Exploring the Effect of Mental Demand in Web Searches: A Pilot Study

Kyoungsik Na (나경식)*

Contents

ABSTRACT

This pilot study explored the effect of mental demand on a Web searcher's thoughts, emotions, efforts, and performance in Web searches in order to address whether or not there is any difference between searchers exposed to mental demand manipulation and searchers not exposed. Research data were collected via think-aloud protocol (TAP) with a dual-task in experiments and interviews with 10 subjects who participated in this study. For the searcher's thoughts, relevance judgment was found to be hindered by mental demand. For the searcher's emotions, the experimental group was more frustrated than the control group. With respect to the searcher's efforts, searchers for the experimental group with mental demand manipulation were more likely to spend more time, make fewer queries, and visit fewer pages but work harder to find more relevant information that they needed. Lastly, with regard to the searcher's performance, it is likely that performance was highly dependent upon the completion of the search tasks for both groups. The NASA-TLX six components and cognitive load scores of searchers did not make a significant difference in the outcome. The findings support the use of a dual-task methodology as a promising approach for the assessment of cognitive load induced by complex Web searches.

Keywords: Mental Demand, Cognitive Load, Web Search, NASA-TLX, Dual-Tasks, Think Aloud Protocol

1. Introduction

1.1 Background and Motivation

As a great amount of information is growing and a variety of information systems and resources are becoming more available, people may experience difficulty about knowing where to go and how to find information that they need in a timely fashion. As people are more engaged in multi-tasking in their daily life, the demands for cognitive activities may grow higher. As the demands for cognitive activities in Web searches grow higher, people may express emotional instabilities, increase error rate, or decrease performance.

Information seeking on the Web involves human cognition and perception (Ingwersen 1996). The information seeking process is iterative, complex, evolving, and dynamic and is considered a learning process that requires cognitive activities (Bates 1989; White and Roth 2009; Rieh and Xie 2006). Information seeking on the Web interacts with information systems and their computing features and devices, which are actions that also require cognitive activities (Pillori and Card 1999). Web searches are associated with various factors such as the nature of a search task, the characteristics of information systems, the characteristics of searchers and the time allocated for searching, to name a few (e.g., Bell and Ruthven 2004; Borgman 1989; Byström and Järvelin 1995; Li and Belkin 2008; Vakkari 1999).

A searcher may experience variant aspects of cognitive load or levels of difficulty during information searches influenced by these factors, either alone or in combination. One kind of affective component of cognitive load is associated with the mental demand imposed on the searcher by the search system or the search task. Therefore, assessing the effect of the mental demand experienced by a user during a search process is important for better designing information systems and for identifying where, and possibly how, to lower the mental effort required for effective interaction. Assessment of mental demands imposed on a user's search tasks is also useful for characterizing user interfaces and tasks with respect to their demands on the user's mental effort.

With that said, understanding the dynamic aspects of cognitive load is also important to the field of library and information science in order to better serve users and develop information systems. However, to date, only a few studies have investigated how cognitive load affects searchers' experiences during the information search process.

Based on the previous research in educational and cognitive psychology, this study assumed that, as a searcher's mental demand increases, his or her ability to perform effectively slowly

decreases as it approaches the maximum point of cognitive load. Once searchers reach the maximum point of cognitive load, they may choose to give up on a particular search, present emotional instability, or lose motivation to pursue their original information seeking goal.

This study addressed whether or not there is any difference in the categories of a searcher's thoughts, emotions, efforts, and performance induced by mental demand in cognitive load during the information search process between two groups.

The research questions addressed in this analysis are as follows:

1) How do subjects differ in their *thoughts* between those exposed to mental demand manipulation and those not exposed while performing Web searches?

2) How do subjects differ in their *emotions* between those exposed to mental demand manipulation and those not exposed while performing Web searches?

3) How do subjects differ in their *efforts* between those exposed to mental demand manipulation and those not exposed while performing Web searches?

4) How do subjects differ in their *performance* between those exposed to mental demand manipulation and those not exposed while performing Web searches?

1.2 Purpose of the Study

The purpose of this study is to facilitate understanding of the effects of mental demand in cognitive load during Web searches in terms of a searcher's thoughts, emotions, efforts, and performance.

1.3 Significance of the Study

The findings of this study are expected to provide information professionals with a better understanding of the interaction between searchers and information systems or user interfaces in regard to the mental demand imposed on a searcher while performing Web searches.

2. Related Studies

As mentioned earlier, only a few studies have investigated how cognitive load affects searchers' experiences during the information search process. Therefore, these few studies are introduced

as follows.

Gwizdka mentioned that assessing cognitive load in Web searches is useful for characterizing search system features, search tasks and task stages with respect to their demands on the searcher's mental effort. In addition, he pointed out that assessing cognitive load is also helpful in examining how individual differences among searchers affect the search process and its outcomes (Gwizdka 2009).

Albers stated that as more and more people facilitate websites into their daily routines, it is important to create a mode of interaction that is as easy as possible. However, the lack of understanding with respect to the major impediments to the user within that structure presents a substantial problem to both designers and users (Albers 2006). He measured how cognitive load affects website usability and considered three different methods of measuring cognitive load, i.e., Sternberg Memory Task, tapping task, and the NASA TLX. He also pointed out that all three tests work at different granularity levels of their measurements and have different implementation issues.

Gwizdka examined the effects of working memory capacity on a user's search efforts with regard to user behavior during information search tasks at two levels of complexity. The results show that, in more demanding task conditions, both user groups changed behavior, but they differed in how they changed it. He stated that high working memory users performed more actions to find more information, while low working memory users switched their search tactics by significantly decreasing the number of individual documents they visited (Gwizdka 2013).

Gwizdka also explored the distribution of cognitive load in Web searches and found that the average cognitive load was found to vary by search task stages, and it was significantly higher during query formulation and user description of a relevant document, as compared to examining search results and viewing individual documents (Gwizdka 2010). He assessed cognitive load using Stroop task in which reaction time on the secondary task is differentiated between the primary task stages and the user interfaces. He stated that the higher cognitive load component of the secondary task performance evidenced primary task stages, while the lower cognitive load component evidenced user interfaces (Gwizdka 2010).

Kim and Rieh examined a dual-task method for the assessment of mental effort during online searching, having the users engage in two tasks simultaneously. They found that the results of a dual-task performance do not reveal much difference between the two systems with respect to mental effort or concentration during searches. Rather, searchers invested mental effort differently when viewing search results and reading retrieved documents (Kim and Rieh 2006). The findings of their study indicated that a dual-task method provides a useful technique for measuring mental

effort in online searching, and it has the great potential to be used to measure other aspects of information retrieval, such as task complexity and multitasking information behavior.

Cognitive load has been underestimated in Information Retrieval research regarding its impact on interaction between searchers and information systems but is gradually growing attention in Library and Information Science field. Although limited in previous studies in LIS field's scope and number, previous studies have shown the potential importance of understanding the effect of cognitive load in better assisting information searchers on the Web.

3. Methodology

3.1 Theoretical Framework

This study is informed by Cognitive Load Theory (CLT). In CLT, human working memory capacity is limited by the "magical number seven plus or minus two," as suggested by Miller in 1956. Cognitive load theory is an instructional design theory with the aim of assisting instructional designers to reduce the load caused by poor design of the learning materials. Cognitive load theory uses an information processing approach to cognition, involving working memory, i.e., short-term memory and long-term memory.

3.2 Research Design

The present study uses an experimental design which includes a pre-task survey, experiment, post-task survey, and post-interview in consecutive order. This study mainly used qualitative data to explore the research questions. In experiments, all task performances were recorded when the participants were doing Think-Aloud Protocol (TAP) while they performed the search tasks assigned, and the results were transcribed as verbal data later. The qualitative analysis of verbal data was then quantified into numbers that were associated with the categories of thoughts, emotions, efforts, and performance. All post-interviews were also transcribed and quantified into numbers.

3.2.1 Participants

Since this study was a pilot study and exploratory, we recruited ten students from a university in South Korea whose mean age was 23, ranging from age of 21 to 25 (M=23, SD=1.4), and

the gender distribution consisted of 3 males and 7 females. In a discussion of exploratory and pilot studies, Isaac and Michael suggested that "samples with N's between 10 and 30 have many practical advantages" (Miller 1995, 101), including simplicity, easy calculation, and the ability to test hypotheses, yet "overlook weak treatment effects."

3.2.2 Experiment Setting and Procedure

Ten individuals who are enrolled in a university in South Korea were randomly assigned to either an experimental group or a control group for two assigned information search tasks. Half of the participants who were randomly assigned to the experimental group were exposed and then manipulated with mental demand. The other half of the participants assigned to the control group had no manipulation with mental demand.

In the experiment procedure, the participants completed a pre-task survey, two search tasks, a post-task survey, and a post-interview in consecutive order. The pre-task survey collected participants' demographic information and previous experience with the Naver search engine.

Two search tasks were given, and both groups were asked to perform the tasks. These two search tasks were ill-structured and problem solving information search tasks (See Table 2). Information search tasks in experiments were all screen-captured and recorded using Camtasia screen capture software while participants performed tasks and TAP to collect participants' cognitive activities and emotional states. The post-task survey collected participants' cognitive load index score facilitated by the NASA-TLX assessment tool. Six components (i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration) of the National Aeronautical and Space Administration Task Load Index (NASA-TLX) were measured to obtain a subjective self-reported cognitive load score based on a weighted average of ratings as developed by Hart and Staveland (Hart and Staveland 1988).

The post-interview collected participants' verbal interviews when they reviewed their tasks performance with the observer. This study expected that collected data from the post-interview would reinforce the data from TAP.

Finally, the following four categories shown in Table 2 were identified from the verbal data through two information search tasks in the experiment and the post-interview. These classifications were derived from Kuhlthau's six stags of Information Search Process (Kuhlthau 1991; 1993; 2008; Kuhlthau, Heinström and Todd 2008) and are reproduced. Based on Kuhlthau's Model of ISP, the coding scheme of this study was modified and derived by the activities that are found in the appendix. The modified coding scheme is introduced in Table 1 below.

〈Table 1〉 Modified Coding Scheme

Category	Activities (Sub-Categories)	Definition	Example
Searcher's Thoughts (Activities that require human cognition and perception were coded into the sub-category of a searcher's thoughts.)	Thinking	The act of processing information during information search process	"생각보다 잘 안나오네 라고 생각하고 있어요 정확한걸…"
	Remembering	The act of recalling to the mind by effort of memory in processing information	"아까 찾았던 결과들을 생각해 보니,…"
	Relevance judgment	The extent to which information retrieved in a search of a library collection or other resource, such as an online catalog or bibliographic database, is judged by the user to be applicable to ("about") the subject of the query.	"내가 찾고자 하는 정보인거 같습니다."
	Disturbance	Something that stops you from processing information or searching for relevant information	"예 집중 하는데 방해를 받았어요."
	Searching	Searching is a retrieval process where what is being sought is known by the searcher	"웹페이지 같은 파일을 검색합니다."
	Retrieval	Retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources	"찾은 거 같습니다. 이걸로 추가하겠습니다."
Searcher's Emotions (Emotion is a subjective, conscious experience characterized primarily by psychophysiological expressions, biological reactions, and mental states.)	Confidence	The feeling that you can trust someone or something to be good, work well, or produce good results.	"정보검색후가 전보다는 자신감이 생겼던 것 같습니다."
	Frustration	The feeling of being annoyed, upset, or impatient, because you cannot control or change a situation, or achieve something	"좌절감이요. 비슷한 건 나 올 줄 알았는데, 너무 안나오니까."
	Stress	The continuous feelings of worrying about your search or task that prevent you from relaxing	"스트레스는 처음부터 있었던 것 있었습니다."
	Confusion	The feeling when a searcher does not know what is happening or what to do because it is not clear	"막상 컴퓨터에 문제가 있는 게 아니라서 이게 통할지 안통할지는 저도 모르겠어요."
Searcher's Efforts (Activities that use a series of actions requiring physical or mental energy to do something are considered to be efforts.)	Time spent	Time that was taken for a searcher to retrieve relevant information from start to end.	"제가 투자한 시간에 비해서 생각하고 머리 굴린거에 비해서 나온 검색결과도 원하지 않고"
	Query	Query languages are computer languages used to make queries into databases and information systems.	"아 이게 키워드가 어려워서 더 많이 하게 된 거 같아요."
	Pages visit	Pages that were visited by a searcher to search relevant information	"페이지를 더 많이 본거 같습니다."
	Focus	The thing or situation that a searcher pays special attention to	"그냥, 집중해서 봤어요."
	Work hard	The activities of working with a lot of efforts	"평상시노력이 더 많았다고 생각합니다."
Searcher's Performance (Performance is considered as the action or process of carrying out something or accomplishments of an action, task, or function in a given task by measuring a searcher's effectiveness, efficiency, and satisfaction.)	Performance	The process of judging the competence with which a searcher has performed the search task	"예, 중정도의 수행평가를 내릴 수 있겠어요."
	Satisfaction	The act of providing what is needed or desired for relevant information	"만족하는 자료를 찾을 수 있었어요."

3.2.3 Increasing Mental Demand

To increase mental demand on the experimental group, a dual-task method was employed. The dual-task methodology is based on the assumption that the processing capacity of working memory is limited, but can be flexibly allocated (Pass, Renkl and Sweller 2003). A secondary task was added to the search task so that the dual-task demand was imposed on the experimental group in order to increase mental demand. In the dual-tasks, a search task was the primary task, and the secondary task was to memorize an 8-digit number (i.e., a number with 8 different digits such as 324871659) and recall the number following the completion of the search task.

〈Table 2〉 Search Tasks

Tasks	Questions
Virus-related search task	당신의 개인컴퓨터가 바이러스 공격을 받아 컴퓨터가 실행파일을 실행시킬 수가 없게 되었다. V3백신 프로그램이나 노턴바이러스 같은 프로그램을 실시하여 해결하려 했으나, 이 역시 실행파일이 실행되지 않는다. 지금 당신의 컴퓨터에 침입한 바이러스의 이름과 그 해결책에 대해서 정보검색을 실행하여 적절한 정보를 찾으시오.
Library-related search task	당신이 목록시스템의 관리자로 있는 공공도서관이 갑자기 사고로 불타버렸다. 그 결과 목록시스템이 손실되었다. 이 목록시스템을 복구하기 위한 방법과 전략에 대해서 정보검색을 실행하여 적절한 정보를 찾으시오.

4. Findings and Discussion

Based on Chi-square, none of the demographic variables yielded significant differences between the experimental group and the control group. When considering participants' Web searching experience, 8 out of 10 participants rated it as 'average' or 'high' in terms of computer knowledge. In terms of Web search knowledge, 9 out of 10 participants rated it as 'average' or 'high'. In terms of Web search skill, 8 out of 10 participants rated it as 'average' or 'high'.

In terms of Naver use frequency, 9 out of 10 participants selected 'daily' or 'more than once a day' using the Naver search engine in general. In terms of query reformulation, 9 out of 10 participants rated query reformulation 'about 2 times' or 'more than 3 times' per a search in general. Lastly, in terms of search success, 4 out of 10 participants rated search success 'rarely' and 4 out of 10 participants rated search success 'sometimes', and 2 participants rated search success 'usually' using the Naver search engine in general.

4.1 Searcher's Thoughts

Activities that require human cognition and perception were coded into the sub-category of a searcher's thoughts, such as *thinking, remembering, relevance judgment, disturbance, searching, and retrieval.* For the search task 1, it is shown that the sub-category of thinking, remembering, relevance judgment, and disturbance in between two groups needs careful attention for further study.

⟨Table 3⟩ Summary of Thoughts Category in Task 1

Task 1 only						
Thoughts	Experimental Group			Controlled Group		
	Total	MEAN(experimental)	STD	Total	MEAN(control)	STD
Thinking	19	3.17	1.7	13	1.86	1.1
Remembering	34	5.67	2.8	7	1.4	0.9
Relevance Judgment	83	9.22	9.6	68	6.8	3.9
Disturbance	17	3.4	1.8	3	1.5	0.7
Searching	76	9.5	5.6	71	7.1	8.0
Retrieval	8	2.67	2.1	10	2.5	1.3

⟨Table 4⟩ Summary of Thoughts Category in Task 2

Task 2 only						
Thoughts	Experimental Group			Controlled Group		
	Total	MEAN(experimental)	STD	Total	MEAN(control)	STD
Thinking	23	3.29	2.5	21	3	2.5
Remembering	24	3.43	1.8	2	1	0.0
Relevance Judgment	67	7.44	6.0	61	6.1	3.6
Disturbance	15	3	1.9	4	2	1.4
Searching	66	7.33	6.7	55	5.5	4.6
Retrieval	5	1.25	0.5	7	1.4	0.9

When considering searchers' thoughts category in search task 1, both Table 3 and Figure 1 illustrate that the score for *thinking* in participants' verbal expression was higher for the experimental group (M=3.17, SD=1.7) than that of the control group (M=1.86, SD=1.1). Similarly, the score for *remembering* in participants' verbal expression was higher for the experimental group (M=5.67, SD=2.8) than that of the control group (M=1.4, SD=0.9). The score for *relevance judgment* in participants' verbal expression in search task 1 was higher for the experimental group (M=9.22, SD=9.6) that that of the control group (M=6.8, SD=3.9). In the *disturbance* sub-category, the number of verbal expression occurrences of participants for the experimental group (M=3.4, SD=1.8) was higher than that of

the control group (M=2.5, SD=1.3). We can see that the participants in the experimental group were more likely to do more thinking, remembering, and relevance judgment and get more disturbed than those in the control group. However, the scores in the *retrieval* sub-category in search task 1 were almost the same for the two groups (M=2.67 vs. M=2.5).

This result could tell us that searchers with more mental demand might have difficulty with respect to relevance judgment, but they need to expend greater efforts on thinking, remembering, and searching in order to find relevant information.

Similarly as in search task 1 above, for searchers' thoughts in search task 2 we can see in Table 4 and Figure 2 that participants in the experimental group were more likely to do more thinking, remembering, relevance judgment, searching and get more disturbed than those in the control group during the information search process.
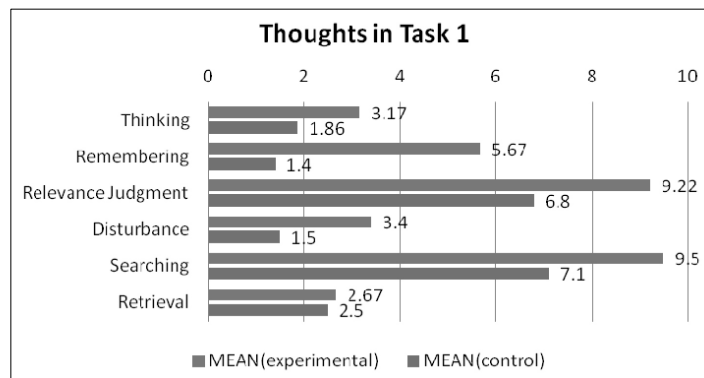
⟨Figure 1⟩ Mean Difference in Thoughts in Task 1
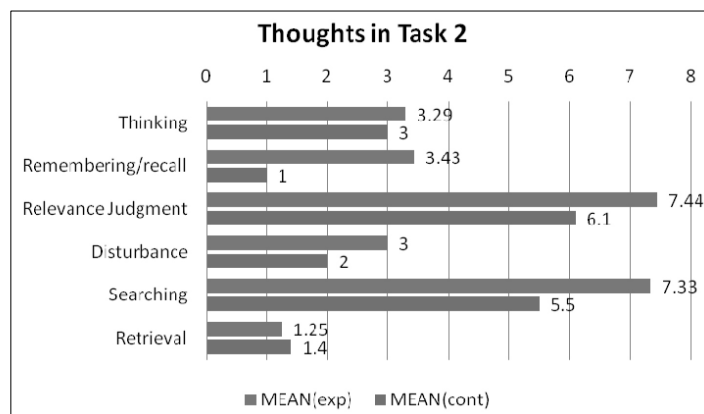
⟨Figure 2⟩ Mean Difference in Thoughts in Task 2

With regard to a searcher's thoughts, the searchers in the experimental group seemed more affected and induced by mental demand. In particular, they seemed to experience the remembering/re-call task more often, searching for information more often, and browsing for information more often than searchers in the control group.

4.2 Searcher's Emotions

In psychology and philosophy, emotion is a subjective, conscious experience characterized primarily by psychophysiological expressions, biological reactions, and mental states. According to Scherer, five elements of emotion exist, including cognitive appraisal, bodily symptoms, action tendencies, expression, and feelings (Scherer 2005). Activities that fall under Scherer's elements were coded into the sub-category of searcher's emotions such as *confidence, frustration, stress, and confusion.*

〈Table 5〉 Summary of Emotions Category in Task 1

Task 1 only						
Emotion	Experimental Group			Controlled Group		
	Total	MEAN(exp)	STD	Total	MEAN(cont)	STD
Confidence	18	3.00	2.00	9	1.8	1.3
Frustration	12	3.00	2.16	8	1.6	0.9
Stress	23	2.30	2.21	38	3.8	4.59
Confusion	42	5.25	6.71	15	1.5	1.51

〈Table 6〉 Summary of Emotions Category in Task 2

| Task 2 only						
Emotion	Experimental Group			Controlled Group		
	Total	MEAN(exp)	STD	Total	MEAN(cont)	STD
Confidence	7	1.17	0.4	22	3.14	1.6
Frustration	16	5.33	3.8	32	4	6.5
Stress	21	2.1	1.85	30	3.33	2.92
Confusion	23	2.3	1.77	36	3.6	3.44

When considering the searchers' emotions category in search task 1, we can see that *confidence* was higher for the experimental group (M=3.00, SD=2.0) than that of the control group (M=1.8, SD=1.3) as shown in Table 5 and Figure 3. In contrast, in search task 2, as shown in Table 6

and Figure 4, *confidence* was lower for the experimental group (M=1.17, SD=0.4) than that of the control group (M=3.14, SD=1.6). This result might show us that there is inconsistency regarding the number of verbal occurrences between the two groups with regard to the tasks, depending upon the participants' mood or will to think aloud during the information search process and performance of the search tasks. Alternatively, the differences could be caused by some other factor that has not yet been recognized.

Similarly, the scores for *confusion* in participants' verbal expression in search task 1 was higher for the experimental group (M=5.25, SD=6.71) than that of the control group (M=1.5, SD=1.51). However, the score for *confusion* in participants' verbal expression in search task 2 was lower for the experimental group (M=2.3, SD=1.77) than that of the control group (M=3.6, SD=3.44).

The level of *frustration* in the participants' verbal expression in search task 1 and search task 2 was higher for the experimental group (M=3.00, SD=2.16 in search task 1 and M=5.33, SD=3.8 in search task 2) that that for the control group (M=1.6, SD=0.9 in search task 1 and M=4.00, SD=6.5 in search task 2), respectively. This result could tell us that searchers with more induced mental demand might get frustrated more often or easily than searchers without such mental demand.

Interestingly, the score in the *stress* sub-category for the control group was higher than that of the experimental group. This sub-category should be considered with care for future studies with more participants. The overall results show that searchers' emotions were likely to be different between the two groups in the sub-categories of stress, frustration and confusion. All of these categories should be examined with care in future studies.
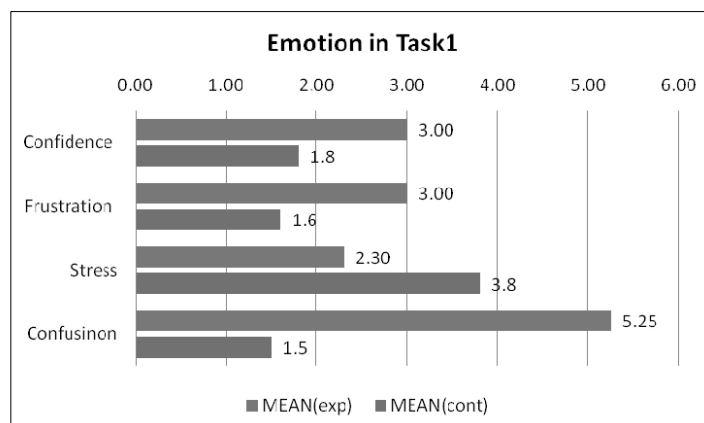
〈Figure 3〉 Mean Difference in Emotion in Task 1
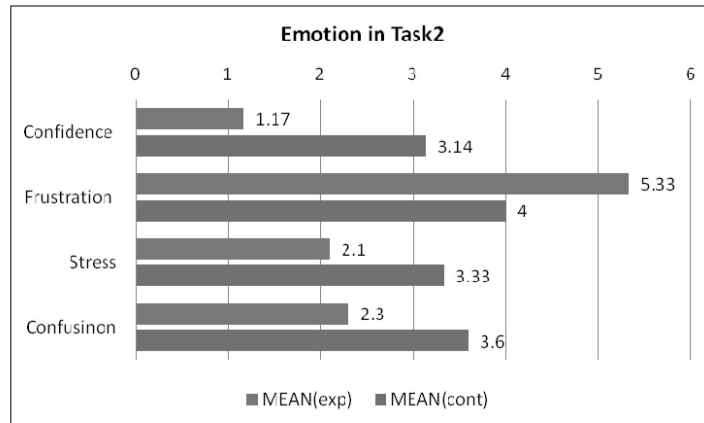
〈Figure 4〉 Mean Difference in Emotion in Task 2

With respect to the searcher's emotions, searchers in both groups tended to show verbal expressions of *frustration* with consistency while they performed the search tasks for this study. However, there was inconsistency in the rankings among the tasks regarding the number of occurrences in these sub-categories. Therefore, we might need to reconsider two dimensional analyses, such as positive vs. negative expressions in these four sub-categories mentioned above.

4.3 Searcher's Efforts

In this study, activities that use a series of actions requiring physical or mental energy to do something are considered to be efforts expended, such as time spent, query, pages visit, focus, and work hard. The results show that searchers' efforts were likely to be different or to be considered with attention between the two groups in the sub-categories of *time spent, query, pages visited,* and *work hard.*

〈Table 7〉 Summary of Efforts Category in Task 1

Efforts	Experimental Group			Controlled Group		
	Total	MEAN(exp)	STD	Total	MEAN(cont)	STD
Time spent	9	3	1.7	8	1.6	0.9
Query	14	2.33	1.4	25	3.13	1.5
Pages visit	3	1	0.0	5	1.25	0.5
Focus	10	2.5	1.3	11	2.75	0.5
Work hard	10	2.5	1.9	3	1.5	0.7

〈Table 8〉 Summary of Efforts Category in Task 2

Efforts	Experimental Group			Controlled Group		
	Total	MEAN(exp)	STD	Total	MEAN(cont)	STD
Time spent	9	3	1.7	9	1.8	0.8
Query	19	3.17	2.2	29	4.14	1.8
Pages visit	6	1.2	0.5	8	1.6	1.3
Focus	9	2.25	1.3	9	1.8	0.8
Work hard	6	1.5	1.0	1	1	

(Table subtitle: Task 2 only)

With respect to the searcher's efforts, in the *time spent* and *work hard* sub-categories in both tasks, it is likely that searchers in the experimental group (M=3.00, SD=1.7 in search task 1 and M=3.00, SD=1.7 in search task 2) tended to spend more time and make more efforts than those in the control group (M=1.6, SD=0.9 in search task 1 and M=1.8, SD=0.8 in search task 2) as shown in Table 7 and 8. This result may indicate that searchers in both groups might take more time to find more relevant information and work hard to make more effort toward accomplishing the task.

In the *query* sub-category in both tasks, it is likely that searchers for the experimental group tended to make fewer queries than those in the control group. In the *focus* sub-category, it is shown that there is inconsistency between tasks.
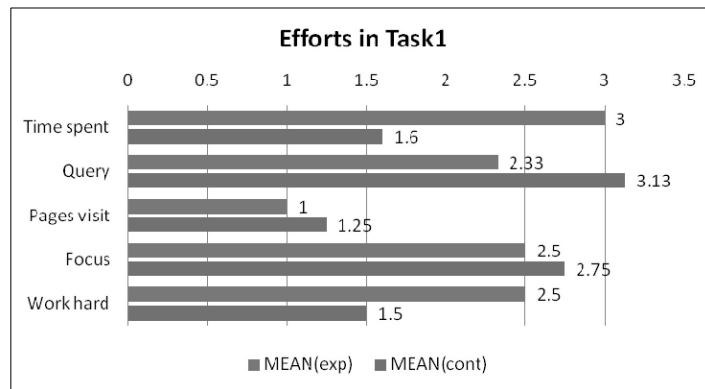
〈Figure 5〉 Mean Difference in Efforts in Task 1
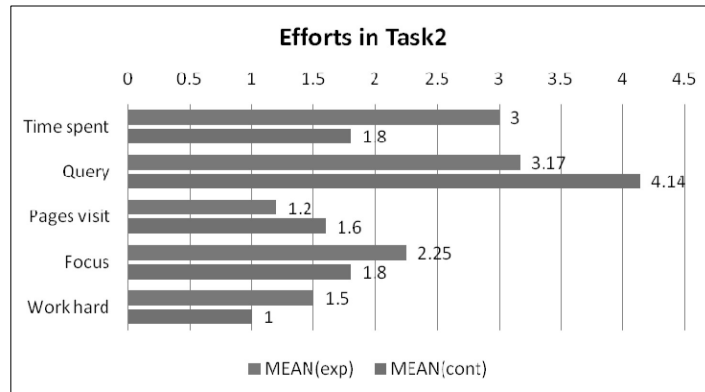
〈Figure 6〉 Mean Difference in Efforts in Task 2

With regard to the searcher's efforts, future research may need to consider not only the occurrences of the searcher's verbal expressions but also the total time spent (h: m: s), the number of queries made, and the number of pages visited.

4.4 Searcher's Performance

Performance is defined as the action or process of carrying out an action or the accomplishments of an action, task, or function in a given task measured. The performance results in this study show that the searcher's performance was likely to be different or to be considered with attention in the sub-category of *performance* and *satisfaction* between the two groups.

〈Table 9〉 Summary of Performance Category in Task 1

Task 1						
Performance	Experimental Group			Controlled Group		
	Total	MEAN(exp)	STD	Total	MEAN(cont)	STD
Performance	41	4.56	2.5	42	4.2	2.7
Satisfaction	7	1.4	0.6	16	2.67	2.3

〈Table 10〉 Summary of Performance Category in Task 2

Task 2						
Performance	Experimental Group			Controlled Group		
	Total	MEAN(exp)	STD	Total	MEAN(cont)	STD
Performance	36	4	2.6	43	4.78	2.1
Satisfaction	1	1		5	1.25	0.5

With respect to the performance assessment sub-category in both tasks, the participants' verbal expressions occurred almost equally in the experimental group (M=4.56, SD=2.5 in task 1 and M=4, SD=2.6 in task 2) and the control group (M=4.2, SD=2.7 in task 1 and M=4.78, SD=2.1 in task 2) as shown in Table 9 and 10. This result may indicate that searchers in both groups might have to check to see if they were on the right track of finding relevant information during the search process.
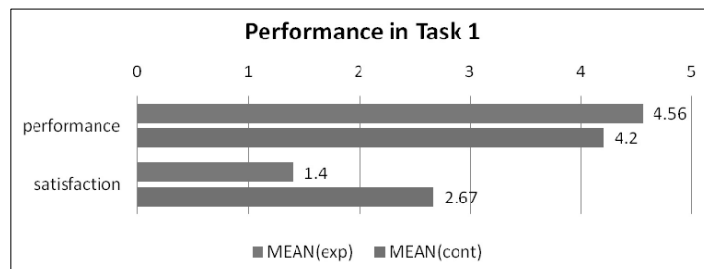
〈Figure 7〉 Mean Difference in Performance in Task 1
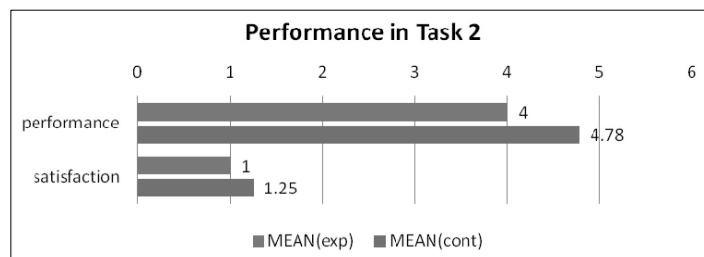
〈Figure 8〉 Mean Difference in Performance in Task 2

With regard to the searcher's performance, searchers in both groups tended to show verbal expressions about *satisfaction* almost at the end of experiments, so there were a low number of occurrences representing performance category verbal expressions.

Finally, NASA-TLX does not show a significant difference between two groups in terms of each component and the total score of cognitive load.

5. Conclusion and Implications

The results of this study seem to show that a higher mental demand imposed on a searcher can interfere with the end users' search experience in four categories: thoughts, emotions, efforts,

and performance. With respect to the searcher's thoughts, relevance judgment was found to be hindered by mental demand. With regard to the searcher's emotions, the experimental group was more frustrated than the control group. With respect to the searcher's efforts, searchers in the experimental group were more likely to spend more time, make fewer queries, and visit fewer pages but work harder to find more relevant information that they needed. Lastly, with regard to the searcher's performance, it is likely that performance was highly dependent upon the completion of the search tasks for both groups. The NASA-TLX cognitive load scores of searchers did not exhibit a significant difference since the study had too few subjects. The overall findings of the study indicate that the use of a dual-task methodology is a promising approach for the assessment of cognitive load induced by complex Web searches.

Given the current nature of the web, higher precision and lower recall is desirable because one can become overwhelmed with millions of hits to assess for relevance. As the Web continues to evolve, the structure of Web contents becomes more complex. However, the lack of understanding of the major impediments to the user within that structure may become a substantial problem to both system developers and users. Therefore, it is certainly important to understand how the mental demand of searchers works during the information search process in order to improve the interaction between the systems and the users. In keeping with this need, information system designers and developers should explore the user's mental demand in order to determine what factors to focus on with the aim of improving the interaction between searchers and systems. One can hope that, in the end, system designers and developers can reduce the mental demand on searchers but take advantage of the results of the effect of mental demand and work to redesign the problem areas in a supportive manner.

6. Limitation of the Study

Since this study is a pilot and exploratory study, our concern in data analysis does not lie in discovering the significance between the two groups and the number of participants; rather, it is important for us to see the trends or likelihood in differences, if any, in between the two groups. The variances between the two groups should be further examined in additional studies. Future research is also needed to explore the correlation between categories.

When using TAP in a study, there are both strengths and weaknesses. The strength of TAP is that it is the closest possible way to draw out the cognitive processes of the participants.

However, this strength appears to be one of the most serious weaknesses, as well, in that only conscious processes are available for verbalization. That said, much of what is going on in the participants' mind remains hidden. Thinking also needs to be slowed down to allow for the additional time that is required for verbalizing thoughts, a reality that may add to the incompleteness of the reports. Therefore, researchers should be aware of this weakness of verbal reports and take it into consideration when designing experiments and analyzing data (Ericsson and Simon 1993).

References

[1] Albers, M. J. 2006. "Measuring Cognitive Load to Test the Usability of Web Sites." *Usability and Information Design.*

[2] Bates, M. J. 1989. "The design of browsing and berrypicking techniques for the online search interface." *Online Review,* 13(5): 407-424.

[3] Bell, D. J. and Ruthven, I. 2004. "Searchers' assessments Of Task Complexity For Web Searching." *In Proceedings of the 26th European Conference on Information Retrieval,* 26: 57-71.

[4] Borgman, C. L. 1989. "All users of information systems are not created equal: An exploration into individual differences." *Information Processing & Management,* 25: 237-252.

[5] Byström, K. and Järvelin, K. 1995. "Task complexity affects information seeking and use." *Information Processing & Management,* 31(2): 191-213.

[6] Ericsson K. A. and Simon H. A. 1993. *Protocol Analysis: Verbal Reports as Data.* MIT Press, Cambridge, MA.

[7] Ford, N., Miller, D. and Moss, N. 2001. "The role of individual differences in internet searching: An empirical study." *Journal of the American Society for Information Science & Technology,* 52(12): 1049-1066.

[8] Ford, N., Miller, D. and Moss, N. 2005. "Web search strategies and human individual differences: Cognitive and demographic factors, Internet attitudes, and approaches." *Journal of the American Society for Information Science & Technology,* 56(5): 741-756.

[9] Gwizdka, J. 2009. "Cognitive load and Web Search tasks." *In proceedings of the third Workshop on Human-Computer Interaction and Information Retrieval.* Washington, DC: Catholic University of America. 54-57.

[10] Gwizdka, J. 2010. "Distribution of Cognitive Load in Web Search." *Journal of the American*

Society for information Science & Technology, 1-23.

[11] Gwizdka, J. 2010. "Using Stroop Task to Assess Cognitive Load." *In Proceeding of the 2010 European Conference on Cognitive Ergonomics.*

[12] Gwizdka, J. 2013. "Effects of Working Memory Capacity on Users' Search Effort." *MIDI 2013 Conference.*

[13] Hart, S. G. and Staveland, L. E. 1988. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

[14] Ingwersen, P. 1996. "Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory." *Journal of Documentation,* 52(1): 3-50.

[15] Isaac, S. and Michael, W. B. 1995. *Handbook in research and evaluation.* San Diego, CA: Educational and Industrial Testing Services.

[16] Kim, Y. M. and Rieh, S. Y. 2006. "Task Performance as a Measure of Mental Effort in Searching a Library System and the Web." *In 68th Annual Meeting of the American Society for Information Science and Technology.*

[17] Kuhlthau, C. C. 1991. "Inside the search process: Information seeking from the user's perspective." *Journal of the American Society of Information Science,* 42(5): 361-371.

[18] Kuhlthau, C. C. 1993. *Seeking Meaning: A Process Approach to Library and Information Services.* Norwood, N.J.: Alex Publishing Corp., 45-51.

[19] Kuhlthau, C. C. 2008. "From Information to Meaning: Confronting Challenges of the Twenty-first Century. Libri: International." *Journal of Libraries & Information Services,* 58(2): 66-73.

[20] Kuhlthau, C. C., Heinström, J. and Todd, R. J. 2008. "The 'information search process' revisited: is the model still useful?" *Information Research,* 13(4): 45.

[21] Li, Y. and Belkin, N. J. 2008. "A faceted approach to conceptualizing tasks in information seeking." *Information Processing & Management,* 44(6): 1822-1837.

[22] Rieh, S. Y. and Xie, H. 2006. "Analysis of multiple query reformulations on the web: The interactive information retrieval context." *Information Processing & Management,* 42: 751-768.

[23] Scherer, K. R. 2005. "What are emotions? And how can they be measured?" *Social Science Information,* 44: 693-727.

[24] Vakkari P. 1999. "Task complexity, problem structure and information actions." *Information Processing & Management,* 35(6): 819-837.

Appendix. Kulthau's ISP Activities

Thoughts	Emotions/Feelings	Efforts/Actions	Strategies
• Contemplating assignment • Comprehending Task • Relating prior experience and learning • Considering possible topics • Weighing topics against criteria of personal interest, project requirements, information available, and time allotted • Predicting outcome of possible choices • Choosing topic with potential for success • Becoming informed about general topic • Seeking focus in information on general topic • Identifying several possible focuses • Inability to express precise information needed • Predicting outcome of possible foci • Using criteria of personal interest, requirements of assignment, availability of materials, and time allotted • Identifying ideas in information from which to formulate focus • Sometimes characterized by a sudden moment of insight • Seeking information to support focus • Defining and extending focus	• Apprehension at work ahead • Uncertainty • Confusion • Sometimes Anxiety • Brief elation after selection • Anticipation of prospective task • Confusion • Doubt • Sometimes threat • Uncertainty • Optimism • Confidence in ability to complete task • Realization of extensive work to be done • Confidence in ability to complete task • Increased interest • Sense of relief • Sometimes satisfaction • Sometimes disappointment	• Talking with others • Browsing the Library • Consulting with informal mediators • Making preliminary search of library • Using reference collection • Locating relevant information • Reading to become informed • Taking notes on facts and ideas • Making bibliographic citations • Reading notes for themes • Using library collect pertinent information • Requesting specific sources from librarian • Taking detailed notes with bibliographic citations • Rechecking sources for information initially overlooked • Confirming information and bibliographic citations	• Brainstorming • Discussing • Contemplating possible topics • Tolerating uncertainty • Discussing possible topics • Predicting outcome of choices • Using general sources for overview of possible topics • Reading to learn about topic • Tolerating inconsistency and incompatibility of information encountered • Intentionally seeking possible focuses • Listing descriptors • Making a survey of notes • Listing possible foci • Choosing a particular focus while discarding others, or • Combining several themes to form one focus • Using descriptors to search out pertinent information • Making comprehensive search of various types of materials, i.e., reference, periodicals, nonfiction, and biography • Using indexes • Requesting assistance of librarian • Returning to library to make summary search Keeping books until completion of writing to recheck information