

9-1-2013

Examining Users' Knowledge Change in the Task Completion Process

Jingjing Liu

University of South Carolina - Columbia

Nicholas J. Belkin

Rutgers University

Xiangmin Zhang

Wayne State University, xiangmin.zhang2@wayne.edu

Xiaojun Yuan

SUNY Albany

Recommended Citation

Liu, J., Belkin, N.J., Zhang, X., Yuan, X. (2013). Examining users' knowledge change in the task completion process. *Information Processing and Management* 49(5): 1058-1074. DOI: <http://dx.doi.org/10.1016/j.ipm.2012.08.006>

Available at: <http://digitalcommons.wayne.edu/slisfrp/96>

This Article is brought to you for free and open access by the School of Library and Information Science at DigitalCommons@WayneState. It has been accepted for inclusion in School of Library and Information Science Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.

NOTICE IN COMPLIANCE WITH PUBLISHER POLICY: This is the authors' final accepted manuscript version ('post-print') of a work that was accepted for publication in *Information Processing and Management*. Changes resulting from the publishing process may not be reflected in this document. This version has been formatted for archiving; a definitive version was subsequently published in *Information Processing and Management* 49(5): 1058-1074. DOI: <http://dx.doi.org/10.1016/j.ipm.2012.08.006>

Examining Users' Knowledge Change in the Task Completion Process

JINGJING LIU School of Library and Information Science, University of South Carolina,
1501 Greene Street, Columbia, SC 29208

NICHOLAS J. BELKIN School of Communication and Information, Rutgers University,
4 Huntington Street, New Brunswick, NJ 08901

XIANGMIN ZHANG School of Library and Information Science, Wayne State University,
106 Kresge Library, Detroit, MI 48202

XIAOJUN YUAN College of Computing and Information, University at Albany,
135 Western Avenue, Albany, NY 12222

Corresponding author JINGJING LIU email:jingjing@mailbox.sc.edu

Abstract This paper examines the changes of information searchers' topic knowledge levels in the process of completing information tasks. Multi-session tasks were used in the study, which enables the convenience of eliciting users' topic knowledge during their process of completing the whole tasks. The study was a 3-session laboratory experiment with 24 participants, each time working on one subtask in an assigned 3-session general task. The general task was either parallel or dependently structured. Questionnaires were administered before and after each session to elicit users' perceptions of their knowledge levels, task attributes, and other task features, for both the overall task and the sub-tasks. Our results support the assumption that users' knowledge generally increases after each search session, but there were exceptions in which a "ceiling" effect was shown. We also found that knowledge was correlated with users' perceptions of task attributes and accomplishment. In addition, task type was found to affect several aspects of knowledge levels and knowledge change. These findings further our understanding of users' knowledge in information tasks and are thus helpful for information retrieval research and system design.

Keywords Knowledge change, pre-task knowledge, post-task knowledge, multi-session task, task type

1. INTRODUCTION

Information tasks involve people searching for information to solve the tasks with which they are in an Anomalous State of Knowledge (ASK) (Belkin, 1980). It is understood that along with the process of working on their tasks, people obtain information and therefore gain knowledge, as Brookes (1980) noted in his "fundamental equation" of information and knowledge (p. 131). As an important factor that could affect users' search behaviors and performance, knowledge has gained much research attention in the field of information science and related areas (e.g., Marchionini, 1989; Allen, 1991; Hsieh-Yee, 1993; Vakkari, Pennanen & Serola, 2003; Wildemuth, 2004; Zhang, Anghelescu, & Yuan, 2005; White, Dumais, & Teevan, 2009).

Despite the assumption that users gain knowledge through working with the information

tasks, so that in later stages users are more knowledgeable about their tasks, there has been no research, to our knowledge, that closely examines users' knowledge and their knowledge change during the task process. We have therefore tested the general assumption about knowledge gain in the information task process, in order both to enhance our understanding of human knowledge gain and to improve the design of information systems.

In everyday life, there are many tasks that cannot be accomplished in one session but require multiple sessions for various reasons, such as the complexity of the task, the difficulty in locating desired information, time constraints, and so on. According to interviews by Spink (1996) with 200 academic end users who searched online public access catalogs or CD-ROM databases for information search, 57% of users conducted multi-session searches. It has also been estimated recently that 25% of the

overall query volume on the web corresponds to this type of task (Donato, Bonchi, Chi, & Maarek, 2010). There have also been studies examining the relationship between users' knowledge and their searching behaviors in multi-stage tasks (e.g., Vakkari, Pennanen & Serola, 2003; Wildemuth, 2004), assuming that users' topic knowledge increases in later stages.

While they happen quite often in everyday life, multi-session tasks also provide a good way to study people's knowledge change in information tasks in laboratory settings. Unlike single session tasks, in which eliciting users' knowledge in the process of task completion is likely to interrupt their tasks, knowledge elicitation during multi-session tasks can be conducted before and/or after sessions, which still measures users' knowledge during the process of task completion.

In the field of human information behavior, many studies have shown that task type is a significant contextual factor that influences the way users search for information (e.g., Li & Belkin, 2008; Liu et al., 2010). Therefore, we explicitly consider the issue of whether task type influences knowledge gain during information seeking.

Thus, in the current study, which is exploratory in nature, we aimed at answering the following research questions in multi-session tasks:

- 1) How does users' knowledge change before and after each information seeking session?
- 2) How does users' knowledge change across multiple information seeking sessions?
- 3) How do users' knowledge and knowledge change relate to task attributes including task difficulty, task success, and task satisfaction?
- 4) Does users' knowledge change show different patterns in different task types?

2. LITERATURE REVIEW

2.1 The effect of knowledge on search

As one of the factors that affect information seekers' search behaviors, knowledge has received much research attention. Previous studies have found, for instance, that knowledge affects users' search term use (e.g., Hsieh-Yee, 1993; Wildemuth, 2004), reading time on documents (e.g., Kelly & Cool, 2002; Kelly, 2006), and search performance measured by precision and recall (e.g., Allen, 1991; Marchionini, 1989). Such studies have generally taken one of two methodological approaches: comparison of behaviors between groups with different levels of knowledge; and, comparison of behaviors of individuals with (presumed) changing levels of knowledge. We review next some studies of the former sort, and follow with a review of studies of the latter type.

Hsieh-Yee (1993) gave participants one search task inside their field (with high familiarity) and one task outside their field (with low familiarity). She found that when users had a certain search experience, subject knowledge affected their search tactics. When working with a less familiar topic, participants used the thesaurus more for term suggestion, made more effort in preparing for the search, and included more synonyms, etc. Hembrooke et al. (2005) took a similar approach, having participants choose 2 tasks that they were familiar with and 2 tasks that they were not. They found that experts with high topic knowledge issued longer and more complex queries, and used elaborations as a reformulation strategy more often as compared to simple stemming and backtracking modifications used by novices. Sihvonen & Vakkari (2004) invited two groups of participants: an expert group of 15 undergraduates in pedagogy and a novice group of 15 students with no previous studies in this field. They found that the thesaurus improved search effectiveness of the experts, but not the novices.

Zhang et al. (2005) divided their participants into high and low groups according to their self-rated familiarity with selected thesaurus terms in the heat and thermodynamics domain. They found that domain experts issued longer queries and had more queries per task. Duggan and Payne (2008) also divided participants into low

and high knowledge groups according to their knowledge scores of knowledge tests. They found that while domain knowledge for the music domain had little effect on search performance, that for the football domain had a large effect on search performance: it was positively correlated with search accuracy, and negatively correlated with time spent on web pages and mean query length. White et al. (2009) separated domain experts and novices according to the frequency of their visits to some representative websites in specific domains. They found that within their domain of expertise, experts search differently than non-experts in terms of the sites they visit, the query vocabulary they use, their patterns of search behavior, and their search success.

Kelly & Cool (2002) looked at the correlation between users' behaviors and their self-rated topic familiarity based on a 5-point scale. They found that with the increase of one's familiarity with topics, his/her reading time tended to decrease and the efficacy, measured by the ratio of the number of saved documents to the total number of viewed documents, increased. Using the same method of participants' self-rated familiarity degrees with task topics, Kelly (2006) found that user topic familiarity, as a contextual factor, had significant effects on user behaviors, specifically, document display time.

Because in these studies, each participant's knowledge was considered to be static at the research time, it was not possible to investigate changes in their knowledge, or changes in their searching with any presumed knowledge change. Another group of studies of knowledge and search behavior has been able to consider differences in behaviors as levels of knowledge change.

In her Information Seeking Process (ISP) model, Kulthau (1991) pointed out that users' cognitive states, i.e., thoughts, change along with the search process. This suggests that users' knowledge changes during the course of the entire process, as one would assume. The literature has seen efforts studying information searchers' behavior changes while their knowledge changes along different search stages. Vakkari, Pennanen, & Serola (2003) examined the search behaviors

of 22 psychology students who searched on PsychINFO while preparing a research proposal. Observation was made at two points in their development of the research proposals: one in the beginning of a 3-month seminar and the other at the end of it. The study found that as students learned more about their research topics, i.e., in the end of the seminar, they used a wider and more specific vocabulary, compared to the beginning of the seminar. Likewise, Wildemuth (2004) conducted a study of 77 medical students working with questions in the domain of microbiology. Three assessment occasions were used in the research design to represent different levels of domain knowledge among the participants: one before entering a medical course, one right after the end of the course, and one six months after they finished the course. The study found that low domain knowledge was associated with less efficient selection of concepts to include in the search and with more errors in the reformulation of search tactics. These studies all assumed that users' knowledge changed at the different stages. However, there were no attempts at describing in detail how users' knowledge changed during, or as a result of, the searching process.

2.2 Knowledge assessment

Due to the fact that knowledge is not easily measured, we provide a summary of knowledge assessment and/or elicitation methods. In general, studies in the literature have used the following methods: 1) scoring by the accuracy or correctness of answers to some test questions in a domain (e.g., Duggan & Payne, 2008); 2) evaluating the output of some tasks, for example, measuring the coverage and depth of the summaries that participants were asked to write as used in Kammerer et al. (2009) and Wilson & Wilson (in press); 3) participants' self-rating of their familiarity with thesaurus terms in a specific domain (e.g., Zhang et al., 2005; Cole et al., 2010); and 4) participants' self-rating of their familiarity with the task topics (e.g., Kelly & Cool, 2002; Kelly, 2006). Administering domain knowledge tests may be costly and it is challenging to correlate the measurements with the topics or tasks in an

experimental setting or real situations where the tasks are not covering the whole domain. In addition, the tests must be carefully selected to maintain reliability. On the other hand, participants' self-rating of their knowledge has been used in many studies due to its convenience and not adding stress to the participants.

One issue about self-reported knowledge level is its reliability. Brantmeier & Vanderplank (2008) claimed that self-assessment questionnaires are good predictors of computer-based tests and classroom performance as measured in terms of sentence comprehension and multiple-choice tests. Cole et al. (2010) studied the relationship between participants' self-ratings of their topic knowledge with genomics tasks and their self-ratings on their familiarity with MeSH terms in the genomics domain. They found that participants' self-rated knowledge levels were significantly and highly correlated with their knowledge levels elicited by their familiarity with the thesaurus terms. Kelly et al. (2006) evaluated alternative information systems for intelligence analysts with involvement of self-assessment and cross-assessment of the reports produced by the analysts. They reported that the analysts producing the best reports, were most accurate in ranking the work of others but most likely to underestimate their own reports in self-assessment; and the analysts who produced the lower ranked reports did the opposite. All these studies indicated that while not an ideal method, in general, participants' self-rating on their knowledge levels or topic familiarity degrees could be a rather reliable estimate of their "real" knowledge levels, especially when there are a group of participants where the possibility of over- or under-estimation could be evened out.

2.3 The effect of task type on search

In the field of human information behavior, many studies have shown that task type as a contextual factor influences the way users search for information. One stream of such research examines the effect of task type on user behaviors in searching and has found significant effects. For example, information gathering tasks are the

most complex out of several types and require long task completion time and viewing more pages (Kellar et al., 2007); complex tasks require users to seek more types and more sources of information, and users working on complex tasks are less likely to predict the types of information they need, and are more dependent upon experts to provide useful information (Byström & Järvelin, 1995; Byström, 2002); mixed-product tasks require longer task completion time, more pages and more search sources than factual tasks (Liu, Gwizdka, Liu & Belkin, 2010). Another stream of research explores whether task type affects modeling of document usefulness or task difficulty. For example, task type information helps predict document usefulness from first dwell time and task stage (Liu & Belkin, 2010). There are also studies which have found that task type should be considered when predicting task difficulty based on observed behaviors (Kim, 2006; Liu, Gwizdka, Liu & Belkin, 2010).

The previous studies on task types were based on various task classification schemes, which makes it difficult to compare across studies. Li & Belkin (2008) developed a comprehensive task classification scheme, with 15 dimensions along which task features can be defined. Based on this scheme, tasks could be designed with some facets controlled and other facets varied to examine their effects. For example, Liu, Gwizdka, Liu & Belkin (2010) found that several task facets, e.g. goal (quality), product, level of document judgment, objective complexity, all had influence upon users' search behaviors when accomplishing search tasks.

2.4 Summary

There is a substantial literature on the relationships between knowledge of task or topic, and information seeking behaviors, and some research on the relationships among task types, knowledge and search behaviors. But there has been little, if any research on how task/topic knowledge changes during the course of information seeking with respect to some task or goal, nor of possible interaction effects between task type and knowledge change. The reason for

wanting to know about this is based on the research that shows how search behaviors change with knowledge level. If we wish to support people in their information seeking, we would like to be able to interpret their behaviors such that we can make accurate predictions of their knowledge levels.

3. METHOD

A 3-session laboratory experiment was conducted which collected data to answer our research questions. The choice of 3 sessions was the same as what has been done in the literature, such as Vakkari, Pennanen, & Serola (2003) and Wildemuth (2004). This section introduces the study design and describes how data were collected in detail.

3.1 Experiment settings

The experiment was conducted in a usability laboratory on campus. A desktop computer with Internet Explorer browser was used as the experimental system in the study. Participants were invited in the study to complete tasks that required them to search the Web for the needed information. Participants were asked to use the browser to find the needed information for the assigned task. They could use any search engine(s) they preferred. Users' interaction with the system was logged by the usability software *Morae*¹.

3.2 Tasks

Tasks were designed to mimic journalists' assignments since they could be relatively easily set as realistic tasks in different domains. Among the many dimensions of task types, this study focused on task structure, i.e., the inter-subtask relation, varying it while keeping other facets in the comprehensive task classification scheme proposed by Li & Belkin (2008) as constant as possible. This makes it reasonable to attribute the task difference to this single factor of task struc-

ture. Two task types were used in the current study: one parallel and one dependent, similar to the two task types of the parallel and the hierarchical used in Toms et al. (2007). They both had three sub-tasks, each of which was worked on by the participants during one session, for three sessions in total.

The tasks asked the participants to write a three-section feature story on hybrid cars for a newspaper, and to finish and submit each article section (also called a report) at the end of each session. At the end of the 3rd session, participants were asked to integrate the 3 sections into one article. In the dependent task (DT), the three sub-tasks were: 1) collecting information on what manufacturers have hybrid cars; 2) selecting three models that you will mainly focus on in this feature story; and 3) comparing the pros and cons of three models of hybrid cars. In the parallel task (PT), the three sub-tasks were finding information and writing a report on three models of cars from the automobile manufacturers renowned for good warranties and fair maintenance costs: 1) Honda Civic hybrid; 2) Nissan Altima hybrid, and 3) Toyota Camry hybrid. It was hypothesized that the sub-tasks in the parallel task were in parallel and independent of one another, but those in the dependent task would be perceived to be having at least some notional order.

3.3 Participants

The study recruited 24 undergraduate Journalism/Media Studies and Communication students (21 female, 3 male) via student mailing list and post board in the School of Communication and Information at Rutgers University. Participants' mean age was 20.4 years. They self-reported to have an average of 8.4 years of online searching experience, and rated their levels of expertise with searching as slightly above average ($M=5.38$) (1=novice, 7=expert). Each of them came 3 times within a 2-week period based on their convenient schedule. Each was assigned to a specific task type: either parallel or dependent. To maintain consistency, sub-task orders in task description in both tasks were rotated and users were allowed to choose whatever order of sub-

¹ <http://www.techsmith.com/morae.html>

tasks they preferred². Each participant obtained a \$30 payment after finishing all 3 sessions. To encourage their serious participation in the study, the participants were offered an incentive of an additional \$20 for the top 6 who submitted the most detailed reports.

3.4 Procedure

Participants came individually to the usability lab to take part in the experiment. Upon arrival in the first session, they completed a consent form and a background questionnaire eliciting their demographic information and search experience. They were then given the general work task (either parallel or dependent) to be finished in the whole experiment. A pre-session task questionnaire then asked them to rate, on a 7-point scale (1=not, 7=very) (note: all questionnaires in the study used the same scale except when otherwise specified), their familiarity with the general task topic, their previous experience with this kind of assignment, and how difficult they think it would be to find information for the task. They were then asked to pick one sub-task to work on in that session. A pre-session sub-task questionnaire followed to collect their familiarity with the sub-task topic, their previous experience with this kind of sub-assignment, and how difficulty it would be to find information for the sub-task. Then participants were given up to 40 minutes to work on the subtask: searching freely on the Web for useful sources and writing the report. After that, participants went through an evaluation process to rate the usefulness of each webpage/document that they had viewed. A post-session sub-task questionnaire and a post-session general task questionnaire were then administered to elicit user perceptions on the difficulty, task success, and satisfaction with the sub-task and the general task, as well as familiarity with

the subtask and general task topics. This ended the first session.

In the 2nd and the 3rd sessions, participants went through the same processes except for filling out the consent form and the background questionnaire. In the 3rd session, after the post-session general task questionnaire, an exit interview asked participants to rate their overall knowledge gain (on a 7-point scale) and to comment on the whole experiment.

3.5 Knowledge variables

As indicated above, in our study, participants were asked to rate their familiarity level with the task topics based on a 7-point scale, where “1” was for “Not at all” and “7” was for “Extremely”. This rating of topic familiarity was treated as measuring one’s topic knowledge level. Topic knowledge in our study was measured for both the general tasks and the sub-tasks, both before and after each session. Thus, we had four knowledge variables that enable us to comprehensively examine users’ knowledge levels and their changes. The following shows the four variables and the questions that participants were asked to rate on³:

- Pre-session general task topic knowledge: “How familiar are you with the topic of this assignment?”
- Post-session general task topic knowledge: “At this point, how familiar are you with the topic of this GENERAL assignment?”
- Pre-session sub-task topic knowledge: “How familiar are you with the topic of this SUB-assignment?”
- Post-session sub-task topic knowledge “At this point, how familiar are you with the topic of this SUB-assignment?”

² It turned out that the 12 dependent task participants had 5 actual sub-task orders, 8 of them were following collect-select-compare, which seemed a logical order to most people. The 12 parallel task participants had 6 task orders, and there was not a single order that appeared to be an overwhelming preference of the participants.

³ The capital words in the questions were as they were in the questionnaires. The use of capital words was to avoid possible confusions between general task and sub-task to the users.

4. RESULTS

To answer our 4 research questions, analyses were conducted for the above listed 4 knowledge variables in both the parallel and the dependent tasks and for all 3 sessions regarding users' knowledge levels, knowledge changes, as well as the relationships between them and other variables. We found that these knowledge variables were not normally distributed, so non-parametric tests were used in our analyses unless otherwise specified.

4.1 Overview of users' knowledge levels

4.1.1 Overview of general task knowledge levels

Although knowledge variables were not normally distributed, it still helps intuitively to look at their means and standard deviations (SDs). Figure 1 shows users' pre- and post-session general task topic knowledge in the 3 sessions when both tasks were considered together.

The within-session knowledge comparison (Table 1) shows that when the 3 sessions were combined, on average, users' post-session general task knowledge was higher than the pre-session one. The same pattern was found in individual sessions 1 and 2. In session 3, the descriptive data shows the same tendency, but the difference is not significant.

Comparison of users' knowledge among sessions (Table 2) found significant differences for the pre-session general task knowledge (Kruskal-Wallis $H(2, N=72)=14.89, p<.005$) and the post-session general task knowledge ($H(2, N=72)=13.90, p<.005$). Post-hoc analysis using the Tukey test revealed that the differences in both types of knowledge were between sessions 1 and 3. In session 3, not only did users have significantly higher levels of pre-session knowledge than in session 1, but they also reached significantly higher levels of post-session knowledge than in session 1. In addition, as one would expect, the difference between users' post-session knowledge in session 3 and pre-session knowledge in session 1 was significant ($H(2, N=24)=4.27, p=.000$). These results all demonstrate that users' knowledge

levels did increase in the process of completing their tasks.

Another point to note is that users' pre-session general task knowledge was a bit lower than the previous session's post-session general task knowledge. This is reasonable considering that when coming back in a later session, participants may have forgotten some of what they had learned in the previous session. It may also have something to do with the learning effect and confidence. After the previous session, users may have felt that they learned a lot and were confident in giving a higher score in the new session for their pre-task knowledge.

4.1.2 Overview of sub-task knowledge levels

Figure 2 shows the changing tendency of the pre- and post-session sub-task knowledge in the 3 sessions. Table 3 shows the results of paired comparison between pre- and post-session sub-task topic knowledge using a paired t-test. As can be seen, in general, users' sub-task topic knowledge did increase.

Table 4 shows the comparison of the pre- and post-session sub-task knowledge among 3 sessions. Unlike the pre-session general task knowledge that increased along sessions, pre-session sub-task knowledge did not have differences among 3 sessions ($H(2, N=72)=0.165, p>.05$). This is reasonable considering that the sub-tasks were different between sessions, and that users could have equal levels of baseline knowledge on the different sub-tasks. On the other hand, users' post-session sub-task knowledge did show a difference ($H(2, N=72)=8.18, p<.05$). Post-hoc analysis using Tukey found that the difference was between sessions 1 and 3. This meant that although for the different sub-tasks, users had equal levels of baseline knowledge, after working on them, they had a higher level of knowledge in session 3 than in session 1. It could be possible that users learned more for the sub-task in the 3rd session given their experiences in the previous sessions with the sub-tasks in the same general task.

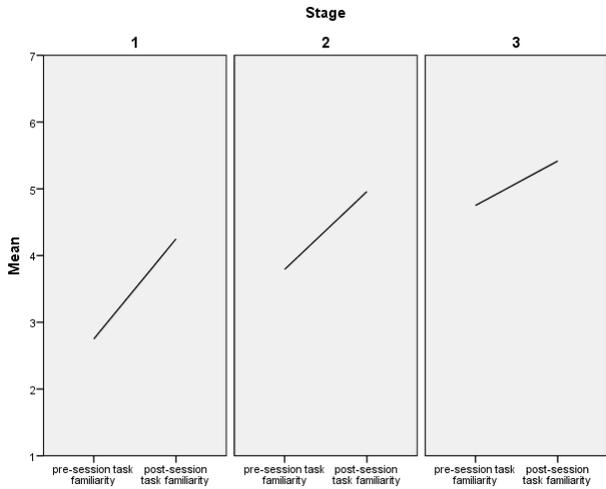


Figure 1. Pre- and post-session general task topic knowledge

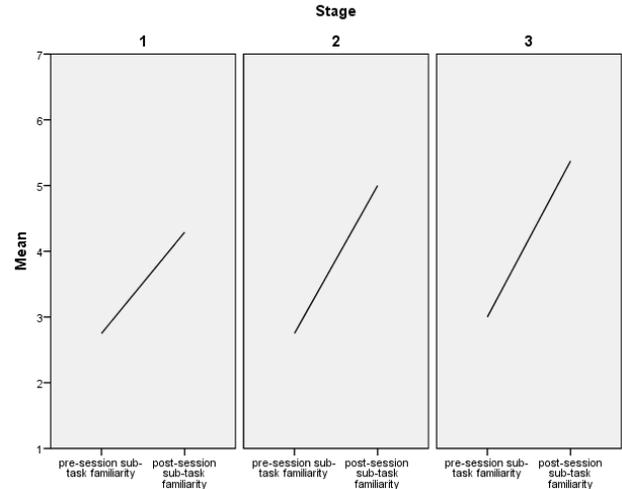


Figure 2 Pre- and post-session sub-task topic knowledge

Table 1 Paired comparison between pre- vs. post-session general task topic knowledge (Mean (SD))

	Pre-session	Post-session	Z(p)
Session 1	2.75 (1.51)	4.25 (1.03)	3.323 (.001)
Session 2	3.79 (1.53)	4.96 (1.12)	3.685 (.000)
Session 3	4.75 (1.80)	5.42 (1.02)	1.620 (.105)
3 sessions together	3.76 (1.80)	4.88 (1.15)	5.045 (.000)

Table 2 Pre- vs. post-session general task topic knowledge across sessions (Mean (SD))

	Pre-session	Post-session
Session 1	2.75 (1.51)	4.25 (1.03)
Session 2	3.79 (1.53)	4.96 (1.12)
Session 3	4.75 (1.80)	5.42 (1.02)
H(p) of 3-session comparison	14.89 (.001)	13.90 (.001)

Table 3 Paired comparison between pre- and post-session sub-task topic knowledge (Mean (SD))

	Pre-session	Post-session	Z(p)
Session 1	2.75 (1.51)	4.29 (1.46)	3.00 (.003)
Session 2	2.75 (1.70)	5.00 (0.93)	3.97 (.000)
Sessoin 3	3.00 (1.91)	5.38 (1.28)	4.04 (.000)
All 3 sessions	2.83 (1.70)	4.89 (1.31)	6.31 (.000)

Table 4 Pre- vs. post-session sub-task topic knowledge across sessions (Mean (SD))

	Pre-session	Post-session
Session 1	2.75 (1.51)	4.29 (1.46)
Session 2	2.75 (1.70)	5.00 (0.93)
Session 3	3.00 (1.91)	5.38 (1.28)
H(p)	0.165 (0.921)	8.177 (.0017)

4.2 Users' knowledge in two tasks

4.2.1 General task topic knowledge in two task types

Figure 3 shows users' pre- and post-session general task knowledge in the 3 sessions in the two tasks. Table 5 shows the within-session comparison of knowledge in individual tasks using paired t-test. As can be seen, when all 3 sessions were considered together, users' post-session general task knowledge level was higher than their pre-session one, and this was also found in both tasks. In the dependent task, post-session task knowledge level was always higher than the pre-session one. In the parallel task, however, only in session 2 was post-session task knowledge level significantly higher than the pre-session one. In sessions 1 and 3, no significant differences were found between users' pre- and post-session knowledge levels, although descriptively, the latter was higher than the former.

Table 6 shows the between-session comparison of users' pre- and post-session general task knowledge levels in the two tasks. As can be seen, in the dependent task, users' rating scores for both the pre-session general task knowledge ($H(2, N=72)=7.97, p<.05$) and the post-session general task knowledge ($H(2, N=72)=16.39, p<.001$) had differences among 3 sessions. Post-hoc analysis using Tukey revealed that the difference for pre-session knowledge was between sessions 1 and 3, and the differences for post-session knowledge were between sessions 1 and 2, and between sessions 2 and 3. In the parallel task, users' rating scores for the pre-session general task knowledge also had differences among 3 sessions ($H(2, N=72)=8.15, p<.05$), specifically, between session 1 and 3 as revealed by Tukey test. However, the post-session general task knowledge levels did not significantly increase along sessions ($H(2, N=72)=3.80, p>.05$).

We did further analysis using General Linear Model (GLM) for the effects of task and session on the pre- and post-session general task knowledge. Results (Table 7) show that in general, users' knowledge level increased across 3 sessions, and this happened with both the pre-task knowledge ($F(2, 70)=9.601, p<.001$) and the

post-task knowledge ($F(2, 70)=8.181, p=.001$). As for the effect of task, results show that users' pre-session general task knowledge level in the parallel task was higher than that in the dependent task ($F(2, 70)=5.336, p<.05$). However, post-session general task knowledge in the two tasks did not show significant differences. This means that although users working with the parallel task showed a higher level of baseline knowledge than those in the dependent task, after sessions, they had equal levels of knowledge. It would be interesting for us to further explore if the same result applies to a study using different kinds of tasks.

4.2.2 Sub-task topic knowledge in two task types

Figure 4 shows the pattern of users' pre- and post-session sub-task knowledge change in the 3 sessions in the two tasks. The comparison between pre- and post-session sub-task knowledge (Table 8) shows that in general, users' post-session sub-task knowledge level was higher than the pre-session one, and this applied to both tasks, in each of the 3 sessions, as well as when 3 sessions were considered together.

Table 9 shows the results of knowledge level comparison across 3 sessions, for the pre- and the post-session sub-task knowledge respectively, in two tasks. No statistically significant differences were found in either pre- or post-session knowledge among the 3 sessions, in either type of task. This means that in individual tasks, users' pre- or post-session sub-task knowledge level in later sessions was not higher than that in the previous sessions, despite that when both tasks were combined for analysis, users' post-session sub-task knowledge level in session 3 was greater than that in session 1 (see Table 5).

The GLM analysis for the effects of task and session on pre- and post-session sub-task knowledge level (Table 10) showed that task was not a significant factor on either of them. As for task session's effect, users' pre-session sub-task knowledge level did not show a difference between sessions ($F(2, 70)=0.166, p>.05$), but their post-session sub-task knowledge was different between sessions ($F(2, 70)=4.632, p<.05$). These were consistent with those displayed in Table 9.

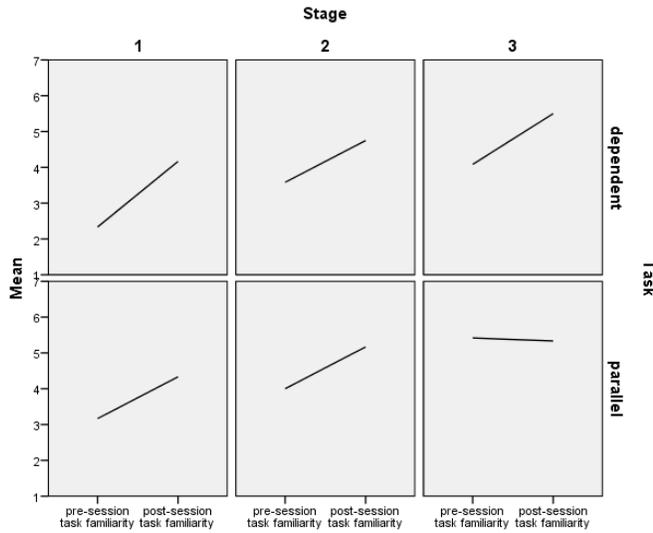


Figure 3 Pre- and post-session general task topic knowledge in three sessions in two tasks

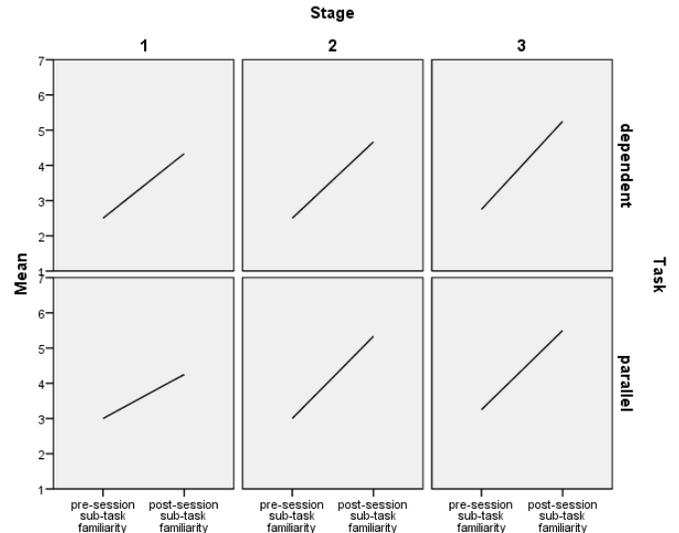


Figure 4 Pre- and post-session sub-task topic knowledge across sessions

Table 5 Paired comparison between pre- and post-session general task topic knowledge in two tasks (Mean (SD))

	Dependent task			Parallel task		
	Pre-session	Post-session	Z(p)	Pre-session	Post-session	Z(p)
Session 1	2.33 (1.07)	4.17 (0.39)	2.969 (.003)	3.17 (1.80)	4.33 (1.44)	1.810 (.070)
Session 2	3.58 (1.44)	4.75 (0.62)	2.401 (.016)	4.00 (1.65)	5.17 (1.47)	2.889 (.004)
Sessoin 3	4.08 (1.68)	5.58 (0.90)	2.388 (.017)	5.42 (1.73)	5.42 (1.17)	.000 (1.00)
All 3 sessions	3.33 (1.57)	4.83 (0.88)	4.451 (.000)	4.19 (1.93)	4.97 (1.40)	3.073 (.002)

Table 6 Pre- vs. post-session general task topic knowledge across sessions in two tasks (Mean (SD))

	Dependent task		Parallel task	
	Pre-session	Post-session	Pre-session	Post-session
Session 1	2.33 (1.07)	4.17 (0.39)	3.17 (1.80)	4.33 (1.44)
Session 2	3.58 (1.44)	4.75 (0.62)	4.00 (1.65)	5.17 (1.47)
Session 3	4.08 (1.68)	5.58 (0.90)	5.42 (1.73)	5.42 (1.17)
H(p)	7.96 (.019)	16.39 (.000)	8.15 (.017)	3.80 (.149)

Table 7 GLM analysis of session and task effects on pre- and post-session general task knowledge

	Pre-task general task knowledge F(p)	Post-task general task knowledge F(p)
Session	9.601 (.000)	8.181 (.001)
Task	5.336 (.024)	0.301 (.585)
Session*task	0.505 (.606)	0.446 (.642)

Table 8 Paired comparison between pre- vs. post-session sub-task topic knowledge in two tasks (Mean (SD))

	Dependent task			Parallel task		
	Pre-session	Post-session	Z(p)	Pre-session	Post-session	Z(p)
Session 1	2.50 (1.38)	4.33 (1.16)	2.325 (.020)	3.00 (1.65)	4.25 (1.77)	2.024 (.043)
Session 2	2.50 (1.51)	4.67 (0.65)	2.971 (.003)	3.00 (1.91)	5.33 (1.07)	2.701 (.007)
Session 3	2.75 (1.77)	5.25 (1.36)	2.947 (.003)	3.25 (2.09)	5.50 (1.24)	2.821 (.005)
All 3 sessions	2.58 (1.52)	4.75 (1.13)	4.716 (.000)	3.08 (1.84)	5.03 (1.46)	4.203 (.000)

Table 9 Pre- vs. post-session sub-task topic knowledge across sessions in two tasks (Mean(SD))

	Dependent task		Parallel task	
	Pre-session	Post-session	Pre-session	Post-session
Session 1	2.50 (1.38)	4.33 (1.16)	3.00 (1.65)	4.25 (1.77)
Session 2	2.50 (1.51)	4.67 (0.65)	3.00 (1.91)	5.33 (1.07)
Session 3	2.75 (1.77)	5.25 (1.36)	3.25 (2.09)	5.50 (1.24)
H(p)	0.11 (0.95)	5.03 (0.08)	0.06 (0.97)	3.95 (0.14)

Table 10 GLM analysis of session and task effects on pre- and post-session sub-task knowledge

	Pre-session sub-task knowledge F(p)	Post-session sub-task knowledge F(p)
Session	.166 (.847)	4.632 (.013)
Task	1.496 (.226)	.886 (.350)
Session*task	.000 (1.000)	.540 (.585)

Table 11 Correlations between pre-task general knowledge, post-task general knowledge, and other variables

	Pre-task knowledge			Post-task knowledge		
	Both tasks	DT	PT	Both tasks	DT	PT
Post-task knowledge	.479* (.018)	.439 (.153)	.571 (.053)	---	---	---
Post-task difficulty	-.439* (.032)	-.108 (.739)	-.682* (.015)	-.217 (.310)	-.340 (.279)	-.210 (.513)
Task success	.335 (.109)	.321 (.309)	.377 (.227)	.674** (.000)	.694* (.012)	.669* (.017)
Task satisfaction	.318 (.130)	.526 (.079)	.290 (.360)	.719** (.000)	.804** (.002)	.673* (.017)

Table 12 Correlations between pre-session sub-task knowledge and other variables

	Pre-session sub-task knowledge											
	Session 1			Session 2			Session 3			3 sessions combined		
	Both tasks	DT	PT	Both tasks	DT	PT	Both tasks	DT	PT	Both tasks	DT	PT
Post-session knowledge	.074 (.731)	.171 (.595)	.031 (.923)	.603** (.002)	.648* (.023)	.578* (.049)	.551** (.005)	.370 (.236)	.716** (.009)	.380** (.001)	.354* (.034)	.381* (.022)
Post-session difficulty	-.175 (.415)	-.062 (.849)	-.285 (.369)	-.181 (.398)	-.094 (.771)	-.285 (.369)	-.239 (.261)	-.020 (.952)	-.579* (.049)	-.198 (.095)	-.037 (.828)	-.396* (.017)
Sub-task success	-.148 (.490)	-.027 (.935)	-.283 (.373)	.000 (1.00)	-.316 (.318)	.198 (.538)	.289 (.171)	.366 (.242)	.182 (.571)	.071 (.555)	.081 (.631)	.069 (.688)
Sub-task satisfaction	-.121 (.574)	-.042 (.897)	-.153 (.636)	-.200 (.348)	-.314 (.320)	-.049 (.879)	.084 (.695)	.085 (.794)	.013 (.969)	-.051 (.668)	-.056 (.746)	-.040 (.816)

** correlation is significant at p<0.01 level (2-tailed); * correlation is significant at p< 0.05 level (2-tailed).

4.3 Correlation between knowledge and other variables

4.3.1 General task knowledge and other variables We also looked at the relations between knowledge and other variables that represent users' perceptions of task attributes and accomplishment. These included three aspects that were elicited in the post-task questionnaires:

- Task difficulty: how difficult was it to find information users needed for the task.
- Task success: how successful users think they were in gathering the information needed for the task.
- Task satisfaction: how satisfied users were with their reports submitted for the task.

As mentioned above, users' knowledge of the general task topic was elicited in each session, however, the post-task difficulty, task success, and task satisfaction were elicited only once in the experiment, at the end of the 3rd session. It would be most reasonable to focus on the pre-session task knowledge in session 1 (could also be called the pre-task knowledge), and the post-session task knowledge at the end of session 3 (could also be called the post-task knowledge), and look at their relationships with other variables.

As can be seen from Table 11, users' pre- and post-task knowledge had a positive correlation when both tasks were combined, and a borderline positive correlation ($p=.053$) in the parallel task, but no correlation in the dependent task. This means that although in general (no task information provided) and in the parallel task, users' post-task knowledge levels (being high or low) could possibly be predicted from their pre-task knowledge levels, this would be hard to do in the dependent task.

Pre-task knowledge was found to negatively correlate with post-task difficulty (when both tasks were combined, and in the parallel task), but not with other variables. It is reasonable to think that the more knowledge users had with their tasks before working with them, the less difficult they found their tasks were. Meanwhile,

one can notice that this happened in general (both tasks combined) and in the parallel task, but not in the dependent task. Again, this seems to indicate that the users' perception of the difficulty of the dependent task, just as users' knowledge after working with the task, could not be easily predicted by the users' baseline knowledge level of the task.

On the other hand, post-task knowledge was found to be correlated positively with task success (when both tasks were combined, and in both individual tasks) and positively with task accomplishment (when both tasks combined, and in both individual tasks). It seems reasonable that if users had higher levels of knowledge after their tasks, they would also feel they were more successful in gathering information for the task, and were more satisfied with their task reports. Meanwhile, one wants to note that the post-task knowledge was not correlated with the post-session assessment of task difficulty.

4.3.2 Sub-task knowledge and other variables

As we did for the general task knowledge, we also looked at the relationships between the pre- and the post-session sub-task knowledge levels and other variables that were elicited through questionnaires, including:

- Sub-task difficulty: how difficult was it to find information users needed for the sub-task.
- Sub-task success: how successful users think they were in gathering the information needed for the sub-task.
- Sub-task satisfaction: how satisfied users were with their reports submitted for the sub-task.

Results (Table 12) show that users' pre- and post-session sub-task knowledge did not correlate in session 1, but they positively correlated in session 2 and 3, and when all 3 sessions combined, except for in the dependent task in session 3. These means that in the beginning session, users' knowledge level of the sub-task after working with it was not likely to be predicted by their

Table 13 Correlations between post-session sub-task knowledge and other variables

	Post-session sub-task knowledge											
	Session 1			Session 2			Session 3			3 sessions combined		
	Both tasks	DT	PT	Both tasks	DT	PT	Both tasks	DT	PT	Both tasks	DT	PT
Post-session difficulty	-.237 (.264)	-.246 (.442)	-.307 (.332)	-.293 (.165)	-.348 (.267)	-.084 (.794)	-.488* (.016)	-.483 (.111)	-.614* (.030)	-.317** (.007)	-.284 (.094)	-.408* (.014)
Sub-task success	.043 (.843)	-.296 (.350)	.397 (.201)	.343 (.101)	-.070 (.830)	.614* (.034)	.616** (.001)	.676* (.016)	.526 (.079)	.338** (.004)	.169 (.324)	.533** (.001)
Sub-task satisfaction	.019 (.929)	-.151 (.640)	.134 (.678)	.091 (.671)	-.132 (.682)	.439 (.154)	.372 (.074)	.330 (.295)	.464 (.128)	.219 (.064)	.069 (.688)	.382* (.021)

** correlation is significant at p<0.01 level (2-tailed); * correlation is significant at p< 0.05 level (2-tailed).

baseline knowledge with the sub-task, but in later sessions, the post-session knowledge levels were likely to be predicted by the pre-session one. Note that these results did not necessarily conflict with the above-reported results that users’ pre- and post-session knowledge with sub-tasks across 3 sessions did not have differences. One possible explanation could be that users were getting familiar with the general task and the format of the subtasks in later sessions, so that their post-session knowledge was in alignment with their baseline knowledge levels: those with higher baseline sub-task knowledge also had higher post-session knowledge, and vice versa.

Table 12 indicates that pre-session sub-task knowledge level was not correlated with other variables, except for being negatively correlated with sub-task difficulty in the parallel task in session 3 and in all 3 sessions combined. Results on the relationship between post-session knowledge and other variables (Table 13) show that post-session sub-task knowledge level correlate negatively with post-session difficulty in session 3 (both tasks combined and in the parallel task) and when 3 sessions are combined (both tasks), positively with sub-task success in session 2 (in the parallel task), session 3 (both tasks combined and the dependent task), and 3 sessions combined (both tasks combined and the parallel task), positively with sub-task satisfaction when 3 sessions combined (in the parallel task). No correlation was found between post-session sub-task knowledge and other variables in session 1. This

was consistent with the findings for the pre-session sub-task knowledge.

4.4 Knowledge change

4.4.1 General task knowledge change There were two ways in this study to measure users’ knowledge change. One was a direct measure: at the end of the 3rd session, participants were asked to self-rate, on a 7-point scale, how much they had learned on the general task in the whole process of the experiment. We call this the *perceived knowledge change*. Examination of the data showed that users’ perceived knowledge change in the dependent task varied between scores 5 and 7, with a mean of 5.75 and an SD of 0.87. Users’ perceived knowledge change in the parallel task varied between 4 and 7, with a mean of 5.42 and an SD of 1.00. When both tasks were combined, users’ perceived knowledge change varied between 4 and 7, with a mean of 5.58 and an SD of 0.93.

The other way to look at users’ knowledge change was the absolute difference between users’ rating scores of the pre- and the post-session general task topic familiarity ratings. We call this the *absolute knowledge change*, calculated as follows:

$$\text{Absolute knowledge change} = \text{post-session topic familiarity rating} - \text{pre-session topic familiarity rating}$$

Examination of the data showed that users' absolute knowledge change in the dependent task varied between scores 2 and 5, with a mean of 3.25 and an SD of 1.06, and that in the parallel task varied between -1 and 4, with a mean of 2.25 and an SD of 1.49. When both tasks were combined, users' absolute knowledge change varied between -1 and 5, with a mean of 2.75 and an SD of 1.36.

While one may think that the above two variables were both measuring users' knowledge change and would correlate with each other, our examination showed that they did not have statistically significant correlation, either in individual tasks, or when both tasks were considered together (Table 15). This suggests that care should be taken when representing users' knowledge change using either one of these two ways.

4.4.2 Sub-task knowledge change In order to avoid making users confused and spending too much effort, our experiment did not ask the participants to self-assess how much they had learned on the sub-task after each session, as we did for the general task knowledge. Therefore, in this study, there was not a measure of *perceived knowledge change* for the sub-tasks. However, we did ask users how much knowledge they had of the sub-tasks both before and after each session, so we were able to compute users' *absolute sub-task knowledge change*, as we did for the absolute general task knowledge.

Examination of the data shows that users' *absolute sub-task knowledge change* in the dependent task varied between scores of -3 to 6, with a mean of 2.17 and an SD of 1.54. We noticed that one user had lower post-session knowledge than pre-session, and that this happened in session 1. In the parallel task, the absolute sub-task knowledge change scores varied between -5 to 5, with a mean of 1.94 and an SD of 1.87. Again, one user had lower post-session knowledge than pre-session, and this also happened in session 1. When both tasks were combined, user ratings varied between -1 to 6, with a mean of 2.06 and an SD of 1.70.

4.5 Relation between knowledge change and other variables

4.5.1 General task knowledge change and other variables We further examined the correlations between perceived knowledge change, absolute knowledge change, and other variables that were elicited in the questionnaires. Results (Table 14) show that while absolute knowledge change and perceived knowledge change were not correlated with each other (as stated in the previous subsection), they were correlated with different sets of other variables.

Specifically, perceived knowledge change was found to correlate positively with pre-task knowledge (in the dependent task), positively with post-task knowledge (in both tasks combined, and in individual tasks), positively with task success (in both tasks combined, and in the parallel task), and positively with task satisfaction (in both tasks combined, and in individual tasks). On the other hand, absolute knowledge change was only found to correlate negatively with pre-task knowledge (in both tasks combined, and in individual tasks), and positively with post-task difficulty (in the parallel task).

These results demonstrate that the perceived knowledge change, which was a retrospective estimate of one's knowledge gain after the task, mostly had correlations with variables that were elicited after the task was completed. The higher the self-rated post-task knowledge was, the higher the perceived knowledge change. Also, the more success users had and the more satisfied they were with the task, the higher the perceived knowledge change. Surprisingly, users' post-task difficulty in either task, or when both tasks were combined, did not show significant correlation with their perceived knowledge change. This indicates that users' assessment of the task difficulty was not necessarily affected by how much they felt they had learned.

On the other hand, the absolute knowledge change, which was the difference between the pre- and the post-task knowledge ratings, was mainly correlated with the pre-task knowledge ratings. It seems reasonable that the higher the pre-task knowledge level, the less the knowledge

Table 14 Correlations of perceived knowledge change, absolute knowledge change, and other variables

	Perceived knowledge change			Absolute knowledge change		
	Both tasks	DT	PT	Both tasks	DT	PT
Absolute knowledge change	.362 (.082)	-.025 (.939)	.538 (.071)	---	---	---
Pre-task knowledge	.201 (.345)	.685* (.014)	.059 (.855)	-.752** (.000)	-.642* (.024)	-.765** (.004)
Post-task knowledge	.779** (.000)	.787** (.002)	.777** (.003)	.219 (.303)	.407 (.190)	.092 (.776)
Post-task difficulty	.000 (1.00)	-.100 (.757)	-.042 (.896)	.325 (.121)	-.181 (.574)	.663* (.019)
Task success	.502** (.013)	.373 (.232)	.621* (.031)	.134 (.532)	.265 (.405)	.067 (.836)
Task satisfaction	.619** (.001)	.632* (.028)	.612* (.035)	.187 (.381)	.150 (.641)	.176 (.585)

** correlation is significant at p<0.01 level (2-tailed).

* correlation is significant at p<0.05 level (2-tailed).

Table 15 Correlations between absolute sub-task knowledge change and other variables

	Absolute sub-task knowledge change											
	Session 1			Session 2			Session 3			3-sessions combined		
	Both tasks	DT	PT	Both tasks	DT	PT	Both tasks	DT	PT	Both tasks	DT	PT
Pre-session knowledge	-.694** (.000)	-.721** (.008)	-.671** (.017)	-.837** (.000)	-.910** (.000)	-.827** (.001)	-.749** (.000)	-.708** (.010)	-.811** (.001)	-.704** (.000)	-.727** (.000)	-.689** (.000)
Post-session knowledge	.667** (.000)	.559 (.059)	.720** (.008)	-.069 (.750)	-.273 (.391)	-.018 (.955)	.140 (.514)	.394 (.205)	-.172 (.592)	.389** (.001)	.386* (.020)	.409* (.013)
Post-session difficulty	-.041 (.850)	-.121 (.708)	-.030 (.927)	.025 (.907)	-.071 (.826)	.291 (.359)	-.104 (.628)	-.348 (.267)	.294 (.354)	-.046 (.702)	-.171 (.318)	.071 (.680)
Sub-task success	.141 (.510)	-.186 (.563)	.491 (.105)	.235 (.269)	.361 (.249)	.181 (.572)	.147 (.494)	.152 (.636)	.183 (.568)	.189 (.112)	.044 (.798)	.350* (.037)
Sub-task satisfaction	.104 (.628)	-.071 (.827)	.205 (.522)	.313 (.136)	.325 (.303)	.363 (.246)	.195 (.361)	.167 (.603)	.371 (.235)	.219 (.064)	.106 (.538)	.339* (.043)

** correlation is significant at p<0.01 level (2-tailed); * correlation is significant at p< 0.05 level (2-tailed).

change was, and vice versa. In addition, in the parallel task, users’ knowledge change was found to be positively correlated with post-task difficulty. Those users in the parallel task who rated it more difficult had more differences between their ratings for their knowledge levels before and after the task, but they did not think that they had learned more (no correlation between post-task difficulty and perceived knowledge change).

4.5.2 Sub-task knowledge change and other variables Table 15 displays the correlations between users’ absolute sub-task knowledge change and other variables in both tasks combined and in individual tasks, in each session, as

well as in the 3 sessions combined. Results show that users’ absolute sub-task knowledge change was negatively correlated with pre-session sub-task knowledge level, and this was true in all sessions and in both tasks. It was also found that this knowledge was positively correlated with post-session sub-task knowledge when all 3 sessions were combined (both tasks combined, and in individual tasks), and in session 1 (both tasks combined and the parallel task). In addition, users’ absolute sub-task knowledge change was positively correlated with task success and task satisfaction in the parallel task when 3-sessions were combined, and positively correlated with pre-task

difficulty in session 2 when both tasks were combined.

In general, these findings were consistent with those for the general task knowledge change as shown in Table 14. Users' absolute knowledge change for sub-tasks was mainly correlated, negatively, with pre-session sub-task knowledge. The greater the users' pre-session sub-task knowledge, the less their absolute knowledge changed. In the cases when absolute knowledge change was correlated with post-session sub-task knowledge, the correlation was positive, meaning that the greater the users' post-session sub-task knowledge, the greater their changes. In a few cases, users' absolute knowledge change was positively correlated with users' post-session assessment of their task success and task satisfaction. This did not happen much, which was in alignment with the similar findings for the absolute knowledge change for the general task, as reported in Table 14.

5. DISCUSSION

5.1 Knowledge increase and the ceiling effect

It makes intuitive sense that users gained knowledge on a task after searching for information and working on it. Our results demonstrate that in general, users' knowledge of the general tasks and that of the sub-tasks increased after a task session. However, there were exceptions. For the general task, in session 3, users did not show significantly greater knowledge after the session than before. This indicates that users may have reached a plateau in their knowledge of the general task after two sessions' working on it. A closer look at individual tasks found that this did not happen in the dependent task, but it did in the parallel task, in session 3, as well as session 1. In session 3, users even showed descriptively equal levels of knowledge before and after the session. One can see that users' general knowledge in the parallel task was relatively high in session 3, with a mean rating score of 5.42 on a 7-point scale. Again, the possible explanation could be that users could have gained a high enough level of knowledge on the parallel task

after two sessions' work, although not on the dependent task. With regard to the case in session 1, one can notice that users showed a pre-session knowledge level that was not very low, with a mean of 3.17 (in comparison, users' pre-session knowledge in the dependent task was only 2.33). This could possibly lead to a no-significant difference with their post-session knowledge level, which did not turn out to be very high (a mean score of 4.33).

These findings seem to indicate that in general, users gained knowledge after a working session, but if they had relatively high levels of baseline knowledge, they may not achieve significantly higher levels of knowledge afterwards. This has implications for information retrieval task design in that tasks in research studies should be designed to avoid the ceiling effect, especially when examining the effects of user knowledge.

Following these findings, it would be interesting to explore the relationship between this no-change in self-assessed pre- and post-session knowledge levels and users' interaction with systems, and see what systems could do to help users gain more knowledge, or better accomplish their tasks, even though they did not think they had gained knowledge, or when they already had relatively high levels of knowledge. These will be done in future studies.

5.2 Knowledge and user perceptions of task attributes and accomplishment

Our results show that knowledge was correlated with a number of user perceptions of task attributes. There was a tendency that pre-session/task knowledge was correlated with users' perception of task difficulty, and post-session/task knowledge was correlated with users' perceived task success and task satisfaction, although there were exceptions.

Specifically, pre-task knowledge with the general task and with post-task difficulty showed a negative correlation in the parallel task and in both tasks combined, but not in the dependent task. Pre-session sub-task knowledge and post-session difficulty had a negative correlation only

in the parallel task in session 3, or when three sessions were combined. Such a correlation between baseline knowledge and task difficulty assessment could help systems to predict how difficult a task would be to users given their knowledge level, and accordingly the system could provide assistance to the users, especially those with lower levels of knowledge. Meanwhile, further exploration is needed to address why this did not happen in the dependent task.

On the other hand, post-task knowledge with the general task was found to be correlated with users' task success and task satisfaction. This indicates a general tendency that if users had higher degrees of knowledge after the general task, they were likely to have more success and satisfaction with the tasks. As for the post-session sub-task knowledge, although the same relationship was not found in early sessions, it did happen in later sessions and when 3 sessions were combined. We also noticed that in later sessions, users' post-session sub-task knowledge tended to be higher, at least descriptively. So it appears that when users' knowledge levels reached a certain degree, they would feel more success and satisfaction with their tasks. This seems to support the idea that search systems should try to assist users gaining more knowledge in order for them to have higher degrees of task success and satisfaction.

5.3 Pre- vs. post-session/task topic knowledge

It was found that users' pre- and post-session/task knowledge had correlations in some but not all cases. Specifically, users' pre- and post-general task knowledge had a significant correlation with each other when both tasks were combined and a nearly significant correlation in the parallel task, but it was not true in the dependent task. A correlation was also found between the pre- and the post-session sub-task knowledge in all situations except for in session 1 and in the dependent task in session 3. Further exploration will be conducted to look into the exceptions, from aspects such as users' behaviors, about why there was no correlation. The expected findings are hoped to be helpful for sys-

tem design, in the sense that for those users with lower levels of baseline knowledge, systems could help them gain knowledge to be comparable to users with higher levels of baseline knowledge.

5.4 General vs. sub-task topic knowledge

It is clearly shown in our results that in multi-session tasks, users' knowledge of the general tasks and that of the sub-tasks in individual sessions are different variables with different patterns across sessions. Our results show that users' pre-session knowledge of the general task increased along sessions, indicating that in the beginning of later sessions, they retained knowledge that was gained in previous sessions. However, users' pre-session knowledge of the sub-tasks did not have differences in different sessions, demonstrating that they had the same levels of baseline knowledge on each sub-task. Meanwhile, users' knowledge of the sub-tasks increased after each session, however, that of the general tasks could stay unimproved if the users had reached a high enough level of knowledge. A search system may not always help increase users' general task knowledge in their multi-session search tasks, but it still could help increase their sub-task knowledge.

Our results also demonstrate that in some cases, the patterns of users' general task knowledge and that of users' sub-task knowledge were the same. For instance, the relationships between users' knowledge and task difficulty, task success, task satisfaction, as well as those between users' knowledge change and other variables were similar. This indicates that general task knowledge could possibly be attributed to sub-task topic knowledge that was at a more detailed/specific level. Accordingly, search systems doing a better job in increasing users' knowledge in their sub-tasks could possibly do a better job in their overall tasks, which would hopefully lead to users' greater success and satisfaction with their tasks.

5.5 Perceived vs. absolute knowledge change

While users' perceived knowledge change and the absolute knowledge change seemed to both measure users' knowledge change, they were found not to correlate with each other. In addition, the perceived and the absolute knowledge change were found to correlate with different sets of other variables that were elicited before and after sessions.

As one would imagine, the perceived knowledge change, which was elicited after task sessions, was usually normally correlated with users' perceptions on the task, e.g., success, satisfaction, that were also elicited after the session. Meanwhile, the absolute knowledge change was found to mainly correlate with pre-task knowledge ratings. One may want to note the differences between this finding and that of Cole et al. (2010), which found that users' self-rated knowledge level are correlated with their knowledge level elicited by their familiarity with the thesaurus terms. While the direct rating of one's knowledge could possibly represent their knowledge level, the direct rating of one's knowledge change after a task was not correlated with the change in their knowledge ratings before and after tasks. Care should be taken in representing how much the users learned or how much their knowledge changed.

5.5 The effect of task type

There were several aspects where we can see task type showed effects on knowledge, or in other words, differences were shown in different task types. The above discussion has addressed the points that task type affected the relationships between users' knowledge level and their perceptions with task difficulty, task success, and task satisfaction, as well as the relationships between knowledge change with users' perceptions on those factors.

In addition, results show that users' baseline knowledge of the general tasks showed differences in different tasks. Specifically, those working on the parallel task had significantly higher pre-session general task knowledge ratings than those on the dependent task. However, the post-session general task knowledge of users working

on the two tasks did not have differences. It would be interesting to examine in future studies users' search behaviors in both tasks, and explore how and why they ended up with equal levels of knowledge despite of the differences in their baseline knowledge.

The task type in the current study focused on one task attribute dimension of task structure. Our future studies will look at more task types that vary along other dimensions, for example, tasks with different levels of difficulty and complexity, factual vs. intellectual tasks, etc., to examine the effect of task type on knowledge and knowledge change in information tasks.

6. CONCLUSIONS

We found in our study that users' knowledge generally increased after sessions, but there were exceptions in which a "ceiling" effect was shown. Knowledge was found to correlate with users' perception of task attributes and accomplishment: pre-session/task knowledge was correlated with users' perception of task difficulty, and post-session/task knowledge was correlated with users' perceived task success and task satisfaction, although there were exceptions. Although having different changing patterns across sessions, some attributes of the general task knowledge could possibly be attributed to those of the sub-task knowledge. We also found that users' perceived knowledge change and their absolute knowledge change were not correlated with each other. In addition, task type was found to affect several aspects of knowledge levels and knowledge change.

These findings help us understand information searchers' knowledge change and are also beneficial for information retrieval research and system design. For example, in designing knowledge domain information visualization systems (e.g., CiteSpace⁴), a reliable prediction of users' knowledge levels and how their knowledge changes in different situations could help improve the effectiveness, efficiency and

⁴ <http://cluster.cis.drexel.edu/~cchen/citespace/>

usability of the system. In designing personalized or generic information systems, the accurate prediction of users' knowledge level and how it changes in users' different searching stages would also assist designers to produce systems that can provide better services for the current users.

As any research, this study has its limitations. First, the study used self-rated topic familiarity to assess knowledge, which was mainly subjective. One's knowledge does not usually drop but at times the self-evaluation of one's knowledge could. Future studies will employ other knowledge assessment methods, such as test-based measures, to assess user knowledge. Second, the study used only two tasks. Although differing in task structure, they were both the category of informational tasks (Kellar et al., 2007) or intellectual tasks (Li & Belkin, 2008), which restricts the generalizability of our findings. Other task types, such as transactional (Kellar et al., 2007), factual tasks (Li & Belkin, 2008), or others as defined in that task classification scheme, may also need to be examined for users' knowledge changes in the search processes. These will be conducted in future studies. Third, user engagement is also another factor that may have influenced users' knowledge gain or retention. Future studies will elicit users' engagement levels and examine if it has any effects.

Despite these limitations, our study has important findings on users' knowledge change during the task completion process, as well as on the relationship between users' knowledge change and task attributes. The results are important in confirming some assumptions about knowledge increase during search and task performance that have been assumed by many researchers, and contributes to understanding how users' knowledge level influences search behavior.

7. ACKNOWLEDGEMENTS

This research was supported by IMLS grant LG#06-07-0105-07.

8. REFERENCES

- Allen, B. (1991). Topic knowledge and online catalog search formulation. *Library Quarterly*, 61(2), 188-213.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133-143.
- Brantmeier, C. & Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System*, 36(3), 456-477.
- Brookes, B. (1980). The foundations of information science, Part I: Philosophical aspects. *Journal of Information Science*, 2(3-4), 125-133.
- Byström, K. (2002). Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology*, 53(7), 581-591.
- Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing and Management*, 31(2), 191-213.
- Cole, M. J., Zhang, X., Liu, J., Liu, C., Belkin, N. J., Bierig, R., and Gwizdka, J. (2010). Are self-assessments reliable indicators of topic knowledge? In *Proceedings of the American Society for Information Science & Technology (ASIS&T) 2010*.
- Donato, D., Bonchi, F., Chi, T., & Maarek, Y. (2010). Do you want to take notes? Identifying research missions in Yahoo! Search Pad. In *Proceedings of WWW 2010*.
- Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In *Proceedings of CHI 2008*, 39-48.
- Hembrooke, H. A., Granka, L. A., Gay, G. K., & Liddy, E. D. (2005). The effects of expertise and feedback on search term selection and subsequent learning. *Journal of the American Society for Information Science & Technology*, 56(8), 861-871.
- Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science & Technology*, 44(3), 161-174.
- Kammerer, Y., Nairn, R., Pirolli, P., & Chi, E. H. (2009). *Signpost from the masses: learning effects in an exploratory social tag search browser*. In *Proceedings of the 27th international conference on Human factors in computing systems (CHI'09)* (pp. 625-634), Boston, MA

- Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science & Technology*, 58(7), 999-1018.
- Kelly, D. (2006). Measuring online information-seeking context, part 1. Background and method. *Journal of the American Society for Information Science & Technology*, 57(13), 1729-1739.
- Kelly, D., Kantor, P., Morse, E., Scholtz, J., & Sun, Y. (2006). User-centered evaluation of interactive question answering systems. In Proceedings of the Interactive Question Answering Workshop at HLT-NAACL 2006 June 9, 2006, New York, New York, 49-56. Stroudsburg, PA: Association for Computational Linguistics.
- Kelly, D. (2006). Measuring online information-seeking context, part 2. Findings and discussion. *Journal of the American Society for Information Science & Technology*, 57(14), 1862-1874.
- Kelly, D., & Cool, C. (2002). The effects of topic familiarity on information search behavior. *Proceedings of ACM and IEEE Joint Conference on Digital Libraries (JCDL) 2002*, 74-75.
- Kim, J. (2006). Task difficulty as a predictor and indicator of web searching interaction. In *Proceedings of CHI 2006*, 959-964.
- Kuhlthau, C.C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science & Technology*, 42, 361-371.
- Li, Y. & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, 44, 1822-1837.
- Liu, J. & Belkin, N. J. (2010). Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In *Proceedings of ACM Special Interest Group on Information Retrieval (SIGIR) 2010*, 26-33.
- Liu, J., Cole, M., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., Zhang, J., & Zhang, X. (2010). Search behaviors in different task types. In *Proceedings of ACM and IEEE Joint Conference on Digital Libraries (JCDL) 2010*.
- Liu, J., Gwizdka, J., Liu C., & Belkin, N.J. (2010). Predicting task difficulty for different task types. In *Proceedings of the American Society for Information Science & Technology (ASIS&T) 2010*.
- Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science & Technology*, 40(1), 54-66.
- Sihvonen, A., & Vakkari, P. (2004). Subject knowledge improves interactive query expansion assisted by a thesaurus. *Journal of Documentation*, 60(6), 673-690.
- Spink, A. (1996). A multiple search session model of end-user behavior: An exploratory study. *Journal of the American Society for Information Science*, 47(8), 603-609.
- Toms, E., MacKenzie, T., Jordan, C., O'Brien, H., Freund, L., Toze, S., et al. (2007). How task affects information search. In *Workshop Pre-proceedings in Initiative for the Evaluation of XML Retrieval (INEX)*, 337-341.
- Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal research. *Information Processing & Management*, 39(3), 445-463.
- White, R. W., Dumais, S., & Teevan, J. (2009). Characterizing the influence of domain expertise on Web search behavior. In *Proceedings of Web Searching and Data Mining (WSDM) 2009*, 132-141.
- Wildemuth, B. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science & Technology*, 55(3), 246-258.
- Wilson, M. J. & Wilson, M. L. (in press). A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. To appear in *Journal of the American Society for Information Science & Technology*.
- Zhang, X., Angheliescu, H. & Yuan, X. (2005). Domain knowledge, search behavior, and search effectiveness of engineering/science students: An exploratory study. *Information Research*, 10(2).