitive effects of inserted questions in learning from scientific text. *The psychology of science text comprehension*, (pp.417–436). Mahwah, NJ: Lawrence Erlbaum Associates.

Rouet, J.-F., Vidal-Abarca, E., Erboul, A. B., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes, 31*(2), 163–186. doi:10.1207/S15326950DP3102_03

Salmerón, L., & Kammerer, Y. (2012). Selecting pages from Google to learn about a controversial topic: The role of epistemic beliefs. In de Vries, E. & Scheiter, K. (Eds.), *Proceedings from EARLI Sig 2: Staging knowledge and experience: How to take advantage of representational technologies in education and training?* (pp. 181–183). Grenoble, France: EARLI.

Salmerón, L., Kintsch, W., & Cañas, J. J. (2006). Coherence or interest as basis for improving hypertext comprehension. *Information Design Journal, 14*(1), 45–55.

Stahl, S., Hynd, C., Britton, B., McNish, M., & Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly, 31*(4), 430–456.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20*(3), 192–204.

Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction, 31*, 176–203.

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology, 102*(4), 817–826.

Vidal-Abarca, E., Martinez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., ... & Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior Research Methods, 43*(1), 179–192.

Vidal-Abarca, E., Mengual, V., Sanjose, V., & Rouet, J. F. (1996, September). *Levels of comprehension of scientific prose: The role of text and task variables*. Paper presented at the International Seminar on Using Complex Information Systems, Poitiers, France.

Vidal-Abarca, E., & Sanjose, V. (1998). Levels of comprehension of scientific prose: The role of text variables. *Learning and Instruction*, *8*(3), 215–233.

Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research*, *27*(3), 255–265.

Wallace, R. M. C., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the Web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences*, *9*(1), 75–104.

Wang, Y. M., & Artero, M. (2005). Caught in the Web: University student use of Web resources. *Educational Media International*, *42*(1), 71–82.

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

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

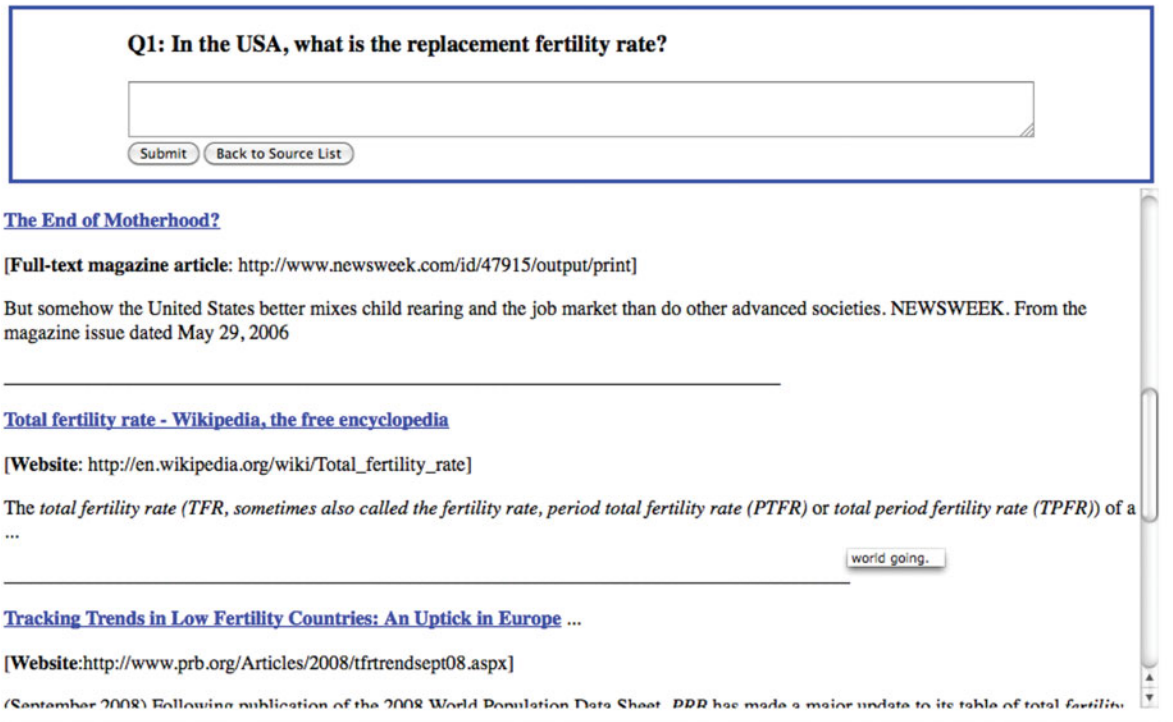
Wilkinson, S., & Payne, S. (2006, September). Eye tracking to identify strategies used by readers seeking information from on-line texts. *Proceedings of the 13th European conference on cognitive ergonomics: Trust and control in complex socio-technical systems* (pp. 115–116). New York: Association for Computing Machinery.

Wilkinson, S. C., Reader, W., & Payne, S. J. (2012). Adaptive browsing: Sensitivity to time pressure and task difficulty. *International Journal of Human-Computer Studies*, *70*(1), 14–25.

Wixson, K. K. (1983). Postreading question–answer interactions and children's learning from text. *Journal of Educational Psychology*, *75*(3), 413–423.

Appendix A

Screen Shot of the Search Task and Source Library Interface



Appendix B

Scoring Rubric for the Open-Ended Question

How is government supported childcare important as a factor in high fertility rates?

0 = irrelevant or inappropriate information/evidence; or "I don't know"
1 = provide a single piece of relevant information/evidence
2 = provides multiple pieces of relevant information/evidence, but does not link them together, may come to different conclusions with the same data, may be inconsistent
3 = provides multiple pieces of relevant information/evidence that are linked together, and conclusions are reasonable given the information in the answer; however, additional information may make conclusions questionable
4 = provides multiple pieces of relevant information/evidence that are linked together; conclusions are reasonable given the information in their answer and hold even when additional information is added; may generalize beyond the text

Students were initially classified according to the type of answer they were attempting to provide according to the rubric above. Then if they were unsuccessful in their undertaking, a half point was deducted. For example, if students attempted to provide multiple reasons for how government sponsored child-care might impact high fertility rates, they were initially classified as a 2. However if one of their reasons was incorrect, unclear or not directly related to high fertility rates they were ultimately scored as a 1.5.