WILEY InterScience® DISCOVER SOMETHING GREAT

Task-Based Information Retrieval: Structuring Undergraduate History Essays for Better Course Evaluation Using Essay-Type Visualizations

John E. Leide, Charles Cole, Jamshid Beheshti, Andrew Large, and Yang Lin
Graduate School of Library and Information Studies, McGill University, 3459 McTavish Street, Montreal, Quebec, Canada H3A 1Y1. E-mail: {john.leide, charles.cole, jamshid.beheshti, andrew.large}@mcgill.ca, yang.lin1@mail.mcgill.ca

When domain novices are in C.C. Kuhlthau's (1993) Stage 3, the exploration stage of researching an assignment, they often do not know their information need; this causes them to go back to Stage 2, the topic-selection stage, when they are selecting keywords to formulate their query to an Information Retrieval (IR) system. Our hypothesis is that instead of going backward, they should be going forward toward a goal state—the performance of the task for which they are seeking the information. If they can somehow construct their goal state into a query, this forward-looking query better operationalizes their information need than does a topic-based query. For domain novice undergraduates seeking information for a course essay, we define their task as selecting a high-impact essay structure which will put the students' learning on display for the course instructor who will evaluate the essay. We report a study of first-year history undergraduate students which tested the use and effectiveness of "essay type" as a task-focused query-formulation device. We randomly assigned 78 history undergraduates to an intervention group and a control group. The dependent variable was essay quality, based on (a) an evaluation of the student's essay by a research team member, and (b) the marks given to the student's essay by the course instructor. We found that conscious or formal consideration of essay type is inconclusive as a basis of a task-focused query-formulation device for IR.

Introduction

Research in library and information science (LIS) often compares the domain novice, who has little or no subject knowledge or background, with the domain expert, who has extensive knowledge of the topic area in which a search will be affected. One such example of a comparison is between the undergraduate student client and the subject specialists in the university library who prepare subject guides for these students. Reeb and Gibbons (2004) found that there is an apparent mismatch between the mental models of undergraduate students and the subject specialists concerning information organization within academic disciplines and the role of library subject guides. The domain novices "undergraduates … focused on courses and coursework" (p. 126) while the domain expert subject specialist who writes the subject guide focuses on the subject discipline. Reeb and Gibbons concluded that the undergraduates' mental models were "not well suited to library subject guides" (p. 126). A result of the mismatch between domain expert and domain novice is that after referring to the subject guide provided by the library, when "undergraduate students turn to the scholarly electronic resources licensed by libraries their search skills are poor" (Troll, 2002, p. 115), leading to a lasting negative impression, especially for those students receiving average or lower than average grades.

The design of subject discipline or document-oriented information retrieval (IR) systems created by domain expert subject specialists is beginning to be influenced by user-centered research that takes into account the way domain novices think when they search IR systems. User-centered research focuses on information seeking, which is a broader picture view of the user and the IR system than information search (Wilson, 1997, 1999). Primary points of interest in information-seeking research are the user's stages of information seeking (Kuhlthau, 1993) and how different groups of users, mainly occupation groups (Case, 2002), seek information differently (Ellis, Cox, & Hall, 1993; Leckie, Pettigrew, & Sylvain, 1996). In the university library, social science and particularly history users have been studied as separate, distinct groups of information seekers (Cole, 1994; Line, 1971; Stieg, 1981). These user-centered studies can be cognitive oriented (e.g., Belkin, Oddy, & Brooks, 1982a, 1982b; Ingwersen, 1996; Saracevic, 1996a, 1996b), sociologically oriented (Fisher, Durrance, & Hinton, 2004; Rieh, 2004; Savolainen, 1995; Spink & Cole, 2001), or both.

Received January 10, 2006; revised September 14, 2006; accepted September 14, 2006

© 2007 Wiley Periodicals, Inc. ● Published online 4 May 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20598

The cognitive and sociological processes involved in user–IR system interaction are linked in the concept of the user's task when seeking and searching for information. User tasks and the effect task has on behavior when using an IR system have received less attention as foci of modeling and theorizing about user–system interaction (Järvelin & Ingwersen, 2004; Vakkari, 2003); however, the influence of the user's task when accessing information from an IR system can be an important variable when designing effective IR systems. Beyond the searching mechanics of making an electronic IR system work for the user, the user's task forms an overlay on the basic process of accessing information using an electronic IR system, coloring and informing the entire information-search activity.

This study forms part of a series of studies that is focused on formulating design concepts for a new IR system specifically designed for undergraduates researching a social science essay (Cole & Leide, 2003; Cole, Leide, Beheshti, Large, & Brooks, 2005; Cole, Leide, Large, Beheshti, & Brooks, 2005; Leide, Large, Beheshti, Brooks, & Cole, 2003; Yi, Beheshti, Cole, Leide, & Large, 2006). In the present study, we focus on university undergraduates in a first-year history course who are in Kuhlthau's (1993) Stage 3, the exploration stage of her six-stage essay-writing process. These first-year university students are just beginning their study of history and are therefore domain novices in the subject area. Furthermore, as first-year history students, they also are novices in the task of researching and writing college-level term papers. Since the students are learning both their task and the subject of history, if one can explicitly articulate important elements of task performance for the students while they are accessing information for that task, it will facilitate an important improvement in educational quality generally and these users' perceptions of the usefulness of IR systems in particular.

The General Problem

Current IR systems are designed either for known-item searching tasks or for exploratory subject searches for domain expert users of these systems who can transfer searching skills and knowledge about the structure of subject disciplines to the their search (Drabenstott, 2003; Stoan, 1984). Domain novice undergraduates researching their social science essay who use an IR system such as the university library's online public-access catalogue, a scholarly database, an online index, or an Internet search engine such as Yahoo! or Google are often conducting exploratory subject searches in a topic domain that is new to them without any knowledge or subject-searching skills. This is the specific situation undergraduates find themselves in at the beginning of the university term. They do not know the subject domain in which they are searching, and therefore do not know the information they will need to perform their task of researching and writing a social science essay. This makes it particularly difficult for these students to select search terms to put in their query to the IR system.

Kuhlthau's (1993) Information Search Process (ISP) model divides the task of researching a course assignment into six stages. The student receives the essay assignment (Stage 1) from the course instructor and then selects a general topic (Stage 2). Stage 3 indicates an exploration of the topic via an exploration of information sources. During this exploration stage, students often select topic terms to form their query to IR systems. These topic elements have been found not to be an effective mine for articulating search terms to find appropriate information; that is, topic terms do not effectively represent this type of user's information need (Belkin et al., 1982a, 1982b). New technology incorporating information organization, collaboration, and management features into the activity of accessing information has the potential to correct this crucial problem encountered by domain novice users of IR systems by joining together information seeking, search, and use into the activity of query formulation centered on the user's task (Bartlett, 2004).

The general problem we address in this article is how to incorporate task-appropriate query-formulation facilitation devices in the design of IR systems for domain novice undergraduates attempting to retrieve information for their social science (i.e., history) essays. After reviewing the literature on task and IR, starting with LIS antecedent terms "Problem" and "Problem Situation," we attempt to determine the nature of the undergraduate's task when writing a history essay. We report findings from both a pilot study and the main study. The pilot study tested various elements of the essay writing task against one another, using essay quality as the performance or dependent variable. The element "essay type" was found to have the greatest positive effect on task performance. We then describe the main study, which tested two different essay types for 78 undergraduates in a first-year history course who were in Stage 3 of writing their essay tests against a control group, again with essay quality as the dependent variable. We conclude the article with a discussion about the future role of task in the design of interactive IR devices.

Literature Review

Task: Antecedents: Problem and Problem Situation

Task in IR is a relatively new term which combines aspects of the antecedent terms "problem" from information search and "problem situation" from information seeking (Wersig, 1973). The user's "problem" labels the initiating conditions of the user's information search, focusing on user–IR system interaction, while "problem situation" labels the wider initiating conditions regarding the purposive information-related behavior of the user. Marchionini (1995), who gave an example of an information search-as-process model, defined the first two steps of information search as (a) recognizing/accepting the information problem and (b) defining the problem. On the other hand, Kuhlthau's (1993) six-stage ISP model, which is an information-seeking model (despite its name), is concerned with the wider picture of all purposive information-seeking behaviors (both user–IR system interaction and other information behavior) from start to finish for a person researching an assignment

DOI: 10.1002/asi

[Kuhlthau's (1993) ISP model predates Wilson's (1997, 1999) definitional distinction between information seeking and information search.]

As a result, there is a definitional problem between what task means in the context of information search and what it means in the context of information seeking. To illustrate the problem, we list four common LIS definitions of task:

- the user's search task of making the IR system work when accessing relevant information using an IR system. These include information-literacy issues such as being able to select the most appropriate database in which to effectuate the search, selecting the most appropriate search strategy and tactics to effectuate the search (Bates, 1979, 1990; Fidel, 1991a, 1991b, 1991c; Harter, 1986; Wildmuth, 2004), and the user's search task of formulating an effective query to the IR system (Belkin et al., 1982a);
- the user's search task as open or closed. A known-item search task (Kim, 2001) is sometimes called a "closed search task" (Fujihara & Miura, 2003), versus a subject-search task or an unknown-item search task (Kim, 2001), which is sometimes called an "open search task" (Fujihara & Miura, 2003);
- the complexity of the user's search task. In information-seeking studies, a subject search is considered a complex search task while a known-item search is considered a simple search task (Vakkari, 1999, 2003). In a variation of this, simple tasks are defined as routine and can be performed with minimal new information while complex tasks are "ill structured" and may need more than one IR session (Vakkari & Hakala, 2000);
- and the user's information-seeking task, operating inside a problem situation that includes organizational and other issues, of which the user's information-related behavior is but a part—for example, writing an undergraduate essay, developing a proposal, or preparing a legal case to be presented before a judge and jury (e.g., Cole & Kuhlthau, 2000).

The first and third definitions are information-search definitions of task describing the task characteristics from the point of view of the effect it will have on the user's search while the last definition is a broader, information-seeking definition concerned with task as a sociological phenomenon.

The migration of the last definition information-seeking concept of user task into the narrower perspective of information search is illustrated in Belkin et al.'s (1982a, 1982b) classic study testing the anomalous states of knowledge (ASK) hypothesis as a basis of interactive IR system design. The ASK hypothesis was innovative because it was designed to solve the user's problem during the user's interaction with the IR system by identifying the user's problem type as being one of five classes and then linking it to an effective information-retrieval strategy devised by the IR system. Originally, the connectivity classes that made up the ASK classification scheme were meant to be algorithmically generated (Belkin et al., 1982b, p. 152); however, when Oddy et al. (1992) revisited ASK 10 years later, they observed that the underlying structures of the user's problem statement and the abstract of the documents created by professional indexers did not in fact match all that well. One of the reasons was that in writing a problem statement, users concentrated too much on topic aspects of the information need, thus ignoring other issues in the larger problem situation of the user, which probably had an effect on the ASK.

In the mid-1990s, hybrid models of information seeking and search were developed that attempted to bring information seeking's wider view of the user's problem situation into the user–system search interaction [e.g., see Saracevic's (1996a) Stratified Model of user–system interaction and Belkin, Cool, Stein, & Thiel's (1995) Information Seeking Strategies Model of Interaction; for further discussion of these two models, see Cole, Beheshti, Leide, & Large, 2005; Spink & Cole, 2005].

The Information-Seeking Notion of "Task" in IR Research

The information-seeking notion of task is only gradually making its way into IR system research and design. A primary motivation for incorporating its larger sociological and cognitive concerns is dissatisfaction with the traditional, relevance-based evaluation measure of IR system search performance. Relevance, defined as the calculation of the equivalence of the match between the user's query and the system output in response to the query (Harter, 1986, 1992, 1996; Hildreth, 2001; Jansen & McNeese, 2005), works properly only within narrowly defined information-search parameters under controlled experimental conditions (Ellis, 1996). It assumes, for example, a single, static user information need throughout a single information-search session, an observation that has been shown to be false in information-seeking and cognitive-oriented search research that looks at the broad picture of the individual and information (Belkin & Marchetti, 1990; Cole, Kennedy, & Carter, 1996; Hearst, 1999; Järvelin & Ingwersen, 2004; Kennedy, Cole, & Carter, 1999; O'Day & Jeffries, 1993; Schamber, Eisenberg, & Nilan, 1990; Spink & Cole, 2005; Vakkari & Hakala, 2000).

In a review for the *Annual Review of Information Science and Technology*, Vakkari (2003) indicated that there have been few LIS studies linking information search (i.e., the user–IR system interaction) to the broader, information-seeking notion of task, forcing IR researchers such as Bartlett (2004) and Crystal and Ellington (2004) to go elsewhere for theory and models of task (to, e.g., Annett, Duncan, Stammers, & Gray, 1971; Benyon, 1992; Card, Moran, & Newell, 1983; Kirwan & Ainsworth, 1992; Leplat, 1989; Taylor, 1911; Tushman, 1978). LIS research on task-based information retrieval can be grouped into six categories:

> *Task Definition*. According to Vakkari (1999), "Tasks are given or identified by the actor. Each task has a recognizable beginning and end, the former containing recognizable stimuli and guidelines concerning goals or measures to be taken (Byström & Järvelin, 1995; Hackman, 1969)" (p. 824). There is some evidence that for undergraduates, their definition of the task affects how they go about seeking information in support of that task (Entwistle & Entwistle, 2003).

DOI: 10.1002/asi

Task Complexity. Task complexity or difficulty is one of the most essential factors affecting task performance as observed in psychological experiments (e.g., Locke, Shaw, Saari, & Latham, 1981; Wood, Mento, & Locke, 1987), organizational studies (e.g., March & Simon, 1967; Van de Ven & Ferry, 1980), and information-seeking studies (e.g., Byström & Järvelin, 1995; Culnan, 1983; Hart & Rice, 1991; Tiamiyu, 1992; Vakkari, 1999, 2003).

Task Knowledge Level. The user's level of domain knowledge before accessing information determines search behavior. Task (domain) novice users need more information than do task (domain) experts, or even users who performed the task a few times before (Belkin, Seeger, & Wersig, 1983; Harter & Hert, 1997; Ingwersen, 1996; Patel & Ramoni, 1997; Sutton, 1994; Vakkari, 1999).

Task Performance in Stages. Kuhlthau's (1993, 1999, 2004) ISP model "is a process model, which connects source selection and the need and use of information to actors' changing understanding of their task. It conceptualizes information seeking as embedded in task performance and in information use" (Vakkari & Järvelin, 2005, p. 125). Vakkari and Hakala (2000) charted changes during problem stages from the beginning to end of task performance.

Task Evaluation: Relevance assessment and problem stage. "Relevance is understood as a task- and process-oriented user construct," which means that the user's relevance assessment criteria vary due to his or her problem stage during task performance (Vakkari & Hakala, 2000, p. 540; cf. Ruthven, 2005; Vakkari, 2000b). Vakkari and Hakala (2000) charted changes in relevance criteria during task performance.

Task Performance as an IR System Performance Indicator. An outcome measure tied to the information-seeking notion of task performance should be used along with the concept of relevance and the relevance-associated measures of precision and recall to evaluate IR system performance (Hersh, Pentecost, & Hickam, 1996). According to Reid (2000), "the effectiveness of an IR system should be measured in terms of its ability to retrieve 'task-relevant' documents, where a task-relevant document may be defined as one which contributes in some way to the successful completion of the task in hand" (p. 115) (cf. Järvelin & Ingwersen, 2004; Markkula, Tico, Sepponen, Nirkkonen, & Sormunen, 2001). For medical IR system performance, Hersh et al. (1996) proposed outcomes-based methods to assess the impact of the system upon the user for particular real-life tasks (e.g., for answering a question or solving a problem) (cf. Vakkari & Järvelin, 2005).

This article reports findings from a study that explores a specific application of task-based IR research. To begin our discussion of variables used in the study, we make the following observations relating the six categories of task-based LIS research from the aforementioned list to the group and task that we are studying:

- The definition of the group and task studied is undergraduates researching a history essay. The task of essay writing has a definite beginning, when the undergraduate is handed the course requirements the first day of the course, and a definite end, when the student is required to hand in the essay to the course instructor.
- Essay researching and writing is a complex task which requires the student to combine information from many sources including from the course instructor's lectures into a coherent information object which demonstrates to the course instructor a mastery of such skills as organization, critical thinking, and writing style.
- These first-year university students are just beginning their study of history and are therefore at the domain novice knowledge level. Furthermore, as first-year history students, they also are novices in the task of researching and writing college-level term papers.
- We interview the undergraduates when they are in Stage 3, the exploration stage of Kuhlthau's (1993) six-stage ISP model.
- Relevance criteria should be geared to the student's stage of researching the essay, which constitutes the parameters of the evaluation variable for IR system performance. According to Kuhlthau (1993), the task of the information seeker in Stage 3 of his or her ISP is to explore information to gain a personal understanding of the issues and problems in the subject area (p. 35).
- The broader, information-seeking notion of task performance is a second IR system performance evaluation variable. As the task of the undergraduate is to write a history essay, performance evaluation for the IR system should be tied to the quality of the essay the student constructs from the collected information.

Task: Undergraduate History Essay Writing

What are the essential elements of the task of undergraduate history essay writing? History essay writing guidebooks outline a stage approach to essay construction. Robertson (1995), for example, divided the process for writing a research essay into the two components of research and presentation. Under the aegis of research, Robertson listed the following necessary steps: 1. selecting a topic, 2. narrowing the focus, 3. searching for sources, 4. defining the purpose, 5. preparatory reading, and 6. recording of information and ideas. Presentation is the writing up of the essay.

To a great extent, Robertson's steps correspond to the Kuhlthau's (1993) ISP model. Both have topic-selection and focus-narrowing stages (for Robertson, Stage 1 and Stage 2, respectively; for Kuhlthau, topic selection and focusing are Stage 2 and Stage 4 of her six-stage ISP model). The Kuhlthau (1993) model, however, gives a more sophisticated view of the role of information seeking in moving the student through the stages, with each stage having its own distinct information-seeking modality (Cole, Cantero, & Sauve, 1998).

In Hellstern, Scott, and Garrison (1998), the task of writing an undergraduate history essay is broken down into 12 steps, with information seeking understood as a necessary component of each step:

1. Planning. During this stage, topic, purpose, and audience are the three elements that must be decided.

DOI: 10.1002/asi

2. Selecting a topic. Any assigned subject may lead the students to explore a range of topics. What they should look for, however, is a topic that really engages their own interests, which eventually can be transformed on paper.
3. Narrowing a topic.
4. Finding a thesis. Establish the assertion or argument, which will reveal the orientation that the paper will attempt to prove.
5. Defining a purpose. It is a process to establish the writing purpose or goal of the paper.
6. Defining audience. Define the notion and attributes of the intended audience.
7. Organizing the writing. Determine what kind of approach to the subject might best capture the attention.
8. Outlining. Lay out the logical structure of the essay.
9. Drafting.
10. Revising.
11. Editing.
12. Proofreading/Proofing.

The Hellstern et al. (1998) list brings in elements of essay writing that are not included in the Robertson (1995) list: a consideration of the audience for which the essay is being written, the acknowledgment that the student must approach the essay's subject in such a way as to gain the attention of that audience, and the necessity of having a purpose for the essay, tied to the notion of a thesis, which the essay will set out to prove. Hellstern et al. gave specific features of the essay-writing task that may be constructively used to operationalize the essay-writing task so that it can be used in an interactive IR system to facilitate student performance:

> (i) the creation and supporting of the essay thesis, which sets out the orientation or purpose of the essay as well as determining its structure; and
> (ii) appealing to the essay's intended audience—in the case of an undergraduate essay, the course instructor evaluating the essay when it is finished (p. 125).

The definition of task that combines (i) purpose and (ii) audience takes into account both the cognitive and sociological aspects of the undergraduate's task of essay writing. The cognitive aspect includes the cognitive problem that the student is faced with when trying to organize vast amounts of information into a coherent essay structure. The sociological aspect includes the organizational or problem situational constraints of essay writing (i.e., the task of essay writing takes place within the constraints imposed by a specific social, organizational, or problem situation framework.

The problem situational framework joins together in the definition of the undergraduate's essay writing task as: "putting his/her learning on display" for the course instructor so that the course instructor will award the essay with a high mark (Cole & Leide, 2003, p. 39). We previously focused on the essay's thesis as the primary structuring device for an undergraduate essay because it fulfills both the purpose and audience constraints of the essay-writing task. An effective thesis puts the undergraduate's learning on display by providing the undergraduate with a mechanism for synthesizing disparate pieces of information into a single unified and supported (by evidentiary information) message that demonstrates to the course instructor that critical thinking has taken place.

A second vehicle or mechanism for facilitating the undergraduate's task of putting his or her learning on display for the course instructor is essay type. There are so-called high-impact and low-impact essay types. A descriptive essay, for example, is low impact because it does not put the student's learning on display as much as a high-impact cause and effect essay type does. A well-thought-out essay thesis and a high-impact essay type are both vehicles for performing the student's task of putting his or her learning on display for the course instructor so that this instructor will give the essay a high mark. We will discuss these two vehicles of the student's essay-writing task in further detail in the next two sections.

Task: Essay Thesis

An undergraduate essay thesis statement is the expression of an undergraduate's argument and point of view on an issue. In an academic essay, it often consists of two, three, or even more sentences (Kennedy & Smith, 1994). Sometimes, it can be implied by a series of statements and never actually be explicitly expressed in a sentence. Most textbooks on research essays, however, seem to emphasize the view of a thesis as "a one-sentence summary of the main idea in [a] paper" (McGuen & Winkler, 1995, p. 507).

The thesis usually appears in the introductory paragraph of an essay, most often as its final sentence, although certain persuasive strategies, which seek to present the evidence first and allow the reader a chance to consider it before trying to persuade him or her of the writer's view, might very well save the thesis for the end of the essay to give the argument greater effect (McGuen & Winkler, 1995). Most importantly, a thesis statement should answer a research question. Therefore, in the research question, "Always try to pose a single, challenging question that demands analysis and argument—a question that can be stated briefly and succinctly in just one sentence" (Robertson, 1995, p. 14). The thesis statement is the answer to the research question, a statement not a question, which "moves toward a single point, not diverge into two or more ideas" (McGuen & Winkler, 1995, p. 85). By keeping the thesis to a single point with one idea, it "narrow[s] and define[s] your subject to workable size" (Baker, 1991, p. 20). "The usefulness of the thesis lies in its effect on a writer's efforts … a well-worded thesis can help point us in one direction and provide us with useful limits for our thinking and writing" (McGuen & Winkler, 1995, p. 82). The student should avoid a biographical, narrative, or descriptive answer in the thesis statement (Robertson, 1995).

A thesis might emerge from the type of information found, but the closer the working thesis is to the final thesis, the less time a future writer will spend doing irrelevant research in the present, and the time spent will be maximally efficient in that it is spent searching sources to strengthen an

DOI: 10.1002/asi

argument, not to find one. Even if the students eventually change their mind about it, having a working thesis will maximize the efficiency of their research and give them a reference point, a center, from which to evaluate the relevance and usefulness of the sources they find. A well-worded thesis has the same efficient delimiting effect in the writing process. Therefore, an IR system that attempts to guide a student's information-seeking behavior based on the student's degree of certainty about a working thesis statement should, in theory, maximize the efficiency of that student's information search for the fulfillment of the essay-writing task at hand.

Task: Essay-Type

A second candidate for the essential nature of the task of undergraduate history-essay writing is the concept of essay type. The undergraduate's task when writing a history essay has been previously defined as putting his or her learning on display in the essay so that the course instructor will give the essay a high mark. Therefore, the essay structure selected by the undergraduate will be a key variable in effectively performing the task. There are eight commonly used essay styles (Howard University, 2003):

- *Narration essay*: transformation of an experience into a story. The writer seeks to make a point and to provide a detailed, descriptive, and sometimes, personal account of an event or experience.
- *Descriptive essay/narrative essay*: yields a word-portrait of a person, place, or thing. Provides factual information about a topic (e.g., an encyclopedic entry) or provides a subjective depiction of an experience (e.g., how racial prejudice affects the author or an audience).
- *Process/analysis essay*: provides a prescribed "menu." Provides instruction and information (e.g., "how to" essays, procedural writing).
- *Compare/contrast essay*: demonstrates how things, views, or issues are similar and/or different. Juxtaposition of elements.
- *Definition essay*: identifies and explains the special nature of an institution, theory, philosophy, or group of people.
- *Cause and effect essay*: engages in speculation, asks "what if" questions, and hypothesizes about the factors which might bring about an event or the consequences of certain actions.
- *Persuasion and argument essay*: presents a case and attempts to change the reader's mind. Sometimes includes a call to action.
- *Ambiguous essay*: an essay with no discernible structure.

From the previous list, there are high-impact and low-impact essay types from the point of view of the student putting his or her learning on display so that the essay will receive a high mark from the course instructor. The narration and the descriptive essay types are low impact, involving the collection and narration of events in a straightforward fashion. On the other hand, the persuasion/argument, the cause and effect, and the compare/contrast essay types are high impact, requiring critical thinking in both collecting information and writing the essay. The compare/contrast essay type, for example, is commonly taught to undergraduates in university handouts on essay writing as a high-impact essay style (University of North Carolina at Chapel Hill, 1998):

> By assigning [compare and contrast] essays, your instructors are encouraging you to make connections between texts or ideas, engage in critical thinking, and go beyond mere description or summary to generate interesting analysis: when you reflect on similarities and differences, you gain a deeper understanding of the items you are comparing, their relationship to each other, and what is most important about them.

Pilot Study

In October 2004, a pilot study investigation was undertaken to determine if thesis statement or essay type was the most effective task intervention to improve the quality of students' essays. Keywords and cognitive mapping also were tested as possible task-based IR system devices. The most effective intervention would be tested later against a control group in a large study with a greater number of participants.

Twenty-eight students from two history courses at McGill University volunteered for the pilot study. Students were randomly assigned to the five groups: (a) the keyword group, whose participants were asked to explore keywords for their essay topic; (b) the mapping group, whose participants were asked to create cognitive maps of their essay topics; (c) the research questions/theses group, whose participants were asked to write down four possible research questions/theses for their essay; (d) the essay-type group, whose participants were asked to categorize their planned essay by essay type; and (e) the control group, whose participants were asked to write down their essay topic and outline how they would research and write their essay.

Students were asked to send us a copy of the final essays that they had submitted to the course instructor at the end of their history course (about 6 weeks after our intervention). Only 21 of 28 students sent us their essays. The essays were evaluated and ranked by a "blind reader" who had had no contact with the students or the design of the intervention schedules. The evaluation criteria were distilled from a number of essay grading criteria (Oregon Department of Education, 2004–2005; Western International University, 2004, p. 53; cf. Collegeboard.com Inc.; Hyperhistory.net; North Central Regional Technology in Education Consortium; University of Toronto Undergraduate Studies in History). The evaluation was based on four criteria that succinctly captured what a student must do to put his or her learning on display: extensive treatment of issues, good essay structure, critical thinking, and style (i.e., reads well). Each criterion was graded on a 5-point scale, and the criteria scores were totaled to determine the essay mark.

The results of the pilot study are shown in Table 1. The evaluator assigned letter grades to all essays, which were converted to numeric indications. The mean mark (3.0) for Group IV (essay type) was higher than that of any other

DOI: 10.1002/asi

TABLE 1. Pilot study results.

	Group				
	I (Keyword)	II (Mapping)	III (Question/Thesis)	IV (Essay type)	V (Control)
Mean score	2.5	2.5	2.6	3.0	2.8

group, indicating that essay type may be the most effective factor for assisting students in their task of putting their learning on display. The blind reader also was asked to designate essay type for each essay. The compare/contrast essay type was most predominant among the essays accorded the highest grades while the descriptive-narrative essay type predominated among essays given a lower score, with 7 of the top 13 essays designated compare/contrast versus 12 of the bottom 13 essays designated descriptive-narrative essay type. Compare/contrast is a high-impact essay type while descriptive-narrative is a low-impact essay type.

Main Study

Specific Problem

The specific problem of the main study is "Does structuring the undergraduate's essay topics using an essay-type visualization diagram facilitate the students' task of putting their learning on display in the essay so that they will receive a high mark?"

Methodology

In the 2005 winter term, 80 student volunteers from a single first-year McGill University history class were randomly assigned to a control group and an essay-type intervention group. Of the 80 students, 78 of them were interviewed when each was in Kuhlthau's (1993) Stage 3, the exploration stage, of researching their essay for the history course. The control group was given the interview schedule shown in Appendix A. The intervention group was given the interview schedule shown in Appendix B. The subject of the interview was their real-life essay.

The two interview schedules made use of two unconnected (and coincidental) bibliographic instruction sessions on *Academic Integrity* and *Electronic Resources at McGill* given to this class by librarians from McGill's Humanities and Social Science Library in the weeks preceding the study. We had the participants evaluate the two library instruction sessions, with the control group concentrating more on the evaluation of the library instruction sessions than did the intervention group. Each interviewee in the two groups then was asked to state a topic or thesis for the paper and to list possible search terms for accessing information using an IR system. The control group ended there. The essay-type intervention group then was presented with a series of questions about essay type.

For essay type, from the list of eight types of essays (discussed earlier), we limited the participants' choice to the compare/contrast essay type and the cause and effect essay type. These two essay types had proven popular in the pilot study and were practical in terms of the visualization requirements of our 3-year research project (cf. Leide et al., 2003). The study participants selected which of the two essay types they themselves believed that they would use in their own essay and were asked to complete a visualization diagram for the essay type that they had selected, either the compare/contrast essay type (Figure 1) or the cause and effect essay type (Figure 2). The compare/contrast essay visualization is a modification of commonly noted visualizations

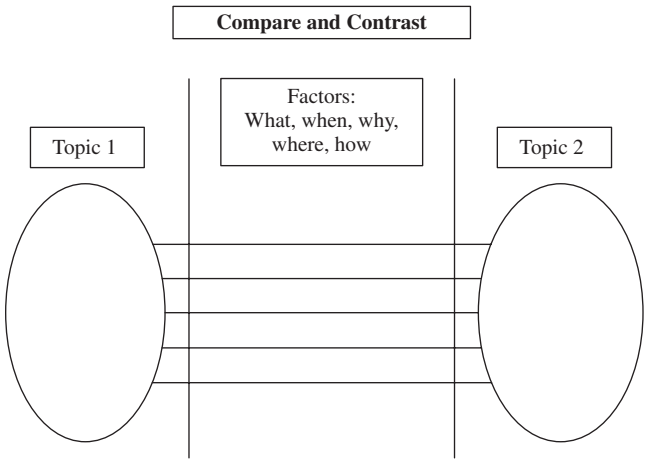

FIG. 1. Compare/contrast essay type.

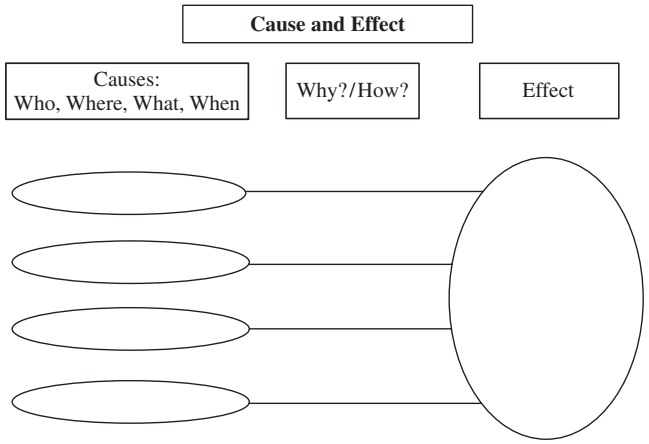

FIG. 2. Cause and effect essay type.

DOI: 10.1002/asi

of this essay type using intersecting circle Venn diagrams (e.g., University of North Carolina at Chapel Hill, 1998). The cause and effect essay visualization is a first attempt on our part to visualize the structure of this essay type.

The dependent variable was an evaluation of the students' essays. There were two separate evaluation procedures used to grade the student papers. In the first evaluation, the term papers were evaluated by a blind reader according to the same criteria as in the pilot study. A second evaluation was carried out by obtaining the essay marks given to the essays by the history-course instructor (A description of the procedures for obtaining these marks is discussed later.)

Data Collection

For the blind reader evaluation, we received 72 essays from the 78 students interviewed (92%). Six students decided not to provide their essays for a variety of reasons (e.g., 1 student withdrew from the class). Appendix C contains the data from the blind reader evaluation. This includes the evaluation ranking for each essay (Column 1), the grade mark the blind reader assigned each essay (Column 2), whether the student was in the control or intervention group, and for the intervention group, whether the student selected a compare/contrast or cause and effect essay type (Column 3), and the number of search terms given by each student during the interview (Column 4).

Results

Blind reader evaluation. The findings from the blind reader evaluation are summarized in Table 2. The letter grades from Appendix C were converted into numeric form to facilitate calculations. Results indicate that those in the control group received an average mark of 3.0 points from our blind reader while the intervention group received an average mark of 2.9. The intervention group was divided into two subgroups: The cause and effect essay group received 2.8 points, and the compare/contrast essay group received 3.3 points.

To cross-check our evaluation procedure, we conducted further data analysis based on the interviews. Two results from the analysis support the evaluation procedure. First, of the 72 essays, 8 (11%) students changed their topics between the interview intervention and the final version of essay. All 8 essays were ranked 32nd or below in their evaluation, indicating that students whose essays ranked highest did not change their topics. We make the assumption here that early and good preparation may be indicative of better quality learning strategies.

Second, during the interview, participants also were asked "What words, search terms, keywords or concepts will you use to look for information to write your paper?" (Appendices A and B). The blind reader did not look at the search terms to evaluate the quality or appropriateness of the search terms themselves (for a study in our research program that did, cf. Yi et al., 2006). However, from Appendix C, students with a high grade listed considerably more search terms than did students with low-evaluated essays. It has not been established that a high number of search terms is an indicator of high-quality student performance in essay writing; however, there may be a positive correlation between these two variables as some studies have found that domain experts use more search terms in their queries than do domain novices (Hembrooke, Granka, Gay, & Liddy, 2005; Holscher & Strube, 2000; Lucas & Topi, 2002; but for counter evidence, see Hsieh-Yee, 1993). There also is the belief among researchers that users employ too few search terms in their queries and that using more would increase search performance (Alemayehu, 2003; Crouch, et al., 2002; Spink & Jansen, 2004). In addition, there is inferential evidence based on longitudinal studies in which participants employed more search terms at the end of the research process than at the beginning (Vakkari, 2000a) or at the end of a 30-min intervention devoted to essay-topic exploration with the participants (Cole, Leide, Large, et al., 2005) because supposedly they knew more about their topic at the end of the process.

Course-instructor evaluation. To collect the course-instructor evaluation data, a McGill graduate student unconnected with the research study and hired only for this specific task sorted data from class marks into (a) the study's research groups and (b) the large group of students in the history course who did not volunteer to participate in our study, called nonparticipants. To protect the anonymity of the students, names were removed by the course instructor before the marks were given to the research assistant; we therefore could not refer each mark to a specific participant. Aggregated data were used to see if there were any statistically significant differences between the groups (Table 3). The 6 participants in our study who were interviewed but did not submit their essays to us were not included in the blind reader evaluation, but they are

TABLE 2. Grade distribution of participants—"Blind Reader" evaluation.

Grade distribution by group[a]	A	A−	B+	B	B−	C+	C	C−	Average Grade
Intervention–Cause & effect (n = 29)	5	0	2	7	6	4	1	4	B/B− (2.8)
Intervention–Compare & contrast (n = 5)	1	1	1	1	1	0	0	0	B+ (3.3)
Intervention: TOTAL (n = 34)	6	1	3	8	7	4	1	4	B/B− (2.9)
Control (n = 38)	7	6	5	7	4	2	3	4	B (3.0)

[a]A = 4.0; A− = 3.6; B+ = 3.3; B = 3; B− = 2.6; C+ = 2.3; C = 2; C− = 1.6.

DOI: 10.1002/asi

TABLE 3. Course instructor evaluations.

Group	No. of students	Average mark (%)
Intervention–Cause & effect	31	78
Intervention–Compare & contrast	7	74
Control	40	77
Nonparticipants	165	73

TABLE 4. Test statistics—student grades (participants and nonparticipants).

Grouping Variable Participants vs Nonparticipants	Student grades
Mann–Whitney *U*	5106.5
Wilcoxon *W*	18801.5
Z	−2.6
Asymptotic significance (two-tailed)	0.009

included here in the course instructor evaluation. One study participant from the cause and effect group and 14 nonparticipants did not submit an essay to the course instructor and thus were excluded from the analysis.

From Table 3, the average mark obtained from the course instructor for each of the groups ranged from 78% for the intervention-cause and effect group to 73% for the nonparticipant group. The control group received an average mark between the two essay-type intervention groups. We decided to first determine if the grades of the students who did not volunteer to participate in the study were significantly different from those who did participate. Due to unequal variances between participant and nonparticipant groups, the Mann–Whitney *U*-test was carried out using SPSS Version 11.0 (Table 4). From Table 4, the results of the analysis indicate that there is a significant difference between the mean grade of the students in the two groups ($p = .009$). This difference may have stemmed from a self-selection process, where students who participated were more interested in their academic progress and achievement than were the nonparticipants. Second, different from the independent variables being studied, the interview process itself may have helped to hone the participants' research and writing practices. Finally, the participants' behavior may have been altered through a Hawthorne effect.

TABLE 5. Test statistic—student grades (intervention groups and control).

Kruskal–Wallis test (grouping variable: interview) ($p = .7$)	Grades
χ^2	0.759
df	2
Asymptotic significance	0.695

TABLE 6. Group statistics.

Student grades	No. of students	*M*	*SD*	*SEM*
Cause & effect group	31	78	8	1.45
Control group	40	77	9	1.50

Intervention groups versus the control group. The central focus of the study was to determine if the two essay-type intervention groups obtained higher evaluations for their essays than did the control group. There were two essay-type interventions: a compare/contrast essay type and a cause and effect essay type. During the interview, only 7 of 78 students indicated that they intended to write compare/contrast essays. Due to the unequal sample size, the Kruskal–Wallis test (Table 5), a nonparametric test, therefore was used to analyze only the rankings of the data points and not their absolute values to determine whether there was a significant difference in the mean grades among the intervention groups (cause and effect: $n = 31$; compare & contrast, $n = 7$) and the control group ($n = 40$). The results show that the mean grades for all the groups are relatively similar ($p = .695$).

Table 6 indicates that there were relatively equal variances between the cause and effect group, which received the highest mean mark of all the groups from the course instructor, and the control group. Therefore, a *t* test was used to further investigate the difference between these two groups, but the results of the test in Table 7 indicate no significant difference between these groups ($p = .3$).

Discussion

The study reported in this article is the final study in a 3-year research project. Here, we discuss our findings and highlight the challenges facing user-centered, field-study IR research.

TABLE 7. Independent samples test—Student grades (cause & effect and control groups).

			t test for equality of means ($p = .3$)						
	Levene's test for equality of variances							95% Confidence interval of the difference	
Student grades	*F*	Sig.	*t*	*df*	Sig. (two-tailed)	*M* Difference	*SE* Difference	Upper	Lower
Equal variance assumed	0.96	0.33	0.646	69	0.52	1.38	2.13	−2.87002	5.62163
Equal variances not assumed			0.66	68.4	0.51	1.38	2.08	−2.78269	5.53431

DOI: 10.1002/asi

The central focus of the main study was to determine if an essay-type visualization could facilitate task performance for a group of undergraduates writing a history essay. Task performance was evaluated by a blind reader from the research project and a real-life evaluation of the essay by the undergraduates' course instructor. The difference in mean marks between the intervention and control groups was negligible for both evaluations; however, for the course-instructor evaluation, students who participated in the study received higher mean marks for their essays than did students in the same course who had not participated in the study. The difference was statistically significant. As previously stated, the difference may have been due to a self-selection process during volunteering for the study, something in the interview besides the variables being studied, or the Hawthorne effect; however, the statistically significant difference between participants and nonparticipants also may indicate that both essay style and control interventions had a positive effect on the performance of the participants. We were ethically constrained from giving the control group a meaningless activity that would disadvantage this group of students vis-à-vis the students who received the essay-style intervention.

Challenge 1

With no equivalent to medicine's placebo in social science human participants studied in real-life settings with real performance-evaluation criteria, how do researchers control for bias in the sampling using a control group?

Challenge 2

What is the optimum design for essay-type IR visualizations whose purpose is to facilitate task performance? In the main study, essay type was selected as the intervention because the pilot study demonstrated that essay type produced the highest impact on the students' task performance. Cause and effect and compare/contrast were selected as essay types in the main study because they were practical to visualize and because they visualize high-impact essay types, facilitating the students' task as we defined it in the study—putting their learning on display for the course instructor. As shown in Figures 1 and 2, the visualizations of these essay types stress the concepts and relational structure of the compare/contrast and cause and effect types, respectively. Such visualizations may eventually serve the purpose of forming query-formulation devices for task-based IR, but we are only in the beginning stages of designing such visualizations and testing the effect of such visualizations on task performance during user–IR system interaction.

Challenge 3

Is there a difference between the essay style selected by students and the essay style rewarded by course instructors? In the main study, it surprised us that so many students in the intervention group selected the cause and effect essay type over the compare/contrast essay type; however, this finding confirms our research evaluation of undergraduate history essays in an earlier part of our research program which showed that the cause and effect essay is the most popular high-impact essay structure (Cole, Leide, Beheshti, et al., 2005). In a previous research project, the second author found that domain experts-in-training, Ph.D. history students, compared and contrasted their interpretation of historical events with the status quo interpretation of the same events, based on new evidence they had collected in their research (Cole, 2000). By comparing and contrasting, the Ph.D. history students, who were learning to become domain experts, gained a deep, original understanding of an event, an issue, or a problem, which according to Gaddis (2002) is essential to the historical method: "to comprehend something is to see it in relation to other entities of the same class" (p. 25). That the domain novice undergraduate does not do history in the same way as a domain expert, perhaps because of the student's own perception or misperception of what the course instructor requires or is looking for in an essay, increases the challenge facing user-centered IR design. An exploratory study using grounded theory or some other qualitative research method may be able to elucidate this question.

Conclusion

The central assumption of this article is that a task-focused visualization device attached to an IR system can be effectively used by undergraduates researching essays if it could focus their information-search efforts toward their central task of creating an essay which most effectively puts their learning on display for the person evaluating the essay, the course instructor. Researchers in the IR field are beginning to think beyond the retrieved document set when evaluating IR system performance to looking at how the user integrates the found information into an essay or some other task document. In task-based IR, the evaluation variable for successful IR-system performance should, we argue, be tied to the user's task performance—in this case, the mark given to the undergraduate's essay by the course instructor. We tested a number of task-performance devices in the pilot study, which determined that an essay-type intervention had the greatest positive effect on undergraduate task performance. The main study tested two essay-type interventions against a control group, with two evaluation procedures as the dependent variable measuring undergraduate task performance. The findings from the study indicate that the two essay-type interventions had no greater impact on undergraduate task performance than it did on the control-group intervention. Both the essay type and control interventions, however, were strongly designed to organize and clarify students' thinking about their essay, and together, the students in the control and essay-type intervention groups scored significantly higher in the course-instructor evaluation of the student essays than did the group of nonparticipants in the same history course.

DOI: 10.1002/asi

An interesting challenge in designing task-based IR systems for specific user groups is the lack of connection between how a domain novice member of the group perceives his or her task and the task itself. The undergraduate essay writer is a specific example of this. The definitional constraints of the undergraduate's task of writing an essay are set by forces outside the undergraduate's control. The undergraduate must learn about these constraints, perhaps through trial and error. But it is not in the undergraduate's interest for there to be more trial than error in information search. Even though the successful performance of the undergraduate's task is a goal state in which trial and error is inevitable—that is, only after much trial and error does the undergraduate come to an understanding of what it is he or she is really doing and what information is really needed to perform the task (Cole, Beheshti, et al., 2005)—it is our contention that trial and error in information-search activities can be greatly reduced by codifying aspects of the undergraduate's task into the interaction with the IR system in advance, thus providing a beacon or organizing compass for the user even when he or she is just starting out in the process of information seeking and search. We still believe that essay type, such as the compare/contrast and cause and effect essay types, or some other task-based visualization device has the potential to structure the undergraduate's history essay for greater efficiency in accessing information from an IR system, and organizing the results list.

Acknowledgments

We acknowledge the financial support of the Social Sciences and Humanities Research Council (SSHRC) of Canada, File No. 410–2002–1394. We also thank research assistants Sarah Jensen, Emily Love, and Stephanie Lemieux.

References

Alemayehu, N. (2003). Analysis of performance variation using query expansion. Journal of the American Society for Information Science and Technology, 54(5), 379–391.

Annett, J., Duncan, K.D., Stammers, R.B., & Gray, M.J. (1971). Task analysis. London: HMSO.

Baker, S. (1991). The practical stylist (3rd ed.). New York: HarperCollins.

Bartlett, J.C. (2004). Connecting bioinformatics analysis to scientific practice: An integrated information behaviour and task analysis approach. Unpublished doctoral dissertation, Toronto: University of Toronto Faculty of Information Studies.

Bates, M.J. (1979). Information search tactics. Journal of the American Society for Information Science, 30, 205–214.

Bates, M.J. (1990). Where should the person stop and the information search interface start? Information Processing and Management, 26(5), 575–801.

Belkin, N., Cool, C., Stein, A., & Thiel, U. (1995). Cases, scripts and information-seeking strategies: On the design of interactive information retrieval systems. Expert Systems With Applications, 9(3), 379–395.

Belkin, N.J., & Marchetti, P.G. (1990). Determining the functionality and features of an intelligent interface to an information retrieval system. Proceedings of the 13th International Conference on Research and Development in Information Retrieval: Universite libre de Bruxelles (pp. 151–177). Brussels, Belgium: Universite libre de Bruxelles.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982a). ASK for information retrieval: Part I. Background and theory. Journal of Documentation, 38(2), 61–71.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982b). ASK for information retrieval: Part II. Results of a design study. Journal of Documentation, 38(3), 145–164.

Belkin, N.J., Seeger, R., & Wersig, G. (1983). Distributed expert problem treatment as a model for information system analysis and design. Journal of Information Science, 5(5), 153–167.

Benyon, D.R. (1992). The role of task analysis in systems design. Interacting With Computers, 4(2), 102–123.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. Information Processing and Management, 31(2), 191–213.

Card, S., Moran, T., & Newell, A. (1983). The psychology of human–computer interaction. Hillsdale, NJ: Erlbaum.

Case, D.O. (2002). Looking for information: A survey of research on information seeking, needs, and behavior. Amsterdam: Academic Press.

Cole, C. (1994). Operationalizing the notion of information as a subjective construct. Journal of the American Society for Information Science, 45(7), 465–476.

Cole, C. (2000). Inducing expertise in history doctoral students via enabling information retrieval design. Library Quarterly, 70(1), 86–109.

Cole, C., Beheshti, J., Leide, J.E., & Large, A. (2005). Interactive information retrieval: Bringing the user to a selection state. In A. Spink & C. Cole (Eds.), New directions in cognitive information retrieval (pp. 13–41). Dordrecht, The Netherlands: Springer.

Cole, C., Cantero, P., & Sauve, D. (1998). Intelligent information retrieval: Diagnosing information need: Part II. Uncertainty expansion in a prototype of a diagnostic IR tool. Information Processing & Management, 34(6), 721–737.

Cole, C., Kennedy, L., & Carter, S. (1996). The optimization of online searches through the labeling of a dynamic, situation-dependent information need: The reference interview and online searching for undergraduates doing a social science assignment. Information Processing & Management, 32, 709–717.

Cole, C., & Kuhlthau, C. (2000). Information and information seeking of novice versus expert lawyers: How experts add value. The New Review of Information Behaviour Research, 1, 103–115.

Cole, C., & Leide, J. (2003). Using the user's mental model to guide integration of information space into information need. Journal of the American Society for Information Science and Technology, 54(1), 39–46.

Cole, C., Leide, J.E., Beheshti, J., Large, A., & Brooks, M. (2005). Investigating the ASK hypothesis in a real-life problem situation: A study of history and psychology undergraduates seeking information for a course essay. Journal of the American Society for Information Science and Technology, 56(14), 1544–1554.

Cole, C., Leide, J.E., Large, A., Beheshti, J., & Brooks, M. (2005). Putting it together online: Information need identification for the domain novice user. Journal of the American Society for Information Science and Technology, 56(7), 684–694.

Collegeboard.com Inc. (2007). Collegeboard Scoring Guide. Retrieved August 18, 2005, from http://www.collegeboard.com/student/testing/sat/about/sat/essay_scoring.html

Crouch, C.J., Crouch, D.B., Chen, Q., & Holtz, S.J. (2002). Improving the retrieval effectiveness of very short queries. Information Processing & Management, 38(1), 1–36.

Crystal, A., & Ellington, B. (2004). Task analysis and human–computer interaction: Approaches, techniques, and levels of analysis. Proceedings of the 10th Americas Conference on Information Systems, New York (pp. 1–9). Retrieved from http://www.ils.unc.edu/Nacrystal/AMC1504_crystal_ellington_final.pdf

Culnan, M.J. (1983). Environmental scanning: The effects of task complexity and source accessibility on information gathering behavior. Decision Sciences, 14, 194–206.

Drabenstott, K.M. (2003). Do nondomain experts enlist the strategies of domain experts? Journal of the American Society for Information Science and Technology, 54(9), 836–854.

DOI: 10.1002/asi

Ellis, D. (1996). The dilemma of measurement in information retrieval research. Journal of the American Society for Information Science, 47(1), 23–36.

Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. Journal of Documentation, 49(4), 356–369.

Entwistle, N., & Entwistle, D. (2003). Preparing for examinations: The interplay of memorizing and understanding, and the development of knowledge objects. Higher Education Research and Development, 22, 19–41.

Fidel, R. (1991a). Searchers' selection of search keys: I. Journal of the American Society for Information Science, 42(7), 490–500.

Fidel, R. (1991b). Searchers' selection of search keys: II. Journal of the American Society for Information Science, 42(7), 501–514.

Fidel, R. (1991c). Searchers' selection of search keys: III. Journal of the American Society for Information Science, 42(7), 515–527.

Fisher, K.E., Durrance, J.C., & Hinton, M.B. (2004). Information grounds and the use of need-based services by immigrants in Queens, New York: A context-based, outcome evaluation approach. Journal of the American Society for Information Science and Technology, 55(8), 754–766.

Fujihara, N., & Miura, A. (2003). Patterns of searching for information on the World Wide Web: A pilot study. Psychological Reports, 92, 1091–1096.

Gaddis, J.L. (2002). The landscape of history: How historians map the past. Oxford, England: Oxford University Press.

Hackman, J.R. (1969). Toward understanding the role of tasks in behavioral research. Acta Psychologica, 31, 97–128.

Hart, P.J., & Rice, R.E. (1991). Using information from external databases: Contextual relationships of use, access method, task, database type, organizational differences, and outcomes. Information Processing & Management, 27(5), 461–479.

Harter, S.P. (1986). Online information retrieval: Concepts, principles, and techniques. San Diego, CA: Academic Press.

Harter, S.P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43(9), 602–615.

Harter, S.P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. Journal of the American Society for Information Science, 47(1), 37–49.

Harter, S.P., & Hert, C. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. Annual Review of Information Science and Technology, 32, 3–94.

Hearst, M.A. (1999). User interfaces and visualization. In B.-Y. Ricardo & R.-N. Berthier (Eds.), Modern information retrieval (pp. 257–323). New York: ACM Press.

Hellstern, M., Scott, G.M., & Garrison, S.M. (1998). The history student writer's manual. Upper Saddle River, NJ: Prentice Hall.

Hembrooke, H.A., Granka, L.A., Gay, G.K., & Liddy, E.D. (2005). The effects of expertise and feedback on search term selection and subsequent learning. Journal of the American Society for Information Science and Technology, 56(8), 861–871.

Hersh, W., Pentecost, J., & Hickam, D. (1996). A task-oriented approach to information retrieval evaluation. Journal of the American Society for Information Science, 47(1), 50–56.

Hildreth, C.R. (2001). Accounting for users' inflated assessments of online catalogue search performance and usefulness: An experimental study. Information Research, 6(2). Retrieved April 18, 2006, from http://InformationR.net/ir/6–2/paper101.html

Holscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. International Journal of Computer and Telecommunications Networking, 33(1–6), 337–346.

Howard University. (2003). Essay styles. Howard University College of Arts and Sciences, Washington, DC. Retrieved July 8, 2003, from http://www.howard.edu/CollegeArtsSciences/Freshman%20Seminar/essay_styles.htm

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 45(3), 161–174.

Hyperhistory.net. (2005). Essay Scoring Guides (Rubrics). Retrieved August 18, 2005, from http://www.hyperhistory.net/apwh/exam/all3rubrics.htm

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. Journal of Documentation, 52(1), 3–50.

Jansen, B.J., & McNeese, M.D. (2005). Evaluating the effectiveness of and patterns of interactions with automated searching assistance. Journal of the American Society for Information Science and Technology, 56(14), 1480–1503.

Järvelin, K., & Ingwersen, P. (2004). Information seeking research needs extension towards tasks and technology. Information Research, 10(1), Paper No. 212. Available: http://informationR.net/ir/10–1/paper212.html

Kennedy, L., Cole, C., & Carter, S. (1999). The false focus in online searching: The particular case of undergraduates seeking information for course assignments in the humanities and social sciences. Reference & User Services Quarterly, 38, 267–273.

Kennedy, M.L., & Smith, H.M. (1994). Reading and writing in the academic community. Englewood Cliffs, NJ: Prentice Hall.

Kim, K.-S. (2001). Information-seeking on the Web: Effects of user and task variables. Library & Information Science Research, 23(3), 233–257.

Kirwan, B., & Ainsworth, L.K. (Eds.). (1992). A guide to task analysis. London: Taylor & Francis.

Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.

Kuhlthau, C.C. (1999). Accommodating the user's information search process: Challenges for information retrieval system designers. Bulletin of the American Society for Information Science, 25(3), 12–16.

Kuhlthau, C.C. (2004). Seeking meaning: A process approach to library and information services (2nd ed.). Westport, CT: Libraries Unlimited.

Leckie, G.J., Pettigrew, K.E., & Sylvain, C. (1996). Modelling the information seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers. Library Quarterly, 66(2), 161–193.

Leide, J., Large, A., Beheshti, J., Brooks, M., & Cole, C. (2003). Visualization schemes for domain novices exploring a topic space: The navigation classification scheme. Information Processing & Management, 39(6), 923–940.

Leplat, J. (1989). Relations between task and activity: Elements of elaborating a framework for error analysis. Ergonomics, 33, 1389–1402.

Line, M.B. (1971). The information uses and needs of social scientists: An overview of INFROSS. Aslib Proceedings, 23(8), 412–434.

Locke, E.A., Shaw, K.N., Saari, L.M., & Latham, G.P. (1981). Goal setting and task performance: 1969–1980. Psychological Bulletin, 90, 125–152.

Lucas, W., & Topi, H. (2002). Form and function: The impact of query term and operator usage on Web search results. Journal of the American Society for Information Science and Technology, 53(2), 95–108.

March, J., & Simon, H. (1967). Organizations (2nd ed.). New York: Wiley.

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge, England: Cambridge University Press.

Markkula, M., Tico, M., Sepponen, B., Nirkkonen, K., & Sormunen, E. (2001). A test-collection for the evaluation of content-based image retrieval algorithms—A user and task-based approach. Information Retrieval, 4(3/4), 275–293.

McGuen, J.R., & Winkler, A.C. (1995). From idea to essay: A rhetoric, reader, and handbook (7th ed.). Toronto: Allyn & Bacon.

North Central Regional Technology in Education Consortium. Scoring Guide for Student Products. Retrieved April 10, 2007, from http://goal.learningpt.org/spsg/GetProd.asp

O'Day, V.L., & Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. In INTERCHI'93 Proceedings of the Conference on Human Factors in Computing Systems, Amsterdam (pp. 438–445). New York: ACM Press.

Oddy, R.N., Liddy, E.D., Balakrishnan, B., Bishop, A., Elewononi, J., & Martin, E. (1992). Towards the use of situational information in information retrieval. Journal of Documentation, 48(2), 123–171.

Oregon Department of Education. (2004–2005). Writing Scoring Guide. Retrieved August 11, 2005, from http://www.ode.state.or.us/teachlearn/testing/scoring/guides/2004–05/writingscoringguide0405.pdf

Patel, V., & Ramoni, M. (1997). Cognitive models of directional inference in expert medical reasoning. In P. Feltovich, K. Ford, & R. Hoffman (Eds.),

DOI: 10.1002/asi

Expertise in context: Human and machine (pp. 67–99). Menlo Park, CA: AAAI Press.
Reeb, B., & Gibbons, S. (2004). Students, librarians, and subject guides: Improving a poor rate of return. Libraries and the Academy, 4(1), 123–130.
Reid, J. (2000). A task-oriented non-interactive evaluation methodology for IR systems. Information Retrieval, 2(1), 115–129.
Rieh, S.Y. (2004). On the Web at home: Information seeking and Web searching in the home environment. Journal of the American Society for Information Science and Technology, 55(8), 743–753.
Robertson, H. (1995). The research essay: A guide to essays and papers (3rd ed.). Montreal: McGraw-Hill.
Ruthven, I. (2005). Integrating approaches to relevance. In A. Spink & C. Cole (Eds.), New directions in cognitive information retrieval (pp. 61–80). Dordrecht, The Netherlands: Springer.
Saracevic, T. (1996a). Interactive models in information retrieval (IR): A review and proposal. Proceedings of the 80th annual meeting of the American Society of Information Science, 33, 3–9.
Saracevic, T. (1996b). Relevance reconsidered '96. In P. Ingwersen & N.O. Pors (Eds.), Proceedings of the 2nd International Conference on Conceptions of Library and Information Science: Integration in Perspective (pp. 201–218). Copenhagen, Denmark: The Royal School of Librarianship.
Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of way of life. Library and Information Science Research, 17, 259–294.
Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A reexamination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26, 755–776.
Spink, A., & Cole, C. (2001). Everyday life information seeking research. Library and Information Science Research, 23(4), 301–304.
Spink, A., & Cole, C. (2005). A multitasking framework for cognitive information retrieval. In A. Spink & C. Cole (Eds.), New directions in cognitive information retrieval (pp. 99–112). Dordrecht, The Netherlands: Springer.
Spink, A., & Jansen, B.J. (2004). Web search: Public searching of the Web. Dordrecht, The Netherlands: Kluwer.
Stieg, M.F. (1981). The information needs of historians. College & Research Libraries, 42, 549–560.
Stoan, S.K. (1984). Research and library skills: An analysis and interpretation. College & Research Libraries, 45, 99–109.
Sutton, S. (1994). The role of attorney mental models of law in case relevance determinations: An exploratory analysis. Journal of the American Society for Information Science, 45(3), 186–200.
Taylor, F. (1911). Scientific management. New York: Harper & Row.
Tiamiyu, M.A. (1992). The relationships between source use and work complexity, decision-maker discretion and activity duration in Nigerian government ministries. International Journal of Information Management, 12, 130–141.
Troll, D.A. (2002). How and why libraries are changing: What we know and what we need to know. Library and the Academy, 2(1), 99–123.
Tushman, M.L. (1978). Technical communication in R&D laboratories: The impact of project work characteristics. Academy of Management Journal, 21(4), 624–645.
University of North Carolina at Chapel Hill. (1998). Writing the paper. Handouts and links. Retrieved April 12, 2006, from http://www.unc.edu/depts/wcweb/handouts/comparison_contrast.html
University of Toronto, Undergraduate Studies in History. The preparation of historical essays. Retrieved August 18, 2005, from http://www.chass.utoronto.ca/history/undergraduate/essays.html
Vakkari, P. (1999). Task complexity, problem structure and information actions integrating studies on information seeking and retrieval. Information Processing & Management, 35, 819–837.
Vakkari, P. (2000a). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. Proceedings of the RIAO 2000 conference (pp. 894–907). Paris: C.I.D. 2000.
Vakkari, P. (2000b). Relevance and contributing information types of searched documents in task performance. Proceedings of the 23rd annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2–9).
Vakkari, P. (2003). Task-based information searching. Annual Review of Information Science and Technology, 37, 413–464.
Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. Journal of Documentation, 56(5), 540–562.
Vakkari, P., & Järvelin, K. (2005). Explanation in information seeking and retrieval. In A. Spink & C. Cole (Eds.), New directions in cognitive information retrieval (pp. 113–138). Dordrecht, The Netherlands: Springer.
Van de Ven, A., & Ferry, D. (1980). Measuring and assessing organizations. New York: Wiley.
Wersig, G. (1973). Informationssoziologie: Hinweise zu einem informationswissenschaftlichen Teilbereich [Information sociology: References to an information science specialty]. Frankfurt aM: Athenäum Fischer.
Western International University. (2004). Writing Style Handbook. Retrieved August 1, 2005, from http://www.wintu.edu/pdf/writingstyles-handbook.pdf
Wildmuth, B.M. (2004). The effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science and Technology, 55(3), 246–258.
Wilson, T.D. (1997). Information behavior: An interdisciplinary perspective. Information Processing & Management, 33(4), 551–572.
Wilson, T.D. (1999). Models in information behaviour research. Journal of Documentation, 55, 249–270.
Wood, R.E., Mento, A.J., & Locke, E.A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. Journal of Applied Psychology, 72(3), 416–425.
Yi, K., Beheshti, J., Cole, C., Leide, J.E., & Large, A. (2006). User search behavior of domain-specific IR systems: An analysis of the query logs from PsycINFO and ABC-Clio's Historical Abstracts/America: History and Life. Journal of the American Society for Information Science and Technology, 57(9), 1208–1220.

Appendix A

Interview Schedule: Control Group Academic Integrity

1. Did you attend the session on Academic Integrity presented by Karen Nicholson? **yes no**
2. How valuable was this presentation for you in preparing to write your term paper. **not at all very little somewhat quite a bit extremely**
3. What is the most important thing you learned from this presentation?
4. If the presentation were to be given again what suggestions do you have to improve it? McGill Databases
5. Did you attend the session on McGill Databases presented by Phyllis Rudin? **yes no**
6. Had you used MUSE prior to the presentation? **yes no**
7. How valuable was this presentation for you in preparing to write your term paper. **not at all very little somewhat quite a bit extremely**
8. What is the most important thing you learned from this presentation?
9. If the presentation were to be given again what suggestions do you have to improve it?
10. In what year of your undergraduate program are you in? **1st 2nd 3rd 4th**

Please write down the topic of your term paper for HIST 215.

What words, search terms, keywords or concepts will you use to look for information to write your paper? (include terms from your topic sentence above)

DOI: 10.1002/asi

Please give the following in one sentence each for your essay:

a. Background information
b. Thesis statement
c. Major topics
d. Minor topics
e. Conclusion

Appendix B

Interview Schedule: Intervention Group Academic Integrity

1. Did you attend the session on Academic Integrity presented by Karen Nicholson? **yes no**
2. How valuable was this presentation for you in preparing to write your term paper. **not at all very little somewhat quite a bit extremely**
3. What is the most important thing you learned from this presentation?
4. If the presentation were to be given again what suggestions do you have to improve it?
   McGill Databases
5. Did you attend the session on McGill Databases presented by Phyllis Rudin? **yes no**
6. Had you used MUSE prior to the presentation? **yes no**
7. How valuable was this presentation for you in preparing to write your term paper. **not at all very little somewhat quite a bit extremely**
8. What is the most important thing you learned from this presentation?
9. If the presentation were to be given again what suggestions do you have to improve it?
10. In what year of your undergraduate program are you in? **1st 2nd 3rd 4th**

Please write down the topic of your term paper for HIST 215.

What words, search terms, keywords or concepts will you use to look for information to write your paper? (include terms from your topic sentence above)

Any topic can be broken down into six essential components: Who, What, Where, When, Why and How. Please categorize your search terms in to these categories below.

i. WHO are the characters?
ii. WHAT is the event?
iii. WHEN is the time period?
iv. WHERE is the location?
v. WHY did the event happen?
vi. HOW did the event happen?

Every good essay has a beginning, a middle, and an end. The research essay is a paper on the research findings on the background of, knowledge and views about, a specific subject. Essays can have a number of different structures depending on their purpose. Two of the most common are:

Compare/contrast essay: demonstrates how things, views, or issues are similar and/or different. Juxtaposition of elements.

Cause and effect essay: engages in speculation, asks "what if" questions, and hypothesizes about the factors which might bring about an event or the consequences of certain actions.

Which type of essay do you plan to write?

For the essay-type you selected, please fill in diagrams (shown in Figures 1 and 2).

Please give the following in one sentence each for your essay:

a. Definition of terms
b. Background information
c. Thesis statement
d. Major topics
e. Minor topics
f. Conclusion

DOI: 10.1002/asi

Appendix C

Main Study Data Table—Evaluation of Term Papers and Interview Analysis

Rank	Grade	Group	No. of Terms[a]	Rank	Grade	Group	No. of Terms[a]
1	A	C	13	37	B	I-CE	5
2	A	I-CE	10	38	B	C	9
3	A	C	4	39	B	C	5
4	A	I-CE	7	40	B	I-CC	6
5	A	I-CE	6	41	B	I-CE	8
6	A	C	5	42	B	C	6[b]
7	A	C	8	43	B	I-CE	8
8	A	I-CE	9	44	B−	I-CE	6
9	A	I-CC	8	45	B−	I-CE	15
10	A	C	11	46	B−	I-CE	4
11	A	C	10	47	B−	I-CC	8
12	A	C	5	48	B−	I-CE	6
13	A	I-CE	8	49	B−	C	5
14	A−	C	4	50	B−	I-CE	4[b]
15	A−	I-CC	7	51	B−	C	7
16	A−	C	6[b]	52	B−	I-CE	5
17	A−	C	7	53	B−	C	3[b]
18	A−	C	8[b]	54	B−	C	7
19	A−	C	6	55	C+	I-CE	5[b]
20	A−	C	6	56	C+	C	7
21	B+	I-CE	8	57	C+	I-CE	12
22	B+	C	4	58	C+	I-CE	10
23	B+	C	7	59	C+	I-CE	6
24	B+	C	5	60	C+	C	6[b]
25	B+	C	8	61	C	I-CE	3
26	B+	C	9	62	C	C	6
27	B+	I-CC	5	63	C	C	8
28	B+	I-CE	8	64	C	C	5
29	B	C	5	65	C−	C	8
30	B	I-CE	7	66	C−	I-CE	2
31	B	C	11	67	C−	C	9
32	B	I-CE	7	68	C−	C	4
33	B	C	9	69	C−	I-CE	6
34	B	I-CE	4	70	C−	I-CE	6[b]
35	B	I-CE	5	71	C−	I-CE	3
36	B	C	8	72[c]	C−	C	14

Note. C = Control; I-CC = Intervention Compare & Contrast; I-CE = Intervention Cause & Effect.

[a]For term count, repeated words were only counted once; words connected by Boolean Operators ["and," "or," (or symbols, e.g., "/," "&,"or "+")] were calculated separately, if they were not repeated; and words connected by "-" (or other special punctuation, e.g., "}" or "<") were counted together as one term.

[b]Dates, such as 1933–1941, were treated as temporal qualifiers and omitted from the term count.

[c]Paper 72 appears to be an outlier with 14 search terms.

DOI: 10.1002/asi