A Day without a Search Engine: An Experimental Study of Online and Offline Search*

Yan Chen Grace YoungJoo Jeon Yong-Mi Kim

November 15, 2010

Abstract

With the advent of the Web and search engines, online search has become a common method of obtaining information. The question arises as to how much time people save by using search engines for their information needs, and the extent to which online search affects search experiences and outcomes. Using a random sample of queries from a major search engine, we conduct an experiment to compare online and offline search experiences and outcomes. We find that participants are significantly more likely to find an answer on the Web. Restricting to the set of queries which participants find answers in both treatments, the average search time is 22 minutes offline, and 7 minute online. While library sources are judged to be significantly more trustworthy and authoritative than the corresponding web sources, web sources are judged to be significantly more relevant and more likely to contain enough information to answer the question. Balancing all factors, the overall source quality is not significantly different between the two treatments. Lastly, post-search questionnaires reveal that online search is more enjoyable than offline search.

Keywords: search, experiment
JEL Classification: C93, H41

*We would like to thank Jeffrey MacKie-Mason, Karen Markey, Betsy Masiello, and Hal Varian for helpful discussions, Donna Hayward and her colleagues at the University of Michigan Shapiro Library for facilitating the non-Web treatment, Ashlee Stratakis and Dan Stuart for excellent research assistance. The financial support from the National Science Foundation through grant no. SES-0720943 and a Google Research Award is gratefully acknowledged. School of Information, University of Michigan, 1075 Beal Avenue, Ann Arbor, MI 48109-2112. Email: yanchen@umich.edu, yjeon@umich.edu, kimym@umich.edu.

1 Introduction

With the advent of the Web and search engines, online search has become a common method of obtaining information. The question arises as to how much time people save by using search engines for their information needs, and the extent to which online search affects the quality of search outcomes and affective experiences.

Researchers from a number of different areas such as consumer studies, health communication, and library and information science (LIS) have examined use of online and offline channels for information seeking, where online search typically involves use of general search engines such as Google. The activities and populations studied include consumers shopping for goods (Zhang 2008) or seeking travel information (Bieger and Laesser 2004) and lay public seeking of health information (Cotten and Gupta 2004). In the LIS field, there have been several studies comparing Web searching to OPAC (online public access catalog) or library system searching including databases (e.g., Fast and Campbell (2004), Brophy and Bawden (2005), Griffiths and Brophy (2005)), but there has been no direct comparison of online and offline information searches. The typical methodology employed is user survey, with some studies employing observation and interviews of study participants.

Some of the questions examined in consumer studies are the information sources used for different types of shopping activities, including travel, as well as source preferences. In the case of travel information, information seeking occurs through online searching as well as offline searching through travel agents, guidebooks, and word of mouth (Pearce and Schott 2005). In shopping for consumer products, there was a preference for online channels when searching for information on these products such as price or product features (Zhang 2008). In the case of health information, people were most likely to turn to health care professionals for information. But, online health information seekers differed from individuals who search for information from online sources, in terms of age, income, and education (Cotten and Gupta 2004).

Studies of students engaged in academic research indicate that students prefer using the internet over the campus library (Jones 2002) and students turn to the internet first for academic research (Griffiths and Brophy 2005). In a qualitative study, Fast and Campbell (2004) found that while students acknowledged the OPAC as a trustworthy source of information and admired its organization, they still preferred using search engines, in this case Google. This preference may come from the perceived ease of use and immediate gratification offered by use of a search engine, compared to the library catalog. A more recent study (Head 2007) examining the academic research process, not just search, found that only about one in 10 students in the survey reported using a search engine first when conducting research. Brophy and Bawden (2005) compared the performance of Google to that of library database services along the four dimensions of quality, relevance, accessibility, and coverage. The two systems were found to be complementary, with Google providing superior coverage and accessibility, while library databases provided superior quality in terms of the documents retrieved.

In this paper, we present an experiment which compares the processes and outcomes of Web search in comparison with the more traditional information search using non-Web resources. Using a random sample of queries from a major search engine, we evaluate the amount of time participants spend when they use a search engine versus when they use the library without access to Web resources; any quality differences in the information found from the Web versus non-Web resources; and any differences in affective experiences between online and offline search processes. We have two treatments, a web and a non-web treatment. In the web treatment, participants can

use search engines and other web resources to search for the answers. In the non-web treatment, participants use the library without access to the Web. We then measure the search time, source quality and affective experience in each of the treatments.

Compared to previous studies where researchers generate search tasks, we use a random sample of queries from a major search engine, representing more realistic search tasks for information which people actually seek. Furthermore, previous studies use observations of search behavior or survey methods, while we use controlled experiments to compare search outcomes and experiences for the same set of questions between online and offline venues. Our approach enables us to obtain time estimates of online and offline search as well as the marginal likelihood of getting an answer over the same set of questions.

The rest of the paper is organized as follows. In Section 2, we present our experimental design. Section 3 presents the analysis and results. Section 4 concludes.

2 Experimental Design

Our experimental design draws from methods in information behavior and experimental economics. To directly compare online and offline search outcomes and experiences using realistic search tasks, we obtained a random sample of queries from a major search engine. Our experiment consists of four stages. First, queries are classified into four categories by three trained raters. The results enables us to eliminate queries which locate a web site or ask for web-specific resources. These queries are not suitable for comparisons between online and offline search. We also eliminated non-English language queries or queries locating adult material. Second, we ask three raters to convert the remaining queries into questions and provide an answer rubric for each question. In the third stage, using questions generated in the second stage, a total of 244 subjects participate in either the web or non-web treatment, randomly assigned as a searcher or an observer in each experimental session. Each searcher is assigned 5 questions, with incentivized payment based on search results. Lastly, answer quality is determined by trained raters, again using incentivized payment schemes. We now explain each of the four stages in detail below.

2.1 Query Classification

We obtained a random sample of 2515 queries from a major search engine, pulled from the United States on May 6, 2009. After eliminating duplicates, we ended up with a total number of 600 unique queries to be classified. Query classification was completed by three raters, who were two graduate students of the School of Information (SI), majoring in Information Science and one undergraduate student of University of Michigan. In particular, two graduate students were recruited from the pool of students who took a course, SI 665 (Online Searching and Databases) or SI 666 (Organization of Information Resources), which offer students a variety of skills and knowledge regarding information search. Thus, we could consider them as semi-professionals. We believe that having the mix of the graduate students and undergraduate student would allow us to have more valid classification result as we could obtain perspectives from both experts and a lay person.

The queries were classified into four categories: Factual, Source, Web, and Other. The intent of a Factual query is to obtain a brief factual answer to a question, while the intent of a Source query is to obtain a list of sources that can help answer the question. The Source category also includes

searching to find out more information on a topic such as in subject searching. The intent of a Web query is to locate a particular website, or obtain a resource on the Web, such as downloading a file. Specifically, this category includes navigational searching (searches for a particular website, including one the searcher believes exists) and transactional searching (searches to locate a website with the goal of carrying out transactions with that site, such as purchase of a product, executing an online application, or downloading files). Lastly, a query of which the intent does not match any of the three categories mentioned above, or cannot be ascertained is classified as Other. Queries in languages other than English, and queries for adult material, also were placed in this category.

The query classification was conducted in two steps. First, we provided each rater with 600 queries and instructions, and each rater classified them independently based on the given criteria. A facilitator compiled all the classified queries and reviewed the classification results. In reviewing the classification results, the facilitator used a majority rule to decide the classification of queries when all the raters failed to reach a consensus. After the review, we had 536 queries which were completely classified and 65 queries which all the raters classified differently. In the second step, a meeting was scheduled and all three raters and the facilitator gathered to discuss the classification results of the 65 queries which showed disagreement among the three raters. The facilitator was present to promote their discussion. For each query, three raters exchanged their opinions regarding the rationale for their classification, and arrived at a consensus classification by agreeing together or using the majority rule if necessary.

After completing the classification of 600 queries, we removed the queries which were classified as either Web or Other, as they were considered inappropriate to be converted into search questions for use in both Web and Non-Web conditions. Consequently, the total number of queries which could be converted into search questions was 356, 105 being Factual and 251 being Source.

2.2 Question Generation

As a result of query classification, we had 356 queries in total, 105 being Factual and 251 being Source. We converted these queries into questions for use in our experiment. Three SI graduate students were responsible for this task. All of the raters were selected among students who had taken either SI 665 or SI 666. We asked the raters to generate, for each query, a question and an answer rubric. The answer rubric is for evaluating the quality of answers provided by our subjects. We provided the raters with instructions for question generation, as well as examples for additional clarification. To make sure that our instructions made sense to all the raters and were interpreted consistently among them, we first asked each rater to generate a sample of ten questions and answers, which were reviewed by one of the authors. Then, each rater was assigned similar numbers of queries for generating questions and answer rubrics.

During the process of question generation, we emphasized to the raters to include both Web and non-Web options in developing the answer rubric, and to make clear from the question wording whether it was a Factual or Source type question. After compiling the questions generated by the three raters, an undergraduate research assistant reviewed questions to make sure that each question was intelligible, and one of the authors reviewed the answer rubrics. In reviewing questions and answer rubrics, reviewers could correct typos, as well as change the original classification of a query, if necessary. During the review process the following types of questions were discarded: questions about computer game cheat codes, software registration or activation codes, uploading/downloading files, and questions seeking information about individuals who are not public figures. Questions with no answer and questions that do not make sense as they were generated

from undecipherable queries were also removed from our question pool. As a result, we had a set of 330 questions that can be used for online and offline search. Of these, 25 queries were used for pilot sessions to test the experiment instructions and protocols, while 305 were used for the formal experimental treatments.

2.3 Experimental Procedure

We use a two-treatment between-subject design, including a web search treatment, and a non-web search treatment. Each subject participated in only one session. Each search session can last up to three hours to avoid fatigue. Subjects in each session first complete a pre-experiment survey (Appendix A), then participate in the search experiment, and then finish a post-experiment survey (Appendix D).

The non-web sessions were conducted at the Hatcher Library at the University of Michigan. Each session consisted of 10-16 subjects, half of whom were designated as searchers and the other half as observers. Subjects are paired into searcher-observer pairs. After experiment instructions are read aloud, each observer in the searcher-observer pair receives five questions and observation forms. The observer then hands the searcher one question. Each searcher-observer pair proceeds to the reference room or library stacks to search for answers to these questions. A searcher logs his or her own search, including call numbers and information in the materials examined, while an observer logs their observations of the searcher's actions. After the completion of each search, the searcher completes a post-search questionnaire (Appendix C)[1] on specific aspects of their search. The observer hands the searcher the next question when this is completed. A searcher can consult the reference librarian up to two times during a session. Instructions for the non-web treatment is included in Appendix B-1.

The web search treatment was conducted at the Behavioral and Experimental Economics Lab at the University of Michigan. Each experimental session consisted of 8-20 participants, randomly matched into searcher-observer pairs. Each searcher is assigned five questions, and can use search engines or other resources on the Web to search for answers. Observers again log the time and search activities. Unlike the non-web treatment, however, we use Morae to record the precise mouse movements and clicks of each searcher. Even though observers are technically not needed, we decide to include them so that the process is comparable across the two treatments. The Morae data also enable us to evaluate the accuracy of the observer logs. Instructions for the web treatment is included in Appendix B-2.

After completing the search tasks, all subjects record their search results and process on a computer, and fill out a demographics and post-experiment survey. While observers are paid a flat rate, searchers are paid based on the quality of their answers.

A total of 244 subjects participated in our experiment, half were randomly assigned a role as searchers while the other half as observers.

2.4 Rating Procedure for Answer Quality

Out of 305 questions used for the formal experimental treatments, 275 questions receive answers in both the non-web and web treatments. To evaluate the source quality of these answers, we recruit 12 raters from the masters students in the School of Information specializing in Library

[1] Questions on the affective aspects of the search process were derived from White, Jose and Ruthven (2003).

and Information Services (LIS) who had taken the course, SI 647 (Information Resources and Services) or SI 665 (Online Searching and Databases)[2] in the past two years. Specifically, a total of 550 question-answer pairs are randomly divided into four sets, each consisting of approximately 137-138 question-answer pairs. We decide to divide these question-answer pairs into four sets to avoid raters' fatigue. Three raters are assigned to each set. Thus, each question-answer pair is rated by three independent raters, and we use the median rating for each answer in our analysis. The first rating sessions uncover that sources are not recorded in 60 of the non-web questions. We then turn to the original hardcopies of search record and find recorded sources for 49 out of the 60 answers. These 49 questions form question Set 5, each of which rated by 3 independent raters. In sum, question sets 1-4 have a mixture of Web and Non-Web answers, and each group rated 110 to 115 question-answer pairs after excluding answers originally marked as having no source, whereas question set 5 consists of the 49 recovered Non-Web question-answer pairs.

Six common evaluation criteria of source quality are applied for each question-answer pair, including trustworthiness, authoritativeness, relevance, accuracy, amount of information, and overall source quality. In addition, different evaluation criteria are applied to factual and source questions. For example, correctness of the answer is evaluated for factual questions, but not for source questions, which often require evaluation and synthesis of multiple information sources without necessarily a definitive correct answer. Instructions to raters are included in Appendix E.

Table 1: Inter-rater reliability using intraclass correlation coefficient: ICC[3,3]

Evaluation Criteria	Set 1	Set 2	Set 3	Set 4	Set 5
Q1. Question difficulty	0.695	0.643	0.72	0.726	0.786
Q2a. Source trustworthiness	0.717	0.816	0.609	0.692	0.523
Q2b. Source authoritativeness	0.698	0.789	0.659	0.697	0.575
Q2c. Source relevance	0.698	0.691	0.519	0.62	0.522
Q2d. Information accuracy	0.538	0.769	0.536	0.441	0.370
Q2e. Enough information	0.663	0.72	0.714	0.636	0.514
Q3. Overall source quality	0.645	0.774	0.640	0.699	0.320

Table 1 reports the intraclass correlation coefficient (ICC[3,3]) as a measure of inter-rater reliability.[3] In general, values above 0.75 represent excellent reliability, values between 0.40 and 0.75 represent fair to good reliability, and values below 0.40 represent poor reliability. Good to excellent reliability is observed for all but two of our ratings.

[2]Both courses prepare them for reference services in settings such as libraries or other information centers, requiring students to work with actual reference questions. Students gain expertise in searching, evaluating information resources, and answering questions submitted online.

[3]There are six main cases of intraclass correlation coefficients (ICC), distinguished by the numbers in parentheses following the letters ICC. The first number indicates the statistical model assumed. Case 3 assumes that judges are fixed and not drawn from a random population. This statistic is appropriate when the research design involves averaging ratings across all raters. The second number indicates the number of raters. More details on ICC computation can be found in Shrout and Fleiss (1979).

3 Results

In this section, before presenting each result, we formulate our alternative hypotheses based on prior empirical research and our own experience. Our general null hypothesis is that outcome does not differ between the two treatments. In our analysis, we use a 5% statistical significance level as our threshold to establish the significance of an effect.

We first compare the likelihood that searchers find an answer when searching with versus without access to the web.

Hypothesis 1. *Searchers are more likely to find an answer using web search.*

While Hypothesis 1 *per se* might not be surprising, we hope to be able to pin down precisely how much more likely a searcher is able to find answers using the web.

Result 1 (Likelihood of Success)**.** *Searchers are significantly more likely to find an answer using web search. Of the 305 questions, 99.7% are answered in the web treatment, while 90.2% are answered in the non-web treatment.*

Support. *A paired t-test of the null hypothesis of equal likelihood against Hypothesis 1 yields* $p < 0.01$ *(one-sided).*

By Result 1, we reject the null in favor of Hypothesis 1 that searchers are more likely to find an answer using the web.

We now compare the search outcome for the 275 questions which the searchers in both treatments find answers to. We are interested in two outcome measures. The first measure is the time it takes to search for an answer. The second measure is the quality of the information source.

Hypothesis 2. *Web search will take less time than non-web search.*

Again, while Hypothesis 2 is not surprising, our study will be able to estimate the precise time it takes to find an answer for a randomly selected query, and thus provide an answer to how much time web search saves on average.

Result 2 (Time)**.** *Online search takes significantly less time than offline search. Online search takes on average 1/3 of the time of offline search.*

Support. *A paired t-test of the null hypothesis of equal time against Hypothesis 2 yields* $p < 0.01$ *(one-sided).*

For the 275 questions with answers in both treatments, the average search time to find an answer is 22 minutes ($\sigma = 14.0$) in the non-web treatment, and 7 minutes ($\sigma = 4.7$) in the web treatment. Again the difference is highly significant. By Result 2, we reject the null in favor of Hypothesis 2 that web search takes less time non-web search. Furthermore, in our sample, web search uses 1/3 of the time to find an answer compared to non-web search.

We next examine source quantity and quality based on the evaluation of independent raters. To determine source quantity, for the non-web condition, we asked raters to review a searcher's answer and a corresponding observer's log, and record the following for each question-answer pair: (1) sources a searcher looked at and their order, (2) whether a searcher consulted a reference librarian, (3) total number of sources a searcher looked at, (4) number of source(s) actually used by a searcher in her answer, and (5) the reasons a searcher failed to answer a question inferred from

the observation log. We also asked raters to describe any noticeable characteristics or pattern of search process for each searcher. The raters were given an initial list of ten source categories for use, including Mirlyn (UM online catalog), reference librarian, ready reference, books, journals, magazines/newspaper, government documents, telephone directory, people, and other.

In comparison, for the web condition, we asked raters to review Morae files of each searcher and record the following for each question-answer pair: (1) queries a searcher typed in, (2) whether a searcher used advanced search features, (3) position of each link clicked on the first search results page, (4) type of each link clicked from the first search results page (e.g., organic, sponsored, and others), (5) number of links a searcher clicked from the first search results page, (6) whether a searcher went beyond the first search results page, (7) number of links a searcher clicked from other search result pages, (8) whether a searcher returned to previous search results page, (9) whether queries a searcher typed in is exactly same as an original query, (10) whether a searcher reformulate the query or entered a new query, (11) the total number of queries a search typed in, (12) whether a searcher changed a search engine used, (13) total number of links a searcher clicked, (14) number of links actually used by a searcher in her answer. We also asked our raters to describe any noticeable characteristics or pattern of search process for each searcher.

Hypothesis 3. *Source quantity is comparable across the two treatments.*

Result 3 (Source Quantity)**.** *Searchers under the web condition looked at and used more sources than those under the non-web condition.*

Support. *On average, a searcher under the web condition clicked 3.9 links during a search to answer a given question, and actually used 1.6 sources in her answer, while a searcher under the non-web condition looked at 3.1 sources and used 1.3 sources in her answer. The difference in the total number of sources looked at ($p < 0.01$, one-sided t-test) and total number of sources actually used between non-web and web conditions are statistically significant ($p < 0.01$, one-sided t-test).*

The top three sources a searcher under the non-web condition first looked at are Mirlyn (72%), ready reference (13%), and telephone directory (9%). Additionally, 70% of searchers sought help from reference librarians. The searchers under the web condition reformulated the query or entered new query in 62% of their searches. On average, a searcher typed in three queries to answer a question. They typed in the same query as the original one in 22% of their search.

For source quality, recall each answer is independently rated by three MSI students. Using the intraclass correlation coefficient to measure inter-rater reliability, we find that that our raters achieve good to excellent levels of inter-rater reliability (Table 1). We thus use the median rating of the three raters for our analysis of source quality. The issue of source quality is largely unresolved when comparing web and non-web search. Thus, we do not have a strong prior belief about the alternative hypothesis.

Hypothesis 4. *Source quality is comparable across the two treatments.*

Result 4 (Source Quality)**.** *While non-web sources are judged to be significantly more trustworthy and authoritative than the corresponding web sources, web sources are judged to be significantly more relevant and more likely to contain enough information to answer the question. Furthermore, the accuracy of answer source is not significantly different across the two treatments. Balancing all factors, the overall source quality is not significantly different across the two venues.*

Support. *Wilcoxon signed-rank tests yields the following results: (1) trustworthiness:* $p < 0.01$ *(one-sided); (2) authoratativeness:* $p < 0.01$ *(one-sided); (3) relevance:* $p < 0.01$ *(one-sided); (4) accuracy:* $p = 0.824$ *(two-sided); (5) enough information:* $p < 0.01$ *(one-sided); and (6) overall quality:* $p = 0.618$ *(two-sided).*

Lastly, we examine the subjective and, in particular, the affective experience of search using post-search questionnaires. The questionnaire and the summary statistics is included in Appendix C. We first note that searchers' self-evaluation of their familiarity with the question topics is not different across the two treatments ($p = 0.53$, two-sided Wilcoxon signed-rank test for matched samples). This indicates that searchers in the two treatments are equally familiar with backgrounds required to answer the questions. Thus, any difference in affective experience should come from the experimental treatments.

Hypothesis 5. *Web search yields more positive affective experience than non-web search.*

Result 5 (Affect)**.** *Web search yields significantly more positive affective experiences than non-web search.*

Support. *Each two-sided Wilcoxon signed-rank test for matched samples yields* $p < 0.01$.

After searching for answers for each question, searchers in the web treatment find the question significantly more interesting, the process of finding information significantly less difficult, the materials found during the search process significantly more useful, and the search process itself significantly less stressful, more enjoyable, and less tiring compared to their non-web counterparts. Finally, searchers using web search are significantly more confident that their answers are correct. By Result 5, we reject the null in favor of Hypothesis 5 that web search yields more positive affective experience.

4 Conclusions

With the advent of the Web and search engines, online search has become a common method of obtaining information. The question arises as to how much time people save by using search engines for their information needs, and the extent to which online search affects search experiences and outcomes.

Using a random sample of queries from a major search engine, we conduct an experiment to compare online and offline search experiences and outcomes. We find that participants are significantly more likely to find an answer on the Web. Restricting to the set of queries which participants find answers in both treatments, the average search time is 22 minutes offline, and 7 minute online. While non-web sources are judged to be significantly more trustworthy and authoritative than the corresponding web sources, web sources are judged to be significantly more relevant and more likely to contain enough information to answer the question. Balancing all these aspects, our raters' overall evaluation of source quality is not significantly different across the two venues. Furthermore, post-search questionnaires reveal that online search is more enjoyable than offline search.

Even though we pre-select questions which are judged to be searchable both online and offline, our results indicate that online search yields higher likelihood of search success, takes on average one-third of the time as the corresponding offline search, and is more enjoyable. The results on

source quality though is more nuanced. Non-web sources from an academic library is judged to be more trustworthy and authoritative, while web sources are judged to be more relevant and more likely to contain enough information to answer the questions. This result point to the importance of credibility for online sources.

Lastly, since all our queries were drawn from actual questions asked on the web, it might not be too surprising that the searchers were more likely to find answers there. In a follow-up study, we would like to look at questions that users ask in a library to investigate whether there is a bias for non-web search.

Appendix A. Pre-Experiment Survey

Please answer the following questions. Your answers will be used for this study only. Individual data will not be exposed.

1. Please enter today's date.
2. Please enter your Participant ID.
3. How often do you search for information for the following purposes? Please check all that apply.

	Everyday	3-4 times per week	1-2 times per week	Less than 1 time per week	Never
Work					
School					
Personal					
Other					

4. How often do you use search engines to help you find information on the Web?
   - Everyday
   - 3-4 times per week
   - 1-2 times per week
   - Less than 1 time per week
   - Never
5. Which search engines have you used? Please check all that apply.
   - Ask.com
   - Bing (formerly MSN Search)
   - Google
   - Yahoo!
   - Other, please specify:
6. Please indicate which search engine you use most often.
   - Ask.com
   - Bing (formerly MSN Search)
   - Google
   - Yahoo!
   - Other, please specify.
7. Please rate your ability to find information using a search engine (1: Poor, . . . 5: Excellent).

   1

   2

3

4

5

8. How often do you use libraries to find information?

    - Everyday
    - 3-4 times per week
    - 1-2 times per week
    - Less than 1 time per week
    - Never

9. Which libraries have you visited? Please check all that apply.

    - Ann Arbor District Library
    - Art, Architecture & Engineering Library
    - Harlan Hatcher Graduate Library
    - Shapiro Undergraduate Library
    - Other, please specify

10. Please indicate which library you use most often.

    - Ann Arbor District Library
    - Art, Architecture & Engineering Library
    - Harlan Hatcher Graduate Library
    - Shapiro Undergraduate Library
    - Other, please specify

11. Please rate your ability to find information using a library. (1: Poor, . . . 5: Excellent).

    1

    2

    3

    4

    5

Appendix B-1. Experimental Instructions: Non-web Treatment

Participant instructions

This experiment studies how people look for information when they do not have access to information resources on the Web. Participants are divided into two groups, which are "Searcher" and "Observer". In this experiment, those of you sitting in the first row are Searchers. Those of you in the second row are Observers. Each Searcher is paired with an Observer, and vice versa - the person sitting directly behind a Searcher is their assigned Observer, and the person sitting directly in front of an Observer is their assigned Searcher. Please confirm who is your Searcher or Observer now.

Searcher

The Searcher is expected to answer five questions, using print resources in the library. Searchers are not allowed to use search engines such as Google or other Internet resources. Imagine you are trying to answer your questions in the days before the Web. The resources you are allowed to use are Mirlyn (the library's online catalog), and materials in the stacks. You can consult a reference librarian for only two of your five questions. To help you, we have included guides to the Hatcher Graduate Library. We will also provide you with a brief instruction on using Hatcher.

First, here is the procedure you, the Searcher, will follow in this experiment:

1. You have received a clipboard labeled "Searcher". Please check it has all of the following: participant instructions, map of Hatcher 2nd floor, map of call numbers, list of allowed resources, payment form, and a Participant ID.

2. You will complete a pre-survey after we finish going over these instructions.

3. After completing the pre-survey, you will be guided to the library by one of the experimenters.

4. The experimenter will provide a brief overview of resources available in the library.

5. Once the experimenter instructs you to start, the Observer will hand you a question and answer form.

6. Please answer this question to the best of your ability, using the allowed resources.

7. The Observer will be logging your start and end times for the task, as well as the actions you take to answer the assigned question.

8. When you have answered the question on the answer form, please complete a post-search questionnaire which is stapled to the answer form, and return it to your Observer.

9. You will receive the next question and answer form from the Observer. Please follow the same procedure you followed for the first question for the remaining four questions.

10. When you have answered all the questions and returned your last question and answer form to the Observer, please return with your Observer to the 2nd Floor Hatcher computer area.

11. The Observer will give you all the answer forms you completed.

12. Please take a seat and log on to a computer with your uniqname, and open a browser and go to the following link and type in your answer to each question with Question ID and submit it. `http://tinyurl.com/lxwr6f`

13. Once you have submitted your answers, please go to the following link and complete the post-survey. `http://tinyurl.com/nlkqbx`

14. Please go to the waiting table and fill out the subject payment form.

15. Please wait until your ID number is called to receive your earnings. Return the completed subject payment form and your clipboard to the experimenter at this time.

16. After receiving payment, you may leave the study session.

Your payment will be based on the number of questions you answer and the quality of the answers you provide. If you determine you cannot answer a question, you may move on to the next one. Once you move on to the next question, you cannot return to a question you worked on previously.

Observer

The Observer is expected to follow their assigned Searcher and observe and record their observations of the Searcher's activities. The Observer also has to perform a number of additional tasks.

Here is the procedure you, the Observer, will follow in this experiment:

1. You have received a clipboard labeled "Observer". Please check it has all of the following: participant instructions, map of the 2nd floor, map of call numbers, list of allowed resources, 5 sets of questions and answer forms and post-search questionnaires, 5 log sheets, a payment form, and a Participant ID.

2. You will complete a pre-survey after we finish going over these instructions.

3. After completing the pre-survey, you will be guided to the library by one of the experimenters.

4. The experimenter will provide a brief overview of resources available in the library.

5. Once in the library, write down the Question ID of the first question on your log sheet and hand your Searcher the first question and answer form. Log on the log sheet the time this was handed out.

6. As the Searcher attempts to answer the question, log their activities on the log sheet (e.g. searching on Mirlyn, speaking with a reference librarian).

7. When the Searcher returns the question and answer form and post-search questionnaire, please log this time as the end time for this search on the log sheet. Please check the Searcher has written their Participant ID and Question ID in the answer form.

8. Hand the Searcher the next question and answer form. Please follow the same procedure you followed for the first question for the remaining four questions.

9. When the Searcher has completed answering all questions and returned the last question and answer form and post-search questionnaire, please return to the 2nd Floor Hatcher computer area with your Searcher.

10. Please return all the answer forms you received from the Searcher back to him or her.

11. Please take a seat and log on to a computer with your uniqname, and open a browser and go to the following link and type in your log for each search task with Question ID and submit it. `http://tinyurl.com/mkby5a`

12. Once you have submitted your logs, please go to the following link and complete the post-survey. `http://tinyurl.com/nlkqbx`

13. After completing the post-survey, please go to the waiting table and fill out the subject payment form.

14. Please wait until your ID number is called to receive your earnings. Return the completed subject payment form and your clipboard to the experimenter at this time.

15. After receiving payment, you may leave the study session.

Payment

As mentioned earlier, participants are assigned to the role of either a Searcher or Observer. There are two different payment schemes for each group. Both Searchers and Observers receive $5 for showing up. A searcher receives 1 point for each question answered, and 1 point for the quality of the answer, for a total of up to two points for each answer to a question. There is an answer rubric for the questions, and an experimenter will evaluate answers based on the rubric. Points will be converted into a dollar-value based on our conversion rate. The conversion rate for this experiment is 1 pt = $2. The Observer will receive a flat fee of $10 per hour. An individual Observer's actual earnings will depend on the actual time it takes to participate in the experiment. Each participant's earnings will be calculated based on these two schemes. Please note that payoff of Searcher and payoff of Observer will be determined independently. Payment will be made individually in the ord! er of completion of assigned tasks.

Are there any questions? Please make sure that your cell phone is on silent mode. Keep in mind that you are not allowed to communicate with other study participants during the experiment, other than with your assigned Searcher or Observer and only for handoff of forms. You may not help each other in the searches. If there are no further questions, let's begin with the pre-survey. Please log on to a computer with your uniqname, open a browser and complete the pre-survey at this URL: `http://tinyurl.com/ksaojw`

Once you have finished the pre-survey, please log out and wait to be guided to the Graduate Library.

Appendix B-2. Experimental Instructions: Web Treatment

This experiment studies how people look for information on the Web. Participants are divided into two groups, which are "Searcher" and "Observer". A Searcher and an Observer are randomly assigned and paired.

Searcher

The Searcher is expected to answer five questions via a web search. The Searcher's entire search process will be recorded by Morae, a software program.

First, here is the procedure you, the Searcher, will follow in this experiment:

1. You have received a clipboard labeled "Searcher". Please check it has all of the following: participant instructions, a payment form, and a Participant ID.
2. You will complete a pre-survey after we finish going over these instructions.
3. After completing the pre-survey, please go to the following link to open an answer form. You will see the answer form to fill out when answering a question. Please keep this window open throughout the session. `http://tinyurl.com/meky7b`
4. Please raise your hand and let an experimenter know you opened the answer form. The experimenter will start Morae on your computer.
5. The Observer will hand you a question sheet.
6. Please answer this question to the best of your ability, using a web search.
7. The Observer will be logging your start and end times for the task, as well as the actions you take to answer the assigned question.
8. Once you have answered the question on the answer form, please complete a post-search questionnaire which is stapled to the question sheet, and return it to your Observer.
9. You will receive the next question sheet from the Observer. Please follow the same procedure you followed for the first question for the remaining four questions.
10. Once you have answered your last question, please click the "Submit" button on the answer form.
11. Please complete the last post-search questionnaire and return it to the Observer.
12. Please wait until the experimenter stops the Morae recording, as the Observer let the experimenter know you are done by raising his or her hand.
13. Please go to the following link and complete the post-survey. `http://tinyurl.com/ya542hp`
14. After completing the post-survey, please fill out the subject payment form.
15. Please wait until your ID number is called to receive your earnings. Return the completed subject payment form and your clipboard to the experimenter at this time.

16. After receiving payment, you may leave the study session.

Your payment will be based on the number of questions you answer and the quality of the answers you provide. If you determine you cannot answer a question, you may move on to the next one. Once you move on to the next question, you cannot return to a question you worked on previously.

Observer

The Observer is expected to observe and record their observations of the Searcher's activities during his or her search process. The Observer also has to perform a number of additional tasks.

Here is the procedure you, the Observer, will follow in this experiment:

1. You have received a clipboard labeled "Observer". Please check it has all of the following: participant instructions, 5 sets of question sheets and post-search questionnaires, 5 log sheets, a payment form, and a Participant ID.

2. You will complete a pre-survey after we finish going over these instructions.

3. After completing the pre-survey, please move your chair to the position where you can observe your Searcher's search process.

4. Please wait until your Searcher opens an answer form on his or her computer, and an experimenter starts Morae on your Searcher's computer.

5. After Morae is started, please write down the Question ID of the first question on your log sheet and hand your Searcher the first question sheet. Log on the log sheet the time this was handed out.

6. As the Searcher works on the question, log your observations of their search process on the log sheet (e.g. time, browser used, search engine used, queries typed in).

7. When the Searcher returns the question sheet and post-search questionnaire, please log this time as the end time for this search on the log sheet.

8. Please write down the Question ID of the next question and hand the Searcher the next question sheet. Please follow the same procedure you followed for the first question for the remaining four questions.

9. When the Searcher has completed answering all questions and returned the last question sheet and post-search questionnaire, log the time as the end time and raise your hand to let the experimenter know the Searcher is done.

10. The experimenter will stop the Morae recording.

11. Please return to the computer you used for a pre-survey and take a seat.

12. Please open a browser and go to the following link, and type in your log for each search task with Question ID and submit it. `http://tinyurl.com/kprfjy`

13. Once you have submitted your log, please go to the following link and complete the post-survey. `http://tinyurl.com/ya542hp`

14. After completing the post-survey, please fill out the subject payment form.

15. Please wait until your ID number is called to receive your earnings. Return the completed subject payment form and your clipboard to the experimenter at this time.

16. After receiving payment, you may leave the study session.

Payment

As mentioned earlier, participants are assigned to the role of either a Searcher or Observer. There are two different payment schemes for each group. Both Searchers and Observers receive $5 for showing up. A searcher receives 1 point for each question answered, and 1 point for the quality of the answer, for a total of up to two points for each answer to a question. There is an answer rubric for the questions, and an experimenter will evaluate answers based on the rubric. Points will be converted into a dollar-value based on our conversion rate. The conversion rate for this experiment is 1 pt = $2. The Observer receives a flat fee of $10 per hour. An individual Observer's actual earnings will depend on the actual time it takes to participate in the experiment. Each participant's earnings will be calculated based on these two schemes. Please note that payoff of Searcher and payoff of Observer will be determined independently. Payment will be made individually in the order o! f completion of assigned tasks.

Are there any questions? Please make sure that your cell phone is on silent mode. Keep in mind that you are not allowed to communicate with other study participants during the experiment, other than with your assigned Searcher or Observer, and only for handoff of forms. You may not help each other in the searches. If there are no further questions, let's begin the pre-survey. Please open a browser and go to the following link, and complete the pre-survey. `http://tinyurl.com/yelu74j`

Once you have finished the pre-survey, Searcher should open the answer form (please refer to Searcher procedure #3) and Observer should wait until Searcher opens the answer form (please refer to Observer procedure #3). You may start the search and records the search process after the experimenter starts Morae on your computer.

Appendix C. Post-Search Questionnaire

(*Mean and standard deviation of responses for each treatment is included in italics.*)

1. Please rate how familiar you are with the question topic. (1: not at all familiar)

   Non-web: $\mu = 1.88$, $\sigma = 1.32$*;*

   Web: $\mu = 2.06$, $\sigma = 1.29$.

2. Please rate how interesting you found this search question. (1: not at all interesting)

   Non-web: $\mu = 2.11$, $\sigma = 1.36$*;*

   Web: $\mu = 2.69$, $\sigma = 1.29$.

3. Please rate how difficult it was to find information in answering the question. (1: not at all difficult)

   Non-web: $\mu = 2.96$, $\sigma = 1.62$*;*

   Web: $\mu = 2.29$, $\sigma = 1.31$.

   Note: correlation with rater estimation of difficulty of questions: $\rho = 0.09$ ($p = 0.11$) for non-web, $\rho = 0.36$ ($p = 0.00$) for web)

4. Please rate how useful the materials were that you found during the search for answering the question. (1: not at all useful)

   Non-web: $\mu = 2.92$, $\sigma = 1.58$*;*

   Web: $\mu = 3.79$, $\sigma = 1.17$.

5. The search process to answer the question was

   (a) stressful (1: stressful, 5: relaxing)
   Non-web: $\mu = 2.419$, $\sigma = 1.277$*;*
   Web: $\mu = 3.347$, $\sigma = 1.07$.

   (b) unenjoyable (1: unenjoyable, 5: enjoyable)
   Non-web: $\mu = 2.20$, $\sigma = 1.24$*;*
   Web: $\mu = 3.10$, $\sigma = 1.17$.

   (c) tiring (1: tiring, 5: restful)
   Non-web: $\mu = 2.42$, $\sigma = 1.28$*;*
   Web: $\mu = 3.35$, $\sigma = 1.07$.

6. Please rate your confidence in your answer to the question. (1: not at all confident)

   Non-web: $\mu = 3.23$, $\sigma = 1.66$*;*

   Web: $\mu = 4.03$, $\sigma = 1.09$.

Appendix D. Post-Experiment Survey

Please answer the following questions. Your answers will be used for this study only. Individual data will not be exposed.

1. Please enter today's date.
2. Please enter your Participant ID.
3. What is your age?
4. What is your gender?
    - Female
    - Male
5. Please indicate your race/ethnicity.
    - American Indian/ Alaska Native
    - Asian/ Pacific Islander
    - Black
    - Hispanic
    - White
    - Other, please specify
    - I do not wish to disclose it.
6. What is your major at the University of Michigan?
7. What year are you in your program?
    - Freshman
    - Sophomore
    - Junior
    - Senior
    - Other, please specify
8. Before today, have you ever used the following advanced features of search engines? Please check all that apply.
    - Phrase search by putting quotes around a set of words
    - Search within a specific web site
    - Include all of the words you typed in the search box by inserting "AND" or "+"
    - Searches for results that match either one or more of the words by inserting "OR"
    - Exclude words from your search by inserting "NOT" or "-"
    - Fill in the blanks by using "*"

- Restrict your search to sites with specific domain suffix
- Restrict your search by a specific file type
- Limit your results to sites that have been updated within a certain amount of time period
- None of the above

9. Please rate your ability to find information using a search engine. (1: Poor, . . . 5: Excellent).

10. Before today, have you ever used the following resources available at a library? Please check all that apply.

    - Document center for government information
    - Mirlyn
    - Reference librarian
    - Serial services (e.g. periodicals, newspapers, microforms)
    - Special collections
    - Subject specialist
    - None of the above

11. Please rate your ability to find information using a library.
    (1: Poor, . . . 5: Excellent).

12. Where do you have Internet access? Please check all that apply.

    - At home
    - At school
    - At work
    - At a coffee shop/ fast-food restaurant/ bookstore
    - Other places, please specify

13. How much do you pay for Internet access at Home? Please check all that apply.

    - I pay $ __ per month.
    - It's free.
    - My parents pay for it.
    - I don't know.
    - Other places, please specify

14. If your answer to the previous question (Question 13) is "I pay $__ per month.", please enter how much you pay per month. If not, please enter "0".

15. Hypothetically, if you do not have Internet access and you would like to search for information such as what you searched for today. How much would you be willing to pay for one day (24 hours) of Internet access?

16. Hypothetically, if you have Internet access and we paid you to give up Internet access, how much money would we have to pay you to give up Internet access for one day?

Appendix E. Rater Training Session Instructions

You are now taking part in a study that compares online and offline search experiences and outcomes. Your participation will take the form of rating questions and corresponding answer source(s) on several factors. There are two types of questions: Factual questions and Source questions. The intent of a Factual question is to obtain a brief factual answer to a question. The intent of a Source question is to obtain a list of sources that can help answer the question or to find out more information on a topic.

You will initially take part in an in-person training session. After the training session, you will rate question-answer pairs using a web-based system, over the following two weeks. We ask you not to discuss your rating activities with others during the course of the two weeks. Should you have any questions during the course of rating, please email us at yjeon@umich.edu.

The purpose of this training session is to familiarize you with the rating procedure, and to ensure a common understanding of the rating factors. We are interested in eliciting objective ratings from impartial raters. We ask you to think how a trained information professional would rate these question-answer pairs when following the rating procedure.

In this training session, you will be asked to rate two question-answer pairs, one with a Factual question, the other a Source question. You will also be asked to fill out a Background Questionnaire at the end of this training session. For each question-answer pair, you will be asked for the ratings shown below:

1. Please rate the difficulty of the question. (1 = very easy . . . 5 = very difficult)
2. Please rate the answer source(s) for the following factors
   (1=strongly disagree . . . 5 = strongly agree)
   (a) The source is trustworthy.
   (b) The source is authoritative.
   (c) The information in the source is relevant to the question.
   (d) The information in the source is accurate.
   (e) There is enough information in the source to answer the question.
3. Please rate the overall quality of the answer source
   (1=very low quality ... 5=very high quality)
4. If a question is factual, please rate the followings:
   (a) Is the answer correct?
      i. Yes
      ii. No
      iii. Don't know
   (b) Please rate your confidence about your judgment (1=not confident at all ... 5=very confident):

The procedure in the training session is as follows:

1. You will each receive two rating sheets, a Background Questionnaire, and your rater ID.
2. You will rate the first question-answer pair, working individually. Please write your rater ID on the rating sheet for Question 1, and enter your ratings on the rating sheet.
3. Please let us know when you have finished rating the first question-answer pair. We will wait until all the raters have completed rating. Do not proceed to the second pair.
4. When all raters have completed rating the first question-answer pair, there will be a brief discussion, no longer than 10 minutes, regarding the rating activity. We will go over each rating, asking all of you for your ratings. You may ask us questions at any time.
5. When all questions have been addressed, we will move on to the second question-answer pair, and repeat the procedure used for the first pair.
6. Please fill out the Background Questionnaire given to you.
7. Please return the completed rating sheets and questionnaire to us.
8. After everyone has returned the completed rating sheets to us, we will use the remaining time to go to the web interface and start the formal rating sessions. This way, if you have any questions, you can ask us.

Are there any questions? We would like to ask you to please take the time to read each question and answer pair carefully when rating.

Rating Session Instructions

The actual ratings will be done using a web-based system. The system has been programmed to show you a predefined number of question-answer pairs for rating per login session. You will rate between 20 - 25 question-answer pairs per session. Once you have rated a predefined number of pairs you will be automatically logged out. If you have to quit your session before rating all the pairs, simply close the browser window.

For a rating session, you will get paid at the rate of $15 per hour. In addition to this base rate, you will receive a bonus of $1 per question-answer, for up to 50 randomly selected question-answer pairs, if the following condition is met: for each question-answer pair in the set of 50, if your rating for overall source quality for a question-answer pair is at the median for that question-answer pair, you will receive a bonus of $1.

Instructions for rating on the web:

To start your rating session, please go to the following web page: `http://yjeon.people.si.umich.edu/login.php`

We recommend using the Firefox browser, as Internet Explorer does not display in a consistent format.

Now follow these steps:

1. Login using the name (Rater ID) and password given to you.
2. Provide the ratings requested for the question-answer pair.
3. When you are done, click 'Submit' - note that you will not be permitted to continue until you have entered all nine ratings.

4. The next question-answer pair will be presented.

5. When you have finished rating the session limit, you will be shown a 'That is all your questions for this session. Thank you.' Screen.

6. Close the browser.

References

Bieger, Thomas and Christian Laesser, "Information Sources for Travel Decisions: Toward a Source Process Model," *Journal of Travel Research*, 2004, *42*, 357–371.

Brophy, Jan and David Bawden, "Is Google enough? Comparison of an internet search engine with academic library resources," *Aslib Proceedings: New Information Perspectives*, 2005, *57* (6), 498–512.

Cotten, Shelia R. and Sipi S. Gupta, "Characteristics of online and offline health information seekers and factors that discriminate between them," *Social Science & Medicine*, 2004, *59*, 1795–1806.

Fast, Karl V. and Grant Campbell, ""I Still Like Google": University Student Perceptions of Searching OPACs and the Web," in "Proceedings of the 67th ASIS&T Annual Meeting," Vol. 41 2004, pp. 138–146.

Griffiths, Jillian R. and Peter Brophy, "Student searching behaviour and the Web: use of academic resources and Google," *Library Trends*, 2005, *53* (4), 539–554.

Head, Alison J., "Beyond Google: How do students conduct academic research?," *First Monday*, 2007, *12* (8).

Jones, Steve, "The Internet Goes to College," Technical Report, Pew Internet & American Life Project 2002.

Pearce, Douglas G. and Christian Schott, "Tourism Distribution Channels: The Visitors' Perspective," *Journal of Travel Research*, 2005, *44*, 50–63.

Shrout, Patrick E. and Joseph L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin*, 1979, *86* (2), 420–428.

White, Ryen W., Joemon M. Jose, and Ian Ruthven, "A task-oriented study on the influencing effects of query-biased summarisation in web searching," *Information Processing & Management*, 2003, *39* (5), 707–733.

Zhang, Aimao, "Examining product and process effects on consumer preferences for online and offline channels," *Business Process Management*, 2008, *14* (1), 85–95.