The Effect of Cognitive Abilities on Information Search for Tasks of Varying Levels of Complexity

Kathy Brennan, Diane Kelly, Jaime Arguello
School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC, 27599-3360
[kbrennan, diane.kelly, jarguello] @ unc.edu

ABSTRACT

Although web search engines are designed as one-size-fits-all tools, people do not come in one size, but instead vary across many different attributes. One such attribute is cognitive ability. Because information search is primarily a cognitive activity, understanding the extent to which variations in cognitive abilities impact search behaviors and outcomes is especially important. We describe a study in which we explore how people's cognitive abilities affect their search behaviors and perceptions of workload while conducting search tasks with different levels of complexity. Twenty-one adults from the general public completed this study. We assessed participants' associative memory, perceptual speed, and visualization abilities and also measured workload. To evaluate the relationship between cognitive ability, task complexity and workload, we conducted three separate mixed factor ANOVAs corresponding to each of the abilities. Our results suggest three important trends: (1) associative memory ability had no significant effect on search behavior and workload, (2) visualization ability had a significant effect on search behavior, but not workload, and (3) perceptual speed had a significant effect on search behavior *and* workload. Specifically, participants with high perceptual speed ability engaged in more search activity in less time and experienced less workload. While the interactions were not significant, the differences were more pronounced for more complex tasks. We also found a significant relationship between task complexity and workload, and task complexity and search behaviors, which corroborates previous research.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval: Search Process.

General Terms

Experimentation, Human Factors

Keywords

Information search, cognitive abilities, user study, search behavior, workload, individual differences

IIiX'14, August 26-29 2014, Regensburg, Germany

ACM 978-1-4503-2976-7/14/08…$15.00.
http://dx.doi.org/10.1145/2637002.2637022

1. INTRODUCTION

Successful information search depends upon effective interactions between the user and the system. Effective interactions can take place only to the extent that both the user and system have the abilities to perform whatever tasks might be necessary to achieve the desired search outcomes. While the abilities of the system are programmable and predictable, those of the user are not. Understanding the complexities of human abilities as they relate to information search presents a great challenge for interactive information retrieval (IIR) research.

In its most general form, search requires the user to estimate a search strategy based upon a perceived information need. The information need can be ambiguous [8], and often changes throughout the course of the search as information is discovered, assessed, and discarded or retained [7]. As the user's knowledge changes as he/she engages in cognitive processing [30, 38], the original information need changes also, requiring the user to adjust his or her search strategy to accommodate the new need [43]. It is also generally accepted that the search process will require the user to exert mental effort in varying degrees depending on individual characteristics [22, 26], the search tasks and their structures [12, 48], and the information system [10].

The intellectual processes engaged during search are controlled by the cognitive abilities possessed by the user. Cognitive abilities are comprised of higher mental functions such as reasoning, remembering, understanding and problem solving [15]. Several cognitive abilities have been studied in information science, including memory [50], perceptual speed [1, 2, 35] and visualization [13, 18], yet none of these studies has investigated the ways in which these abilities impact the user's perception of workload or interact with task complexity. While there have been studies examining these elements in part, there has not been a study that has examined all of these elements and their relationships with each other in one research design. In this study, we investigate how users' different cognitive abilities affect their search behaviors and perceptions of workload while conducting search tasks of varying levels of complexity. If variations in cognitive abilities are found to impact search behavior, workload and search outcomes, this might lead to the design of search interfaces or information literacy instruction that are tailored to individual users' cognitive strengths. Furthermore, understanding if and how variations occur according to task complexity, might allow for the development of task-specific interventions.

2. RELATED WORK

This section describes four areas of prior work related to our study: individual differences, cognitive abilities, workload and task complexity.

2.1 Individual Differences

Cognitive abilities have primarily been investigated in information science under the umbrella of individual difference research. While a great deal of research has studied how individual differences, such as gender and cognitive style affect search behavior, fewer studies have focused on cognitive abilities. For example, Ford, Miller, & Moss [22] focused on cognitive *style* (among other differences such as prior knowledge, Internet perceptions, age, and gender). In their work they defined cognitive styles in three ways: wholist-analyst bias, verbalizer-imager bias and cognitive complexity. They found that the verbalizer-imager style was correlated with information retrieval effectiveness. That is, participants with verbalizer styles exhibited poor retrieval performance whereas those participants with imager styles demonstrated strong retrieval performance. Another study of individual differences was that of Palmquist and Kim [42], in which the researchers looked at predictive indicators of web search performance that were related to cognitive style (field dependent and field independent). They found that cognitive style had a strong influence on novice online searchers, but not as much effect on people who had experience with online database searching. In another study, Borgman [10] sought to understand search performance differences based on end-users' academic disciplines. The researcher found strong relationships between characteristics of technical aptitudes and IR performance to academic orientation and also between characteristics of personality and IR performance to academic orientation. This was especially the case when people who had changed academic majors were taken out of the analysis. This study evolved from previous work [11] where the researcher found that science and engineering majors were more likely to pass a benchmark search test after training than social science and humanities majors.

In general, the individual difference research shows that these types of variations impact user behavior and success. However, less is known about how cognitive abilities impact search. One interesting exception is the work of MacFarlane, et al. [36] who, while not studying cognitive ability, studied the effects of cognitive impairment on search. They developed a logging tool based on a dyslexia-related cognitive profile and were able to distinguish the differences between the dyslexic and non-dyslexic searchers. The non-dyslexic users engaged in more search interactions, searched more quickly, and viewed more results and relevance pools (i.e., document sets relevant to a topic). In their work they suggested that slower reading speeds might have contributed to a memory deficit that hindered the dyslexic users from being able to absorb information during the search session.

2.2 Cognitive Abilities

In this study, we focus on three cognitive abilities: associative memory, perceptual speed and visualization ability. Each is described in more detail below.

2.2.1 Associative Memory

While working (short-term) memory is of central importance in search tasks, intermediate memory comes into play as search tasks become more complex. Intermediate memory engages when a person intentionally or actively remembers specific information. It is also called associative memory and is defined as "the ability to recall one part of a previously learned but otherwise unrelated pair of items when the other part of the pair is presented" [21]. To our knowledge, there have been no studies in the field that investigate the role of associative memory on both search behaviors and mental workload.

Several studies have looked at the effect of individual differences, working memory and search behavior. Building upon their previous work in [36], MacFarlane, et al. [37] further examined the search behaviors of dyslexic and non-dyslexic students. On average, non-dyslexic users judged more documents irrelevant than the dyslexic users (41.1 for non-dyslexic versus 27.1 for dyslexic). This difference between the two groups was significantly correlated with phonological working memory test scores, in which a digit span subtest [49] was used. Gwizdka [25] studied the effect of memory span and verbal closure on the formulation of queries using word tags and assessment of search results in the comparison of two different interface styles (list view and tag cloud). There were no main effects for cognitive abilities; however, there were performance differences between the high and low memory span individuals for behavior with word tags (new word tagging differences were significant while word tag deletions were borderline significant). Westerman et al. [50] investigated the impact of cognitive abilities (including several types of memory) and age as predictors of user performance in an hierarchical database in which node selection method was manipulated (i.e., explicit versus embedded menu). While this study found significant effects for both visualization and memory, the equipment and technology of the experiment is now outdated. Findings of these three studies suggest memory is an important factor in search.

2.2.2 Perceptual Speed

Perceptual speed determines a person's ability to efficiently view and identify differences and similarities, patterns, and anomalies when conducting tasks involving symbols and figures [15]. It draws on the ability to scan information effectively, make choices for response and is said to be related to automatic mental processes [19]. Studies have investigated perceptual speed and its effect on user effectiveness and user satisfaction in the context of TREC search tasks [1], its effect on learning and search performance in searching bibliographic abstracts of an experimental IR system [2], and its effect on web search behaviors on four different search engines [35]. For example, in Al-Maskari and Sanderson [1], users with high perceptual speed spent significantly less time finding the first relevant document when conducting TREC search tasks. Results were not significant, however, for number of relevant documents found. A much earlier study by Allen [2] showed that an experimental IR system designed to optimize subject descriptors of bibliographic reference abstracts enabled users with high perceptual speed abilities to exploit that design feature and achieve better search performance than users with low perceptual speed abilities. Importantly, he found that this advantage to the high ability users did not come at a cost to low ability users; there was no difference in performance for the low users for either the optimized information display or the normal display. In one of the only studies of web-based searching conducted to date, Kim and Allen [35] compared undergraduate students' search performance across four different web search engines on a term paper-type task (i.e., search for relevant items to write a term-paper on a given topic) and a news article-type task (i.e., search for relevant items to write a news article on a given topic). While the findings did not support their specific hypothesis, they did find a clear pattern that users with higher levels of perceptual speed and other cognitive abilities showed lower levels of search activity (average time spent, average web sites visited, average number of bookmarks, keyword searches, use of vocabulary suggestion terms, and clicking embedded links) for the newspaper task versus the term paper task. Users with low perceptual speed showed the opposite pattern. They attributed the strong performance by the high perceptual ability users on the term paper to social norms which exist for high ability students that reinforce behavior related to high achievement. In other words, the high ability students invested more effort than the researchers hypothesized into the term paper task because in real life, their strong achievement with writing term papers is rewarded with good grades. As a result, the individual difference of ability and the social influence of academic rewards combined to create a powerful influence on the behavior of high ability users.

2.2.3 *Visualization Ability*

Visualization refers to ". . . the ability to manipulate or transform the image of spatial patterns into other arrangements" [21]. It requires the ability to situate one's self in relationship to a static object, such as being able to imagine a piece of paper in its various stages from being folded to the end, being completely unfolded. It therefore also requires the ability to think sequentially, with a strategy. Studies investigating the ability of visualization have shown mixed results when determining the relationship of visualization to search performance or other outcomes. In search tasks of a hierarchical database system, Downing, Moore, and Brown [18] found that individuals with low spatial visualization ability took longer to find the first relevant document and also found fewer documents than individuals with higher levels of spatial visualization. Campagnoni and Erlich [13] conducted a small-scale user study (N=9) in which they found that spatial abilities related to memory and visualization were highly correlated with navigating and editing screen-text. Pak, Rogers, and Fisk [41], on the other hand, found mixed results for different types of relationships between visualization abilities and search performance under the task and display conditions in their study. They investigated the influence of visualization and other abilities on performance under conditions in which the presentation of display results was manipulated (one display was a graphical map representation of the path required to get to the correct document and the other was a bulleted list of instructions of the same information). There results indicated that spatial orientation was significantly related to search performance in the navigationally-intense condition (i.e., map condition). Swan and Allan [47] found that visualization ability was not related to users' success using a 3-D interface.

2.3 Workload

The term "workload" is used in different ways throughout different research communities. In some cases it is used to mean a cognitive construct, such as cognitive load, mental workload, or cognitive effort. In other cases, when combined with the physical effort of a task, it may be used to mean task workload. The lack of precise terminology for the name of this construct can create confusion [54] and so we have done our best to clearly describe the construct we are measuring. The amount of workload an individual experiences during a task is a result of that individual's working memory capacity [6], cognitive abilities [15], and the context of the person's situation [32]. Search activities that contribute to the user's workload include formulating and re-formulating queries, evaluating search results, viewing and selecting relevant documents, and navigating web pages and sites (see [24] for a review).

The measurement of workload has been used to understand search tasks in different ways by IIR researchers. In a study of visual perception and cognitive speed, Haapalainen, et al. [27] measured mental demand to verify the levels of workload in their study. They combined workload scores with pilot test results as well as task performance analysis (i.e., time-on-task) to determine the success of their manipulation of task difficulty. Di Stasi, et al. [17] compared different search tasks and found that goal-oriented (fact-finding) shopping tasks required more workload than experiential ones (information-gathering).

Workload has also been investigated for other effects it has on users' experiences. For example, it was found that a text- versus visually-based query interface had a significant effect on workload [46]. Information presentation format has been found to affect users' decision-making performance when using electronic versus paper information presentation formats [34]. The workload imposed by specific interactions has also been studied; for example, in the context of tagging [23, 33]. Santos et al. [44] measured the workload of an experimental medical information system that was designed to adapt to users' queries in a medical information application. They found that the experimental system imposed no extra demands on user workload than the standard system. Schmutz et al. [45] found significant correlations among mental demand, primary task completion time, and general user satisfaction in an experiment comparing four online bookstores.

2.4 Task Complexity

Task complexity, as opposed to task *difficulty*, is an inherent property of the task and is independent of the task doer. Different characterizations of task complexity have been proposed. Early work by Wood viewed task complexity as depending on the number of desired outcomes, the number of actions required to produce the outcomes, and the quality of the information cues processed during the task [51]. Campbell later characterized task complexity as depending on the number of desired outcomes, the number of paths to achieving the outcomes, the degree of uncertainty about the paths and outcomes, and the interdependence between outcomes [14]. Byström and Järvelin defined task complexity as a function of the *a priori* determinability about the outcomes, the information requirements, and the processes associated with the task [12]. Similarly, Bell and Ruthven defined task complexity as a function of the *a priori* determinability of the required information, the search strategy, and the ability to judge relevance [9]. Finally, Jansen et al. [31] (and later Arguello et al. [5] and Wu et al. [53]) defined task complexity in terms of the amount of cognitive effort and learning required to complete the task. To this end, they adopted a taxonomy of learning outcomes originally proposed by Anderson and Krathwohl [3] for characterizing educational materials. In this work, we used search tasks that were created using this *cognitive* view of task complexity. Prior work found that task complexity affects search behavior [31, 53] and post-task assessments of task difficulty [4, 53]. In this work, we examine the effects of task complexity (in conjunction with three cognitive abilities: associative memory, perceptual speed and visualization) on mental workload and search behaviors, which to our knowledge, has not been done in a single study.

3. METHOD

3.1 Participant Recruitment

Our interest in the relationship between cognitive abilities and information search necessitated recruiting participants from our local community, rather than relying on university students, to increase our chances of obtaining a sample with diverse cognitive abilities. In addition, most laboratory studies of information search behavior have been conducted with university students, and we wanted to add to the growing body of research that focuses on the search behaviors of the general public. To achieve these goals, we recruited participants from a nearby city and made arrangements with the local public library to conduct the study on their premises in a private study room. We recruited participants by posting fliers at the public library and its northern branch, a nearby technical community college, and several small businesses including coffee and sandwich shops. We also used word of mouth advertising by encouraging participants to refer their friends and family. We recruited a total of 21 participants for the study.

3.2 Demographic Questionnaire

A demographic questionnaire was used to gather information about participants and their search experiences. We collected data about age, education, race, and sex of the participants, as well as information about their use of computers, the Internet and search engines. In addition to these main demographic questions, a set of search self-efficacy questions were included to gauge participants' perceptions of their search abilities. Participants were asked to rate on a scale of 1 to 10, where 1 meant "totally unconfident" and 10 meant "totally confident," their confidence in their abilities to conduct activities related to online search. Items asked about participants' abilities to correctly develop search queries, use Boolean syntax, evaluate results lists, find sufficient number of articles for their queries, efficiently structure their time to complete a search task, and distinguish between relevant and irrelevant articles. These items were based on a previously developed scale of search self-efficacy [16] and have been shown to have high internal consistency in our previous work [40].

3.3 Cognitive Abilities Tests

Participants' associative memory, perceptual speed, and visualization abilities were assessed using tests from the Ekstrom Kit of Factor-referenced Cognitive Tests [20]. For each of the three factors there was a practice test and two actual tests. All tests were timed, with a 30 second verbal warning before time expired. The order of the tests was rotated across participants. The instructions for each test and practice test were read out loud while participants followed along on a print copy.

To evaluate associative memory, we used the Picture-Number Test (MA-1) (7 minutes per test). This memory test required participants to memorize a set of 21 picture-number combinations during an allotted time and then to fill in the numbers for the pictures, which were located in different places on the following page. To assess perceptual speed, we used the Number Comparison Test (P-2) (1.5 minutes per test). In this test, the participant was asked to compare sets of numbers, marking sets in which the two strings differed. This test measures speed and accuracy. Visualization ability was evaluated using the Paper Folding Test (VZ-2) (3 minutes per test). In this test, participants were asked to identify the correct diagram of the unfolded piece of paper that had been previously folded and hole-punched. The total time for the cognitive abilities tests was about 28 minutes, including time for reading the instructions aloud, administering the practice tests, and administering the actual tests.

Tests were scored using guidance provided in the Ekstrom Kit manual and correspondence with the Educational Testing Service in Princeton, NJ. For MA-1 (associative memory), correct answers were counted and summed for the two tests, with 21 points being the perfect score for each test. Incorrect answers and answers left blank were not counted. P-2 scores (perceptual speed) were calculated by summing the correctly marked answers plus the answers correctly left blank. Items left unmarked (omits) because the respondent ran out of time were not counted in the score. The VZ-2 scores (visualization ability) were calculated as the number correct minus the fraction of number incorrect. Omits were not counted in the score. Participants' two test scores for each ability were added together resulting in a single score per participant for each ability.

3.4 Search Tasks

The search tasks from Wu, et al. [53] were used in our study. In the original study, twenty search tasks were created representing five levels of complexity (*remember, understand, analyze, evaluate* and *create*) and four domains (Health, Commerce, Entertainment, and Science and Technology). In the current study, we only used the *remember*, *analyze* and *create* tasks from two domains: Entertainment and Science & Technology. *Remember* tasks required a person to find a specific answer, *analyze* tasks required a person to generate a list of items and explain them and *create* tasks were open and required a person to generate a novel solution. Our choice to restrict task type was based on results from the previous study, which showed strong, significant differences in participants' behaviors for task type, especially among the *remember*, *analyze* and *create* tasks. Our choice to restrict the domain was based on a desire to limit the number of search tasks each person had to perform and subsequently the time required to participate. Examples of tasks from the Science and Technology domain are displayed in Figure 1. Tasks were presented to participants in a rotation.

Remember.

You recently watched a show on the Discovery Channel about fish that live so deep in the ocean they're in darkness most or all of the time. This made you more curious about the deepest point in the ocean. What is the name of the deepest point in the ocean?

Analyze.

You recently became involved with a conservation group that picks up trash from local waterways. One of the group members said your work is important because it helps keep pollution out of the ocean. What are some different types of ocean pollutants? What environmental risks are associated with each pollutant?

Create.

After the NASCAR season opened this year, your niece became really interested in soapbox derby racing. Since her parents are both really busy, you've agreed to help her build a car to enter a local race. First, figure out how to build a car. Then identify basic designs you might use and a plan for constructing the car.

Figure 1. Remember, analyze and create tasks from the Science and Technology domain from [53].

3.5 Workload

The NASA-Task Load Index (TLX) was used to measure workload. The NASA-TLX measures the mental, physical, and temporal demands imposed on individuals by work tasks along with individuals' evaluations of their performance, effort and experienced frustration [29]. The TLX has been described as one of the most widely used mental workload measurement scales [39] and a recent study found more than 500 studies citing the original 1988 research article [28]. Participants completed the NASA-TLX after completing each of the six search tasks. Participants used a 20-point scale to indicate their responses from low to high. To create a total workload score, participants' responses to each of the six NASA-TLX items were summed.

3.6 Search Behaviors

Participants' searches were logged with the Lemur Query toolbar. This toolbar logs queries, clicks, and time spent on pages. Measures taken from the log are in Table 1.

Table 1. Measures taken from search logs.

session_length	Session length in minutes
queries	Number of queries
abandons	Number of queries without SERP clicks
query_length	Number of terms per query
serp_clicks	Number of total clicks on search results
urls_viewed	Number of URLs visited
urls/query	Number of URLs visited per query
serp_dwell_time	Average time in seconds between a query and first click on SERP (if any)

3.7 Procedure

We began the study by giving the participant an information sheet explaining the study and voluntary consent, along with several copies of our advertising flier for distribution. After reading the information sheet, the participant filled out a demographic questionnaire. Next, we administered the Ekstrom Kit tests. Following this, participants completed the search tasks. The searches were performed on a Lenovo ThinkPad X201 laptop computer with Windows 7 Enterprise operating system. We included a wired mouse for participants who were not familiar with the touchpad mouse. Participants were given six search tasks, one at a time. For each task, the participant was asked to record his or her answers by typing or cutting and pasting onto a Microsoft Word document that was open on the laptop. Tasks were typed in large print on half sheets of paper, which the researcher read out loud to the participant before each task began. The sheet was then handed to the participant for reference throughout searching and a countdown timer only visible to the researcher was set to 12 minutes. After each search task the participant filled out a paper and pencil version of the NASA Task Load Index (TLX) questionnaire. This process was repeated six times. At the end of the study, the participant wrote his or her name and address on an envelope, which was used to mail the participant the $20 USD study honorarium.

3.8 Participants

A total of 21 participants, who ranged in age from 27 to 70, completed this study. The average age was 45.4 years and the median was 43.5 years. There were 13 females and 8 males. Participants identified their race or ethnicity as black (n=9), white (n=9), Hispanic (ethnicity) (n=2), and Indian (n=1). All held high school diplomas. Four held associate's degrees, three held bachelor's degrees, and three held master's degrees. A range of occupations was represented including bus driver, cashier, day laborer, copywriter, and communication specialist. Four participants were unemployed and one was retired.

All participants reported having at least seven years of computer experience, with most reporting more than 10 (n=19). Most participants reported using computers and the Internet on a daily basis (n=17) with the remainder reporting usage of 2-3 times per week (n=4). All participants said they had regular access to a computer in a number of locations, including school (n=3), work (n=4), home (n=16), or the library (12). Search engine use included Bing (n=10), AOL Search (n=2), Yahoo! (n=3) and Google (n=21). Common tasks performed on the Internet included work (n=9), school (n=8), general searching for information (n=21), watching movies (n=9), browsing/surfing the Internet (n=18), accessing email (n=20), accessing social media (n=19) and managing day-to-day tasks (n=13).

4. RESULTS

4.1 Cognitive Abilities

Table 2 displays the means and standard deviations of the scores for the three cognitive ability tests. To better understand these scores, we compare them to several reference scores provided in the Ekstrom Kit Manual (EKM) that accompanied the tests as well as in a technical report describing their use to evaluate U.S. Air Force (USAF) enlistees [52]. These figures are provided in the last two rows of the table. For associative memory, participants' mean score was similar to that observed in the USAF sample. It was slightly lower than what is reported in the EKM, but these scores were generated using college students. For perceptual speed and visualization, our sample scored lower than the USAF sample, but in both cases they were within one standard deviation of the USAF means. Our sample was within two standard deviations of the EKM sample for visualization.

Table 2. Descriptive statistics of test scores for three sets of cognitive ability tests from the Ekstrom Kit of Cognitive Reference-Factored Tests (2 each of MA-1, P-2, VZ-2) and benchmarks provided in the manual and from an additional report (last two rows of table).

	Associative Memory	Perceptual Speed	Visualization
Possible Range	0-42	0-96	0-20
Mean (SD)	19.57 (11.58)	44.38 (10.58)	6.48 (3.40)
Median	17	44	6.667
Min, Max	0, 39	25, 73	0, 14.78
EKM Mean (SD)	24.40[1] (8.70)	N/A	10.95 (3.70)
USAF Mean (SD)	20.38 (10.19)	47.94 (12.32)	10.17 (4.41)

[1]college student sample

For analysis, we divided participants into low- and high- groups using a median split for each ability. Associative memory scores were divided at 17 (low=0-17, high=18-39). Perceptual speed scores were divided at 44 (low=24-44, high=45-73). Visualization scores were divided at 6.667 (low=0-6.667, high=6.668-14.778). These binary classifications necessarily result in some loss of information, but given our low number of participants, we felt this would allow more reliable analyses.

We examined the correlation coefficients among participants' scores for associative memory, perceptual speed and visualization to make sure that the high and low groupings for each ability did not consist of the same participants. In other words, did participants who scored low on one ability also score low on the other abilities too? We found no significant correlation between visualization and memory scores (r=-0.110, p=0.64) or visualization and perceptual speed scores (r=0.224, p=0.33). However, there was a significant positive correlation between the associative memory and perceptual speed scores (r=0.523, p<.05). To better understand the groupings, we examined each participant's group membership (high or low) for these two abilities and found that only seven participants' scores (and group membership) were correlated. We also computed the correlations between search self-efficacy (averaged across all items) and the different cognitive ability scores and found no significant correlations.

4.2 Effects on Workload

To evaluate the relationship between cognitive ability, task complexity and workload, we conducted three separate mixed model ANOVAs for each of the cognitive abilities. This allowed us to report main effects for each of the three cognitive abilities on workload (each cognitive ability is treated as a between subjects factor), main effects for task complexity (repeated measures, or within subjects factor), and interaction effects. Because the main effects analyses for task complexity were the same across the three ANOVAs, we first report these results and then report the analyses of main effects for cognitive abilities and the interaction effects between cognitive abilities and task complexity. Because of the exploratory nature of this research, for each analysis, we consider participants' responses to each TLX item, as well as the overall score in the hopes that this might provide insight into any differences we observe. Although we did not expect differences in workload scores to vary according to domain, as a check, we ran these analyses using domain instead of task type and found no significant main effects for domain, or interaction effects between domain and cognitive ability group.

4.2.1 Task Complexity Effects

Table 3 displays the means and standard deviations of all items from the NASA-TLX, including the overall workload score for the different types of tasks, and the results of the ANOVA. There is a clear trend with respect to the remember tasks and the other two tasks with remember tasks always receiving lower ratings. With respect to analyze and create tasks, for some items, create tasks received higher ratings, while for others analyze did. The total workload scores for these two types of tasks were nearly the same. Significant main effects were detected for each item and post-hoc tests showed that the significant differences were always between the remember tasks and the other two tasks. We also provide a measure of effect size in Table 3, η^2, which shows the strength of the relationship. Overall, many strong effects were found, most notably with respect to impact of task complexity on mental demand and effort.

4.2.2 Cognitive Ability Effects and Interaction

Table 4 displays the means and standard deviations of the NASA-TLX items according to associative memory group. ANOVAs revealed no main effects for associative memory. The F-statistics are not reported here to conserve space and were all less than 1. A significant interaction effect was detected between associative memory and task complexity on temporal demand (F=5.09, p=0.01), but not for other items (all F-statistics < 1.5).

Table 3. Mean (standard deviation) responses to the NASA-TLX Workload items according to task type, F-statistics (df=2, 21) and effect size.

	Remem.	Analyze	Create	F	η^2
MenD	4.29 (3.52)	8.50 (4.82)	9.40 (5.01)	18.77**	.50
PhyD	2.74 (2.53)	4.93 (4.65)	5.24 (4.53)	5.91**	.24
TemD	3.62 (3.54)	7.31 (4.68)	6.67 (4.69)	8.40**	.31
Perf	3.79 (4.39)	6.41 (4.67)	6.38 (4.71)	3.89*	.17
Effort	3.71 (2.99)	7.81 (4.47)	7.57 (4.57)	10.49**	.36
Frus	3.50 (3.85)	6.93 (4.28)	6.57 (4.43)	5.32*	.22
All	21.64 (17.54)	41.88 (23.78)	41.83 (22.91)	11.28**	.37

***p<0.05; **p<0.01**

Table 4. Mean (standard deviation) responses to the NASA-TLX Workload items according to associative memory group (L=low; H=high).

		Remem.	Analyze	Create	Total
MenD	L	4.68 (5.25)	8.33 (5.68)	8.73 (5.53)	7.23 (5.71)
	H	3.85 (3.79)	8.55 (5.27)	10.15 (6.20)	7.52 (5.77)
PhyD	L	2.86 (3.40)	4.90 (5.20)	4.68 (5.33)	4.14 (4.73)
	H	2.60 (2.58)	4.90 (5.03)	5.85 (5.40)	4.45 (4.64)
TemD	L	4.05 (4.71)	7.57 (5.14)	4.41 (3.62)	5.31 (4.73)
	H	3.15 (3.30)	7.10 (5.46)	9.15 (6.00)	6.47 (5.56)
Perf	L	4.05 (5.21)	7.90 (6.00)	6.95 (5.74)	6.28 (5.79)
	H	3.50 (3.90)	4.85 (4.38)	5.75 (5.19)	4.70 (4.54)
Effort	L	4.18 (4.67)	7.86 (6.21)	6.36 (5.13)	6.11 (5.49)
	H	3.20 (2.24)	8.00 (5.32)	8.90 (6.32)	6.70 (5.48)
Frus	L	4.05 (5.32)	7.71 (5.74)	6.18 (5.20)	5.95 (5.54)
	H	2.90 (4.23)	6.25 (4.89)	7.00 (5.32)	5.38 (5.36)
All	L	23.86 (24.15)	44.29 (30.68)	37.32 (25.28)	35.02 (27.72)
	H	19.20 (14.44)	39.65 (24.70)	46.80 (28.15)	35.22 (25.65)

Table 5 displays the means and standard deviations of the NASA-TLX items according to perceptual speed group. Significant differences were detected for most items according to perceptual speed ability, with participants in the low group experiencing greater demands than those in the high group. The F-statistic and effect size for each item was: mental demand (5.67; 0.23), physical demand (5.09; 0.21), temporal demand (1.41; 0.07), performance (5.94; 0.24), effort (4.44; 0.19), frustration (5.47; 0.22) and overall (6.75; 0.26). All of these were significant at the p<0.05 level, except for temporal demand which was not significant. No significant interaction effects were detected between perceptual speed and task complexity (most F-statistics < 1.2 except for frustration= 2.71).

Table 5. Mean (standard deviation) responses to NASA-TLX Workload items according to perceptual speed group (L=low; H=high). Shaded values are significantly different *p*<0.01.

		Remem.	Analyze	Create	Total
MenD	L	5.45 (5.21)	10.19 (5.20)	11.64 (5.43)	9.08 (5.85)
	H	3.00 (3.45)	6.60 (5.12)	6.95 (5.36)	5.52 (4.98)
PhyD	L	3.64 (3.42)	6.43 (5.02)	7.27 (5.52)	5.77 (4.92)
	H	1.75 (2.15)	3.30 (4.68)	3.00 (4.18)	2.68 (3.83)
TemD	L	4.32 (4.46)	8.19 (4.32)	7.59 (5.13)	6.68 (4.89)
	H	2.85 (3.56)	6.45 (6.04)	5.65 (5.61)	4.98 (5.33)
Perf	L	5.05 (5.69)	8.29 (5.55)	8.32 (5.45)	7.20 (5.69)
	H	2.40 (2.42)	4.45 (4.59)	4.25 (4.70)	3.70 (4.08)
Effort	L	4.09 (3.79)	9.57 (5.54)	9.41 (6.07)	7.66 (5.75)
	H	3.30 (3.66)	6.20 (5.52)	5.55 (4.87)	5.02 (4.83)
Frus	L	3.64 (4.54)	8.24 (5.28)	9.14 (5.63)	6.98 (5.64)
	H	3.35 (5.20)	5.70 (5.19)	3.75 (4.05)	4.27 (4.87)
All	L	26.18 (22.85)	50.90 (26.48)	53.36 (25.77)	43.37 (27.62)
	H	16.65 (15.43)	32.70 (26.38)	29.15 (22.17)	26.17 (22.52)

Table 6 displays the means and standard deviations of the NASA-TLX items according to visualization group. There were no significant main effects for visualization group and no significant interaction effects (all F-statistics < 1.4).

Table 6. Mean (standard deviation) responses to the NASA-TLX Workload items according to visualization group (L=low; H=high).

		Remember	Analyze	Create	Total
MenD	L	3.73 (4.45)	7.86 (5.70)	9.73 (6.42)	7.09 (6.06)
	H	4.90 (4.75)	9.05 (5.12)	9.05 (5.25)	7.67 (5.35)
PhyD	L	2.00 (1.38)	4.52 (4.08)	5.55 (5.52)	4.02 (4.26)
	H	3.55 (4.01)	5.30 (5.98)	4.90 (5.23)	4.58 (5.11)
TemD	L	2.86 (3.20)	6.14 (4.21)	6.18 (5.23)	5.05 (4.51)
	H	4.45 (4.81)	8.60 (6.00)	7.20 (5.64)	6.75 (5.68)
Perf	L	3.86 (4.60)	6.90 (5.49)	6.55 (5.70)	5.75 (5.37)
	H	3.70 (4.70)	5.90 (5.41)	6.20 (5.31)	5.27 (5.18)
Effort	L	3.23 (3.37)	7.19 (5.71)	7.73 (6.46)	6.03 (5.63)
	H	4.25 (4.06)	8.70 (5.78)	7.40 (5.14)	6.78 (5.31)
Frus	L	2.86 (3.66)	7.38 (4.92)	7.09 (5.89)	5.75 (5.26)
	H	4.20 (5.85)	6.60 (5.82)	6.00 (5.34)	3.50 (4.81)
All	L	18.55 (13.41)	40.00 (25.09)	42.82 (29.43)	33.69 (25.70)
	H	25.05 (25.36)	44.15 (30.68)	40.75 (24.26)	36.65 (27.75)

4.3 Effects on Search Behaviors

In this section we examine differences in search behaviors between participants with different cognitive abilities and for tasks of different complexity. As in the previous section, we conduct three mixed model ANOVAs using each of the cognitive abilities as independent, between-subject variables and evaluate main effects for both cognitive ability and task complexity, as well as interaction effects. We first present data examining potential main effects for task complexity.

4.3.1 Task Complexity Effects

Table 7 displays the mean and standard deviations of search behaviors according to task complexity. Significant main effects for task complexity were found for session length [$F(2, 21)=29.05$, $p<0.01$, $\eta^2=0.61$]; number of queries issued [$F(2, 21)=4.02$, $p<0.05$, $\eta^2=0.18$]; query length [$F(2, 21)=9.88$, $p<0.01$, $\eta^2=0.34$]; and number of SERP clicks [$F(2, 21)=3.87$, $p<0.05$, $\eta^2=0.17$].

Table 7. Mean (standard deviation) for search behaviors according to task complexity. Significant differences are indicated with **$p<0.01$ and *$p<0.05$. Significant pair-wise comparisons ($p<0.05$) are shown in last column.

	Remem	Analyze	Create	Pair-wise
session_length (minutes)**	3.81 (2.52)	9.11 (3.27)	7.03 (3.22)	R< A, C C < A
num_queries*	1.98 (2.02)	3.43 (3.51)	2.79 (2.23)	R<A, C
num_abandons	0.33 (0.87)	1.05 (2.21)	0.76 (1.32)	-
query_length**	5.71 (2.56)	4.35 (1.69)	5.54 (2.05)	A < R, C
#serp_clicks*	2.00 (2.06)	3.33 (2.94)	2.69 (1.89)	R < A, C
#urls_viewed	2.67 (3.18)	4.21 (3.66)	3.48 (2.48)	-
#urls/query	1.41 (0.88)	1.69 (1.62)	1.53 (1.29)	-
serp_dwell time (seconds)	26.41 (30.35)	24.44 (36.53)	18.73 (15.03)	-

4.3.2 Cognitive Ability Effects

Table 8 displays search behaviors according to associative memory group for each task type. The ANOVA showed no significant main effects for associative memory group and no significant interaction effects.

Table 9 displays the means and standard deviations for the search behavior measures according to perceptual speed group. Participants in the low perceptual speed group were found to take significantly longer to complete the search tasks [$F(1, 21)=29.75$, $p<0.01$, $\eta^2=0.61$]. Those in the high perceptual speed group issued significantly longer queries [$F(1, 21)=5.74$, $p<0.05$, $\eta^2=0.23$], made more SERP clicks [$F(1, 21)=12.29$, $p<0.01$, $\eta^2=0.39$], viewed more URLs [$F(1, 21)=11.87$, $p<0.01$, $\eta^2=0.39$] and viewed more URLs per query [$F(1, 21)=7.58$, $p<0.05$, $\eta^2=0.29$]. There were no significant interaction effects. In general, it seems that people with high perceptual speed engaged in more search interaction, while those with low perceptual speed spent more time examining SERPs.

Table 10 displays the means and standard deviations for the search behavior measures according to visualization group. Participants in the high visualization group issued significantly more queries [$F(1, 21)=9.59$, $p<0.01$, $\eta^2=0.34$], abandoned more queries [$F(1, 21)=6.46$, $p<0.05$, $\eta^2=0.25$], had more SERP clicks [$F(1, 21)=6.86$, $p<0.05$, $\eta^2=0.27$] and viewed more URLs [$F(1, 21)=7.58$, $p<0.05$, $\eta^2=0.23$]. There were no significant interaction effects. The general trends were similar to those related to perceptual speed: people in the high group seemed to interact more, while those in the low group spent more time examining SERPs.

Table 8. Mean (standard deviation) of search behaviors according to associative memory group (L=low; H=high). Session length is reported in minutes, dwell time in seconds.

		Remember	Analyze	Create	Total
Session length	L	3.91 (1.65)	9.24 (2.83)	6.60 (2.68)	6.58 (3.83)
	H	3.71 (2.10)	8.96 (2.15)	7.50 (2.79)	6.72 (3.61)
Queries	L	1.59 (0.77)	3.23 (2.36)	2.14 (1.43)	2.32 (2.23)
	H	2.40 (2.01)	3.65 (3.07)	3.50 (2.20)	3.18 (3.12)
Abandon ments	L	0.23 (0.41)	0.68 (0.78)	0.50 (0.74)	0.47 (0.98)
	H	0.45 (0.86)	1.45 (2.23)	1.05 (1.19)	0.98 (2.03)
Query length	L	5.93 (1.59)	4.36 (1.34)	5.57 (2.01)	5.30 (2.27)
	H	5.48 (1.69)	4.34 (1.48)	5.51 (1.39)	5.11 (2.13)
SERP clicks	L	1.68 (0.75)	3.36 (2.67)	2.14 (1.53)	2.39 (2.36)
	H	2.35 (1.99)	3.30 (1.97)	3.30 (1.36)	2.98 (2.40)
URLs viewed	L	2.00 (0.87)	4.00 (3.49)	2.68 (1.63)	2.89 (3.01)
	H	3.40 (3.48)	4.45 (2.43)	4.35 (1.96)	4.07 (3.27)
URLs per query	L	1.31 (0.58)	1.22 (0.50)	1.55 (1.22)	1.37 (1.09)
	H	1.51 (0.45)	2.16 (1.29)	1.51 (0.66)	1.73 (1.46)
SERP dwellT	L	31.64 (21.15)	28.86 (38.24)	21.41 (12.50)	28.28 (36.14)
	H	19.16 (17.83)	18.18 (9.72)	15.55 (7.10)	17.65 (16.07)

Table 9. Mean (standard deviation) of search behaviors according to perceptual speed group (L=low; H=high). Shaded values are significantly different *p*<0.01.

		Remember	Analyze	Create	Total
Session length	L	4.15 (1.82)	8.86 (2.30)	7.55 (3.03)	6.85 (3.71)
	H	3.45 (1.88)	9.38 (2.75)	6.46 (2.31)	6.43 (3.73)
Queries	L	1.64 (1.14)	2.96 (3.06)	2.50 (1.92)	2.36 (2.69)
	H	2.35 (1.83)	3.95 (2.18)	3.10 (1.97)	3.13 (2.70)
Abandon ments	L	0.14 (0.32)	1.05 (2.20)	0.59 (0.80)	0.59 (1.79)
	H	0.55 (0.86)	1.05 (0.80)	0.95 (1.19)	0.85 (1.33)
Query length	L	5.00 (1.76)	3.83 (1.55)	5.01 (2.05)	4.62 (2.24)
	H	6.50 (1.02)	4.92 (0.91)	6.13 (1.01)	5.85 (1.98)
SERP clicks	L	1.55 (1.04)	2.05 (1.25)	2.14 (1.48)	1.91 (1.78)
	H	2.50 (1.76)	4.75 (2.42)	3.30 (1.42)	3.52 (2.68)
URLs viewed	L	2.05 (1.40)	2.32 (0.98)	2.77 (1.66)	2.38 (1.96)
	H	3.35 (3.30)	6.30 (3.05)	4.25 (2.02)	4.63 (3.81)
URLs per query	L	1.21 (0.49)	1.22 (0.62)	1.34 (0.70)	1.26 (0.88)
	H	1.62 (0.48)	2.16 (1.24)	1.74 (1.20)	1.84 (1.58)
SERP dwellT	L	29.71 (21.90)	29.74 (38.16)	20.54 (12.63)	27.95 (36.65)
	H	21.29 (18.21)	17.22 (8.70)	16.50 (7.57)	18.34 (16.36)

Table 10. Mean (standard deviation) of search behaviors according to visualization group (L=low; H=high). Shaded values are significantly different *p*<0.01.

		Remember	Analyze	Create	Total
Session length	L	3.47 (1.59)	9.09 (2.43)	7.23 (2.97)	6.60 (3.77)
	H	4.19 (2.08)	9.12 (2.65)	6.81 (2.52)	6.71 (3.68)
Queries	L	1.27 (0.41)	2.14 (1.19)	2.18 (0.90)	1.86 (1.35)
	H	2.75 (1.90)	4.85 (3.15)	3.45 (2.52)	3.68 (3.44)
Abandon ments	L	0.05 (0.15)	0.50 (0.59)	0.41 (0.49)	0.32 (0.71)
	H	0.65 (0.85)	1.65 (2.20)	1.15 (1.27)	1.15 (2.10)
Query length	L	5.49 (1.69)	4.40 (1.57)	5.00 (1.25)	4.98 (2.00)
	H	5.96 (1.58)	4.29 (1.21)	6.14 (1.98)	5.46 (2.38)
SERP clicks	L	1.32 (0.78)	2.73 (2.27)	2.09 (1.02)	2.05 (1.89)
	H	2.75 (1.72)	4.00 (2.27)	3.35 (1.78)	3.37 (2.69)
URLs viewed	L	1.82 (1.03)	3.41 (2.54)	2.64 (1.23)	2.62 (2.23)
	H	3.60 (3.32)	5.10 (3.28)	4.40 (2.22)	4.37 (3.79)
URLs per query	L	1.30 (0.49)	1.67 (0.75)	1.32 (0.72)	1.44 (0.99)
	H	1.52 (0.54)	1.66 (1.36)	1.75 (1.18)	1.65 (1.55)
SERP dwellT	L	25.95 (17.16)	22.89 (30.17)	21.48 (11.85)	24.34 (25.33)
	H	25.42 (24.05)	24.75 (27.79)	15.46 (8.17)	21.88 (31.64)

5. DISCUSSION

Several trends in our results are worth noting. Task complexity had a significant effect on search behavior and mental workload. In terms of search behavior, more complex tasks were associated with significantly longer search sessions, more queries, longer queries, and more SERP clicks. This is largely consistent with previous work [4, 31, 53]. In terms of mental workload, more complex tasks were associated with significantly lower levels of performance and greater levels of mental demand, physical demand, temporal demand, frustration, and effort. Prior work found that more complex tasks were associated with greater levels of post-task *difficulty*, defined as the user's subjective assessment about the amount of effort expended during the search [4, 53]. Our results complement this prior work by showing that task complexity also affects the different dimensions of workload considered in the NASA-TLX. It is interesting to note that the greatest effect was for the mental demand dimension, as we used tasks with varying levels of *cognitive* complexity. Future work might consider whether different characterizations of task complexity have a greater influence on different workload dimensions.

We considered the effects of three cognitive abilities (i.e., perceptual speed, visualization ability, and associative memory) on search behaviors and workload. Perceptual speed had the strongest effect on both search behavior and workload. Participants in the high perceptual speed group exhibited more search activity (more queries, longer queries, more clicks, more page visits, and more page visits per query) *and* also experienced lower levels of workload across *all* dimensions. The effect size was the greatest for physical demand and frustration. In terms of physical demand, participants in the low perceptual speed group experienced more than twice the physical demand while completing our *analyze* and *create* tasks (though this experience was still only rated near the midpoint of the scale). In terms of frustration, participants in the low perceptual speed group experienced more than twice the level of frustration while completing our *create* tasks (with the experience again rated only near the midpoint of the scale). Although we used different tasks and conducted our study using the open web, our results related to perceptual speed reflect similar findings from prior research [1, 2, 35]. Users with stronger perceptual speed abilities exhibit advantages in information searching. Our research has also found that the performance difference between the high and low groups manifests in significantly increased workload across all dimensions for the users with low abilities, especially on more complex tasks. This new finding from our work warrants further investigation, possibly toward the goal of developing user-adaptive systems whose features capitalize on the cognitive strengths of users without penalizing their weaknesses. For instance, people with low perceptual speed might benefit from additional tools to help them navigate documents, and keep track of, and integrate, their findings. Such people might also benefit from alternative layouts of search results, as well as documents.

Visualization ability had a significant effect on a few of our search behavior measures, but no significant effect on any of the workload dimensions. Participants in the high visualization ability group exhibited more search activity (longer queries, more clicks, more page visits, and more page visits per query). However, they experienced comparable levels of workload. It is possible that the small sample size of our study may have reduced the likelihood for finding a significant difference in perceived workload between the two groups.

Associative memory did not have a significant effect on search behavior or workload. Based on Table 7, our tasks did not require a very long time to complete. Even our *analyze* tasks (arguably the most demanding) took only an average of 9.11 minutes to complete. One possible explanation for why associative memory did not have a significant effect is that our tasks and our study design did not place a heavy demand on associative memory.

While this research makes important contributions to IIR research, it also has some limitations we wish to acknowledge. This study was conducted as a laboratory experiment and so it is not certain to what extent it captures the naturalistic behaviors of users. In addition, the size of our sample was fairly small, which potentially increases the likelihood of Type II errors in our analysis, although we did observe many large effect sizes. The small sample also limits our ability to generalize. We did not include analysis of participants' answers to the task questions in this paper. We plan to analyze the responses in future work.

6. CONCLUSION

The combination of the user's cognitive abilities and perception of mental workload create a powerful set of characteristics for consideration in IIR research. In this study, we explored the ways in which users' cognitive abilities affect their search behaviors and perceptions of workload while conducting search tasks of varying complexity.

Our work makes the following contributions to the IIR research community. First, it provides findings about a population not extensively studied in IIR abilities research: the general adult population. Second, it identifies specific cognitive abilities that impact users' search behaviors, some of our findings were consistent with past research and some were new. Third, it establishes levels of mental demand for specific tasks of specific complexity levels. Fourth, it continues the use of a set of previously developed information search tasks of varying levels of complexity. Finally, it refines the use of additional measures of ability and demand for use in IIR research.

We tested the effects of three cognitive abilities on search behavior and workload: associative memory, perceptual speed, and visualization ability. Perceptual ability had the greatest effect on search behavior and workload, and we believe this presents an opportunity for future research. It may be possible to detect perceptual ability from a user's prolonged interaction with a search system and to provide customized interactions that alleviate the level of workload for this user population. Future work might also consider whether a different set of tasks or a different experimental design reveal greater differences in search behavior and workload for users with varying levels of associated memory and visualization ability. Finally, future research might also investigate other types of cognitive abilities.

7. ACKNOWLEDGMENTS

We thank Joel White and the staff at Durham County Public Library for allowing us to use their facility to recruit and study participants. We also thank UNC-SILS for financial support.

8. REFERENCES

[1] Al-Maskari, A. and Sanderson, M., 2011. The effect of user characteristics on search effectiveness in information retrieval. *IP&M*, *47*, 5, 719-729.

[2] Allen, B., 1994. Perceptual speed, learning and information retrieval performance. In *Proc. SIGIR*, 71-80.

[3] Anderson, L.W., Krathwohl, D.R., and Bloom, B.S., 2001. *A taxonomy for learning, teaching, and assessing*. Allyn & Bacon, New York.

[4] Arguello, J., 2014. Predicting search task difficulty. In *Proc. ECIR*, 88-99.

[5] Arguello, J., Wu, W.-C., Kelly, D., and Edwards, A., 2012. Task complexity, vertical display and user interaction in aggregated search. In *Proc. SIGIR*, 435-444.

[6] Baddeley, A.D., 2003. Working memory: Looking back and looking forward. *Neuroscience*, *4*, 829-839.

[7] Bates, M.J., 1993. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, *13*, 5, 407-424.

[8] Belkin, N.J., 1980. Anomalous states of knowledge as a basis for information retrieval. *Can J Inf Sci*, *5*, 133-134.

[9] Bell, D.J. and Ruthven, I., 2004. Searcher assessments of task complexity for Web searching. In *Advances in Information Retrieval*, S. McDonald and J. Tait Eds. Springer, 57-71.

[10] Borgman, C.L., 1989. All users of information retrieval systems are not created equal: An exploration into individual differences. *IP&M*, *25*, 3, 237-251.

[11] Borgman, C.L., 1996. Why are online catalogs still hard to use? *JASIS*, *47*, 7, 493-503.

[12] Bystrom, K. and Jarvelin, K., 1995. Task complexity affects information seeking and use. *IP&M*, *31*, 191-213.

[13] Campagnoni, F.R. and Erlich, K., 1989. Information retrieval using a hypertext-based help system. In *Proc. SIGIR*, 212-220.

[14] Campbell, D.J., 1988. Task Complexity: A Review and Analysis. *Acad Manage Rev*, *13*, 1, 40-52.

[15] Carroll, J.B., 1993. *Human cognitive abilties: A survey of factor-analytic studies*. Cambridge University Press, NY.

[16] Debowski, S., Wood, R.E., and Bandura, A., 2001. Impact of guided exploration & enactive exploration on self-

regulatory mechanisms & information acquisition through electronic search. *J Appl Psychol*, *86*, 6, 1129-1141.
[17] Di Stasi, L.L., Antolí, A., Gea, M., and Cañas, J.J., 2011. A neuroergonomic approach to evaluating mental workload in hypermedia interactions. *Int J Ind Ergonom*, *41*, 3, 298-304.
[18] Downing, R.E., Moore, J.L., and Brown, S.W., 2005. The effects and interaction of spatial visualization and domain expertise on information seeking. *Comput Hum Behav*, *21*, 2, 195-209.
[19] Ekstrom, R.B., 1973. *Cognitive factors: Some recent literature.* Office of Naval Research, Report No. PR-73-30; TR-2.
[20] Ekstrom, R.B., French, J.W., Harman, H.H., and Dermen, D., 1976. *Kit of factor-referenced cognitive tests*. Educational Testing Service, Princeton, NJ.
[21] Ekstrom, R.B., French, J.W., Harman, H.H., and Dermen, D., 1976. *Manual for kit of factor-referenced cognitive tests.* Educational Testing Service.
[22] Ford, N., Miller, D., and Moss, N., 2001. The Role of Individual Differences in Internet Searching: An Empirical Study. *JASIST*, *52*, 12, 1049-1066.
[23] Gao, Q., 2011. Empirical study of tagging for personal information organization: Performance, workload, memory, consistency. *Int J HCI*, *27*, 9, 821-863.
[24] Gwizdka, J., 2009. Assessing cognitive load on web search tasks. *Ergonom Open J*, *2*, 114-123.
[25] Gwizdka, J., 2009. What a difference a tag cloud makes: effects of tasks and cognitive abilities on search results interface use. *Info Res*, *14*, 4, paper 14.
[26] Gwizdka, J., 2010. Distribution of cognitive load in Web search. *JASIST*, *61*, 11, 2167-2187.
[27] Haapalainen, E., Kim, S.J., Forlizzi, J.F., and Dey, A.K., 2010. Psycho-physiological measures for assessing cognitive load. In *Proc. UbiComp*, ACM, 301-310.
[28] Hart, S.G., 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum Fac Ergonom Soc*, *50*, 9, 904-908.
[29] Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, P.A. Hancock and N. Meshkati Eds. NH Press, Amsterdam.
[30] Ingwersen, P., 1996. Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *J Doc*, *52*, 11, 3-50.
[31] Jansen, B.J., Booth, D., and Smith, B., 2009. Using the taxonomy of cognitive learning to model online searching. *IP&M*, *45*, 6, 643-663.
[32] Jex, H., 1988. Measuring mental workload: Problems, progress, promises. In *Human Mental Workload*, P. Hancock and N. Meshkati Eds. Elsevier, 5-39.
[33] Kammerer, Y., Nairn, R., Pirolli, P., and Chi, E.H., 2009. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Human Compu*, ACM, Boston, MA, 625-634.
[34] Kelton, A.S. and Pennington, R.R., 2012. Internet financial reporting: The effects of information presentation format and content differences on investor decision making. *Comput Hum Behav*, *28*, 4, 1178-1185.
[35] Kim, K. and Allen, B., 2002. Cognitive & task influences on web searching behavior. *JASIST*, *53*, 2, 109-119.
[36] MacFarlane, A., Al-Wabil, A., Marshall, C., Albrair, A., Jones, S.A., and Zaphiris, P., 2010. The effect of dyslexia on information retrieval: A pilot study. *J Doc*, *66*, 3, 307-326.
[37] MacFarlane, A., Albrair, A., Marshall, C., and Buchanan, G., 2012. Phonological working memory impacts on information searching: An investigation of dyslexia. In *Proc. IIiX 2012*, 27-34.
[38] Marchionini, G., 1995. *Information seeking in electronic environments*. Cambridge University Press, New York.
[39] Megaw, T., 2005. The definition and measurement of mental workload. In *Evaluation of human work*, E.N. Corlett and J.R. Wilson Eds. Taylor & Francis, Boca Raton, FL, 525-551.
[40] Niu, X. and Kelly, D., 2014. The use of query suggestions during information search. *IP&M*, *50*, 1, 218-234.
[41] Pak, R., Rogers, W.A., and Fisk, A.D., 2006. Spatial Ability Subfactors and Their Influences on a Computer-Based Information Search Task. *Proc. Hum Fac Ergonom Soc*, *48*, 1, 154-165.
[42] Palmquist, R.A. and Kyung-Sun, K., 2000. Cognitive style & online database search experience as predictors of web search performance. *ASIS*, *51*, 6, 558-566.
[43] Pirolli, P. and Card, S.K., 1999. Information Foraging. *Psycho Rev*, *106*, 4.
[44] Santos, E., Nguyen, H., Zhao, Q., and Pukinskis, E., 2003. Empirical Evaluation of Adaptive User Modeling in a Medical Information Retrieval Application. In *User Modeling 2003*, P. Brusilovsky, A. Corbett and F. De Rosis Eds. Springer Berlin / Heidelberg, 148-148.
[45] Schmutz, P., Heinz, S., Metrailler, Y., and Opwis, K., 2009. Cognitive load in eCommerce applications: Measurement and effects on user satisfaction. *Adv HCI*.
[46] Speier, C. and Morris, M.G., 2003. The influence of query interface design on decision-making performance. *MIS Quart*, *27*, 3, 397-423.
[47] Swan, R.C. and Allan, J., 1998. Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. In *Proc. SIGIR*, ACM, 173-181.
[48] Toms, E.G., O' Brien, H.L., Mackenzie, T., Jordan, C., Freund, L., Toze, S., Dawe, E., and Macnutt, A., 2007. Task effects on interactive search: The query factor. In *Proc. INEX 2007*, 359-372.
[49] Wechsler, D., 1997. *Wechsler Adult Intelligence Test*. The Psychological Corporation, San Antonio, TX.
[50] Westerman, S.J., Davies, D.R., Glendon, A.I., Stammers, R.B., and Matthews, G., 1995. Age and cognitive ability as predictors of computerized information retrieval. *Behav Inform Technol*, *14*, 5, 313-326.
[51] Wood, R.E., 1986. Task complexity: Definition of the construct. *Organ Behav Hum Dec*, *37*, 1, 60-82.
[52] Wothke, W., Curran, L.T., Augustin, J.W., Guerrero, C., Jr., Bock, R.D., Fairbank, B.A., and Gillet, A.H., 1991. Factor Analytic Examination of the ASVAB and Kit of Factor-Referenced Tests, A.F. Systems Ed., Brooks Air Force Base, Texas, 76.
[53] Wu, W.-C., Kelly, D., Edwards, A., and Arguello, J., 2012. Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity. In *Proc. IIiX 2012*.
[54] Xie, B. and Salvendy, G., 2000. Prediction of mental workload in single and multiple task environments. *Int J Cog Ergonom*, *4*, 3, 213-242.