LOAN ME THE MONEY: HOW COGNITIVE ABILITIES AND FINANCIAL KNOWLEDGE INFLUENCE CONSUMERS' INFORMATION BEHAVIORS

Kathleen N. Brennan

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science.

Chapel Hill
2018

Approved by:

Diane Kelly

Jaime Arguello

Robert Capra III

Jacek Gwizdka

Javed Mostafa

ABSTRACT

Kathleen N. Brennan: Loan Me the Money: How Cognitive Abilities and Financial Knowledge Influence Consumers' Information Behaviors
(Under the direction of Diane Kelly)

For most people, financial well-being depends on the ability to make sound decisions about many aspects of personal finance. This is especially true in the United States (U.S.), when it comes to consumer loan products such as mortgages and student loans. Consumers who lack strong financial knowledge can unwittingly expose themselves to bad information when searching online. Without understanding people's searching behaviors, information professionals cannot know whether personal finance-related information systems adequately meet the needs of the people using them. Interactive Information Retrieval (IIR) is well-suited to study this, yet there has been little research in this area. One approach that makes sense for studying debt-related information searching is to investigate the role individual differences play in people's searching. This is because individual differences are testable constructs that can be associated with differences in search performance outcomes.

The purpose of this dissertation research is to understand influences that cognitive abilities and financial knowledge have on outcomes related to search, assessment, and mental workload of adults searching online for debt-related personal finance information. A theoretical model is proposed in which financial knowledge acts as a moderating variable on the effect that cognitive abilities have on search and evaluation behaviors as well as mental workload.

The results of the study were mixed. The testing of hypotheses on the model were unsuccessful and provide information for informing future model designs and hypothesis development. The qualitative portion of the study provided numerous insights, including that the topic of personal finance, specifically in the realm of financial loans such as mortgages, student loans, and payday loans, is more challenging for people than they realize. Participants reported low prior knowledge of all task topics and used simple search strategies such as avoiding advertisements on search engine results pages (SERPs), relying heavily on the first SERP result, and reformulating queries rather than investigating SERPs at deeper levels. Participants rated most webpages they found as relevant or very relevant but expert assessors rated most of those same pages as only somewhat relevant or not relevant. The findings have numerous implications and point to key areas for further research.

To my beautiful wife, Alaina.

Thank you for the support, encouragement, and inspiration you give me every day.

ACKNOWLEDGEMENTS

I am very grateful to many people who helped me earn my doctorate at UNC's School of Information and Library Science (UNC-SILS).

Thank you to my advisor, Dr. Diane Kelly. Your mentorship started on my first day at UNC-SILS and now, years later, I find it almost impossible to sum up its great value in just a few words. On the days I feel most competent as a researcher, I know it is because of your training I can feel this way. Thank you for helping me become a scholar. Thank you for your guidance, support, loyalty, and patience. Thank you for your friendship.

Thank you to the members of my dissertation committee members, Drs. Jaime Arguello, Rob Capra, Jacek Gwizdka, and Javed Mostafa. You have each played a special and unique role in guiding me through the dissertation process. As I go forward, I bring your insights with me into new directions and approaches for this research.

It is not surprising to me that UNC-SILS consistently receives the #1 spot in national and worldwide ILS program rankings because I know how exceptional many of its faculty and staff really are. I owe special thanks to several of these individuals in particular. My deepest thanks go to my good friend and teaching mentor, Dr. Evelyn Daniel, for supporting me through so much of the daily doctoral student grind and for sharing your own stories with me as we sat many nights in Memorial Hall awaiting a CPA concert event opening. Thank you also to Dr. Barbara Wildemuth for your consistency, fairness, and strength of intellect that sometimes gave me pause and always made me think more deeply about whatever topic was at hand. Thank you to Dr. Gary Marchionini for inspiring me to learn how to answer the "so what?" question about my research. Thank you to Dr. Ron Bergquist for your helpful teaching guidance and for trusting me to teach in your place. Thank you to Dr. Amelia Gibson for the guidance you gave me about qualitative coding and for being an example of authenticity and strength. Thank you to all the administrative staff at UNC-SILS but especially to Lara Bailey – somehow you manage to help all the doctoral students with every aspect of university minutiae without ever losing your cheerful demeanor. That is a true gift that you have.

I was fortunate to be able to attend numerous research conferences during my doctoral program, thanks to student travel scholarships from ACM-SIGIR, ASIS&T, and others. During these events I was often humbled by the amount of time, attention, and interest that international scholars graciously gave me to talk about my research, even in its most nascent stages. My thanks go to these researchers for that: Drs. Leif Azzopardi, Pia Borlund, Katriina Byström, Pertti Vakkari, Heather O'Brien, Luanne Freund, George Buchanan, Cathy Smith, and Howard White.

I was also fortunate to have many excellent current and former peers from the doctoral program at UNC-SILS who shared this journey with me, providing valuable input and feedback throughout the milestones in my PhD program. These peers included Dr. Ashlee Edwards-Brinegar, Anita Crescenzi, Leslie Thomson, Dr. Earl Bailey, Dr. Wan-ching Wu, Dr. Laura Sheble, Sandeep Avula, Shenmeng Xu, Yinglong Zhang, Yuan Li, Bogeum Choi, and Austin Ward. I hope I have not left anyone out. Thank you, to all of you.

There are numerous funders to thank for their financial support. These include UNC-SILS, the Kalp Fellowship, the Lester Asheim Fellowship, the Louis Round Wilson Fellowship, and the Monning Family Fund. I thank Phi Beta Mu for the Eugene Garfield Dissertation scholarship during the last year of my program. I thank my parents, Edward and Sheila Brennan, for financial help during some lean semesters. I also thank my wife, Alaina Kupec, who supported me financially through the last few years of my program.

I thank my two expert assessors, Melissa Eggleston and Marianne Sullivan, for their extensive efforts to evaluate more than 600 webpages each for the relevance part of my study. I know it was a bigger task than you expected and I greatly appreciate your work. I also thank my 47 anonymous participants.

I am grateful to numerous friends who gave me so much moral support, including Marne, Melissa, Amber, Kendra, Tom, and many others. I am also indebted to many of my family members. I thank my parents for teaching me to value education through their inspiring examples and for instilling in me the values of persistence and patience. I thank all the rest of the Brennans and Kupecs who put up with me working on vacations, working while they were visiting, and generally not always being fully present. Your patience and support throughout my doctoral program was nothing short of amazing. Finally, I thank my wife, Alaina Kupec, because most everything good that occurs in my life today is in some way because of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1. Background

For most people, financial well-being depends on the ability to make consistently sound, informed decisions about many aspects of personal finance. This is especially true in the area of consumer loan products such as mortgages and student loans in the United States (U.S.), where consumer debt holdings are currently more than $10.4 Trillion[1]. Even though there is an abundance of high quality information about personal finance available online, the growing array of financial concepts and debt products has become so complex that it is increasingly challenging for people to sort through the broad expanse of choices to make the best-informed financial decisions for themselves and their families. The growing alternative financial services (AFS) industry (e.g., payday lending, rent-to-own leasing, and subprime mortgage lending) makes the situation worse, with its often aggressive marketing practices that can crowd-out safer competing products (McCoy, 2009). The potential for making suboptimal financial choices in this environment is high.

Debt needs to be properly managed. This is backed up by evidence from research indicating that too much debt can be a source of financial stress (Kim, Sorhaindo, & Garman, 2006) and can negatively impact physical health (Drentea & Lavrakas, 2000). Not all debt is the same, however, with research showing that debt related to credits cards causes people greater psychological stress than debt related to owning a home, such as mortgage debt (Brown, Taylor,

[1] Federal Reserve Bank of New York, 2017 Q1 Results

& Wheatley-Price, 2005). At the heart of good personal debt management is the need to understand core aspects of debt and debt products. This is a kind of financial literacy that includes grasping financial concepts such as compound interest rates and inflation risk and being able to discern the advantages and disadvantages of specific kinds of debt products based on one's own particular financial needs and circumstances. Access to accurate, up to date, contextually appropriate information is paramount for individuals to maintain adequate levels of financial literacy for success when it comes to managing personal debt.

1.2. Statement of the Problem

While there are many ways that people can acquire information about debt products and debt management, the main way people go about doing this is by searching the Internet (Bricker et al., 2014). People seek out information sources online before they talk to family members, associates, accountants and financial advisors (Bricker et al., 2014). This is problematic because the Internet is an unregulated, open environment where information can be posted by anyone and where financial transactions are instantaneous. The problem is that consumers who lack strong financial knowledge can unwittingly expose themselves to bad information, whether it simply be advice that does not fit their own personal financial situations or worse, the schemes of bad actors seeking to prey upon them. While the exact size of the consumer population lacking adequate financial knowledge is not known, a recent survey of financial capability in the U.S. (N = 27,564) found that two-thirds of Americans could not pass a basic financial literacy test (Lin, Bumcrot, Ulicny, Lusardi, Mottola, Kieffer, & Walsh, 2016). Financial illiteracy is also more prevalent in some demographic groups than others, such as women and the less-educated. Therefore, this seems like sufficient evidence to argue for learning more about consumers' online searching behaviors in areas like debt and debt management.

The main approach to date for understanding consumers behaviors around debt and debt management has been for consumer finance and behavioral economics researchers to conduct secondary analyses of large public surveys like the Survey of Consumer Finance[2] or the American Housing Survey[3]. These analyses are used for developing basic insights about high-level trends in generic consumer behaviors, but they tell us nothing about the actual behaviors of people who open up their Internet browsers on a daily basis and search for everything from the cost of a pair of shoes to the interest rate they might qualify for on a mortgage loan. This should matter to Information Science and all its subdisciplines because without understanding people's online searching behaviors in this realm, there is no way to know whether personal finance-related information systems adequately meet the needs of the people using them. Interactive Information Retrieval (IIR) is particularly well-suited to study consumers' online debt-related information searching because the foundations of IIR research methods are aimed specifically at evaluating human interactions with information systems. Yet there has been very little research on this in IIR or information and library science (ILS). Occasional studies have appeared in the ILS literature but they have either been related only to personal investing (e.g., (Kuhlthau, 1997; Mezick, 2002)) or, if debt-related, have been about consumers in countries other than the U.S., where regulatory requirements can differ greatly (e.g., (Biza-Khupe, 2014)).

One approach to studying information searching online in IIR has been to look at the role that individual differences play in search behaviors. This makes logical sense as an approach for studying debt-related information searching because individual differences are testable constructs of cognitive, demographic, and interpersonal characteristics that can be associated

[2] https://www.federalreserve.gov/econres/scfindex.htm

[3] https://www.census.gov/programs-surveys/ahs.html

with differences in search performance outcomes. At the same time, investigating individual differences in the context of debt-related information searching opens a new pathway for contributing to the existing body of individual differences research and heeds the recent call for more research of general or specialized adult populations to understand unique considerations of people with different levels of literacy (O'Brien, Dickinson, & Askin, 2017), in this case, financial literacy.

In addition, many studies investigating individual differences look for behavioral signals that can be used to predict outcomes for users, and it is still largely unclear if these signals are valid and reliable. There is no overarching body of research that tells us a coherent story about individual differences (O'Brien et al., 2017). Part of the problem is that actions taken during information searching, such as the number of users' clicks, search queries, and webpage views can be ascribed divergent meanings depending on the user's circumstances (e.g., the user is experiencing difficulty versus the user is actively engaged). Relevance judgments are also tricky because the construct of relevance can be defined from more than one point of view (e.g., system, user, or topic domain) and also because webpages retrieved based on the system's algorithm of relevance may not match what the user judges as relevant, nor what an expert views as relevant (e.g., consider the case of ambiguous queries such as "jaguar," "zenith," or "mont blanc"). Mental workload is another area on which individual differences have an influence, but research to date has only uncovered parts of the picture when it comes to what creates, adds to, and reduces mental workload for users as well as whether or not there may be a need to distinguish between "good" workload (i.e., user engagement) versus "bad" workload (i.e., user fatigue or boredom).

Even in cases where researchers seem to consistently determine a particular individual difference is responsible for a set of actions in study after study, the value of that knowledge outside the search lab has yet to be determined. For example, in the case of individual difference measures for cognitive speed, there is strong evidence that users with high levels of cognitive speediness interact more with search systems (c.f., (Al-Maskari & Sanderson, 2011; Brennan, Kelly, & Arguello, 2014)) and exhibit differing eye gaze patterns (Steichen, Carenini, & Conati, 2013), but there is little evidence of the real-world impact these differences might make in terms of outcomes like relevance judgments, decision-making accuracy, or time savings.

1.3. Purpose of the Study and Research Questions

The purpose of this dissertation is to introduce the study of debt-related information searching to IIR by investigating the ways in which individual differences are manifested in behavioral signals during search and assessment tasks. Similar to early efforts by researchers to increase knowledge about health information seeking to information research, the larger goal of this work is to introduce personal finance information seeking to IIR by offering an empirical approach for studying its dimensions within established research frameworks. The specific aspect of personal finance for this study is consumers' information searching about debt-related information. The context for understanding this kind of searching is through the investigation of the role that individual differences play, specifically cognitive abilities and financial knowledge. Through this approach, it will hopefully be possible to distinguish ability-related influences from domain knowledge-related influences on search and assessment behaviors.

In summary, the objective of this dissertation research is to investigate the influence of cognitive abilities and financial knowledge on adults searching the Internet for information about different kinds of financial loans. It focuses on two main areas of inquiry, which can be expressed as the research questions for the study:

1. How do cognitive abilities and financial knowledge influence the search performance, relevance assessments, and mental workload of adults searching the Internet for information about different kinds of financial loans?
2. What are users' strategies for finding and evaluating information on the Internet about different kinds of financial loans? How do users' cognitive abilities and financial knowledge influence these strategies?

The approach of the two questions is to use both quantitative and qualitative research methods. Since this topic and combination of variables has not been attempted prior to this dissertation research, data collection and analyses are broadly focused so as to understand general areas of debt-related information searching behavior that will be meaningful for future study. Research goals may reflect exploratory, descriptive, and explanatory methods to gathering and analyzing data (Kelly, 2009) and the research for this dissertation addresses all three. The exploratory aspects are related to the qualitative analysis of the stimulated recall sessions as well as some search interactions captured in the eye tracking logs that have to do with time, search engine results pages (SERPs), and eye fixation count. The qualitative coding and analysis is also in part descriptive because it documents and describes the phenomenon of users searching online for personal finance tasks. The numerous hypotheses related to research question #1 are explanatory.

By studying how people search for and evaluate information about consumer debt, this research can generate insights about user behaviors that can be used to create clearer, easier to understand online financial information worlds for consumers. In addition, the work of the dissertation seeks to introduce to information and library science a set of models and specific to the domain of personal finance.

1.4. Organization of the Dissertation

The dissertation is organized into seven chapters, including this one. Chapter two summarizes the literature related to the current dissertation research study. Chapter three provides the research questions, models, and hypotheses of the study. The fourth chapter provides a detailed explanation of the research methods, including the study design and measurements, study tasks, recruitment strategy, and study procedures. Chapter five explains the data analysis and results of the questionnaire and test data, the search interaction and evaluation data, eye tracking data, and stimulated recall data. The sixth chapter discusses the findings, their implications, and the limitations of the research. Chapter seven concludes with the contributions of this research and directions for future work. Appendices at the back of the dissertation contain copies of the questionnaires, tests, and other items that were used to conduct the research.

CHAPTER 2: LITERATURE REVIEW

This chapter contains a review of literature related to cognitive abilities, domain knowledge and the domain of personal finance, mental workload, and information searching and evaluation. It reviews studies of variables that have been found to influence aspects of users' information searching and evaluation, specifically those related to cognitive abilities, topic domain knowledge, and mental workload.

2.1. Cognitive Abilities

Certain cognitive abilities have been found to play important roles in users' search and evaluation behaviors. In this section, I define the term *cognitive abilities*, explain the particular theory of cognitive abilities used in this dissertation, and provide a review of studies that investigated the impact of two abilities, perceptual speed and memory span, on the search and evaluation behaviors of users searching the Internet.

2.1.1. Definition of cognitive abilities. Cognitive ability is defined in this dissertation as follows: *a cognitive ability is a person's inherent and acquired intellectual capacity to comprehend the requirements of a cognitive task within its context and successfully achieve the task's desired outcome*. This original definition is derived from several sources. The first source focuses on the information processing aspect of cognitive abilities: "the term cognitive ability generally refers to the capacity to mentally process, comprehend, and manipulate information" (Reeve, 2007, p. 77). The second emphasizes the need to include cognitive tasks as part of the definition, because in order to observe the information processing capacities of individuals there has to be some kind of task performed and a way to measure its successful completion. Thus, a cognitive task is one in which proper mental functioning is required for understanding the expected outcomes of the task and for performing the task successfully toward that end (Carroll, 1993). The third addresses the context of the task, that the task should be part of a group of tasks sharing similar attributes (Carroll, 1993). The final sources stress that cognitive abilities "are generally assumed to be fairly stable, to have a biological basis, and to be both learned and innate" (Dik, 2007, p. 2) and that cognitive abilities vary across individuals (Carroll, 1993).

2.1.2. Theory of cognitive abilities. The theoretical basis for the measurement of cognitive abilities in this study is the Three-Stratum Theory (Carroll, 1993). It depicts cognitive abilities in a hierarchical structure using a factor-analytic model of human intelligence belonging to three strata or layers. The technique of factor analysis reduces large correlational matrices of original variables, called first-order factors, to a smaller set of factors, called second-order factors, that may then be reduced again to a third-order. Figure 1 shows the Three-Stratum Theory model.

The model depicts the first-order factors at the bottom of the diagram as Stratum I, labeled "Narrow". It shows the name of each original variable and its abbreviation in parentheses. First-order factors are "distinct, but correlated concepts" (Cattell, 1966, p. 225) divided in this figure into their groupings based on a correlation matrix (not shown). In the factorial rotation the correlated first-order factors load onto the eight factors shown above them on Stratum II. These eight "Broad" second-order factors load onto a single factor above them on Stratum III, the "General" factor of general intelligence. The "*g*" represents the "*g*" general factor in Spearman's (1904) two-factor theory of intelligence. The psychometric test batteries produced by Educational Testing Service (ETS) strongly influenced Carroll's work on Stratum I.

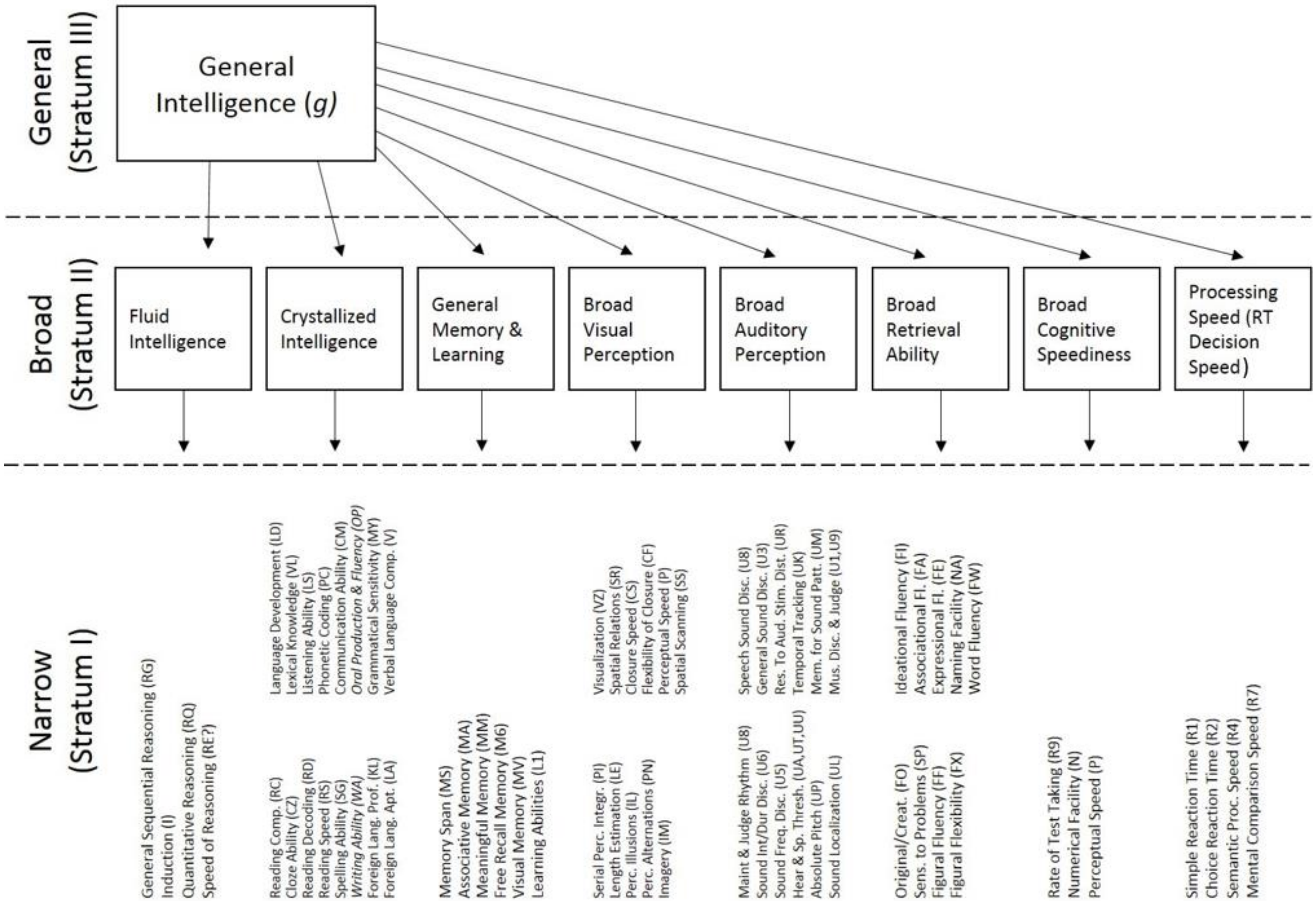


Figure 1. Three-Stratum Theory Model (Carroll, 1993).

Three Stratum Theory is useful for studying individual differences because the approach makes different abilities separable, measurable, and distinguishable. By using an approach based on a factored method, it is possible to select and test only those attributes of ability that are believed most important to specific activities of information search and relevance assessment. For example, while some abilities may be more important in relevance assessment, others may be more important in selecting search results or navigating information systems. This method of understanding abilities can allow for testing of specific kinds of abilities which can then be applied in different contexts of tasks, topic domains, interfaces, and systems in interactive information retrieval (IIR). Based on a review of the literature, the two cognitive abilities studied in this dissertation are perceptual speed and memory span (i.e., as the operationalization of working memory).

2.1.3. Perceptual speed. Perceptual speed has been found to play an important role in search and evaluation behaviors. It is defined as an ability "characterized by the task of finding in a mass of distracting material a given configuration which is borne in mind during the search" (Carroll, 1993, p. 308). In other words, perceptual speed enables us to stay focused on finding the thing we want to find.

Allen (1992, 1993) investigated the influence of cognitive abilities on search performance and effectiveness of a CD-ROM periodical index. Participants were asked to read a stimulus article about the influence of television violence on aggression in young children and then search the periodical index to find articles related to the topic, as if they were going to be writing a college term paper. Three search options allowed participants to search by browsing a subject heading list, entering query terms, or issuing search commands. Regardless of the search option chosen, participants with higher perceptual speed ability achieved higher precision and found higher quality information.

To better understand the role of perceptual speed, Allen conducted further studies. Allen (1994) created a system designed to enhance users' learning of subject heading vocabulary. He found that participants with higher perceptual speed learned more vocabulary and completed higher quality searches than those with lower score perceptual speed. Learning was found to be a mediator variable between perceptual speed and searching performance. Allen also compared systems with different information display formats and found roles of perceptual speed (Allen, 1998a, 1998b, 1998c). One system displayed subject headings either as an alphabetical list or a visual word map. Another system displayed subject headings in a single or multi-window display. Search tasks followed the same procedure as in previous studies, that is, participants were asked to read a brief article and once finished, find as many articles as possible by searching the bibliographic databases subject headings, from which they were to select promising headings, view the article abstracts presented, and then indicate whether or not they would select that article for use by answering a yes/no pop-up screen to print the article. The context of one task was to select articles for writing a ten-page term paper assignment and the context for the second was for writing an article for the student newspaper. Participants with higher perceptual speed scores viewed and printed more records overall than those with lower perceptual speed. They also viewed more references in the linear subject heading display condition (versus the word map condition) and learned more vocabulary in the single window display than they did in the multiwindow display. Participants with lower perceptual speed viewed more references and learned more vocabulary in the word map condition versus the linear subject heading list. While the word map helped all users with lower abilities, it actually hindered vocabulary learning for people with higher perceptual speed.

More recent studies have also investigated perceptual speed within the context of information searching. Al-Maskari and Sanderson (2011) investigated the impact of participants' perceptual speed and search experience on search effectiveness. Search effectiveness was defined using four variables: total number of relevant documents found, time taken to find the first relevant document, self-assessed satisfaction with the search, and self-assessed familiarity with the search topics. Participants searched a Text REtrieval Conference (TREC) collection using TREC topics on an interface called the Query Performance Analyzer, an experimental system which provided access to three retrieval systems (InQuery, Lemur, and Terrier). The researchers found that there was a significant correlation between perceptual speed and the time to find first relevant document variable of user effectiveness. Participants with higher perceptual speed spent less time finding the first relevant document than participants with lower perceptual speed (1.71 minutes versus 2.19 minutes). However, there was no difference in the total number of relevant documents found by participants based on their perceptual speed levels. So while participants with higher perceptual speed were faster at identifying relevant information, this did not contribute to them finding more instances (documents) of relevant information.

Finally, there are several studies that investigated searching on the open Web. Brennan, Kelly, and Arguello (2014) explored how people's cognitive abilities affected their search behaviors and perceptions of mental workload while conducting search tasks on the open Web. Participants were 21 adults from the general public. The three search tasks were taken from the research project that was later published in Kelly, Arguello, Edwards, and Wu (2015), who developed tasks based on Bloom's Taxonomy (Anderson, Krathwohl, & Bloom, 2001). The cognitive complexity levels ranged from least complex *remember* tasks which required participants to find a specific answer, to more complex *analyze* tasks which required participants to generate a list of items for comparing and contrasting, to the most complex *create* tasks that required participants to create new or original solutions from the results of their information searching. While there were no significant interaction effects between cognitive complexity and perceptual speed, there were significant main effects for perceptual speed group and search behaviors. Participants in the high perceptual speed group exhibited more search activity – they submitted more queries, wrote longer queries, made more clicks, visited more web pages, and visited more web pages per query than participants in the low perceptual speed group.

Turpin, Kelly, and Arguello (2016) compared search behaviors of those with high and low perceptual speed who used blended and non-blended search engine results page (SERP) interfaces. The blended interface used two application program interfaces (APIs) to present results from specialized search engines (known as *verticals*) of web pages, news stories, images, videos, and shopping sites. The non-blended interface only allowed participants to access vertical search engines by clicking on separate tabs. Participants (N=16) completed two search tasks using the blended interface and two search tasks with the non-blended interface. Participants with higher perceptual speed rated both interfaces as having higher usability and ease of use than the participants with lower perceptual speed. Participants with lower perceptual speed also took longer on the tasks on the blended interface, while those with higher perceptual speed showed no difference in performance between the two interfaces.

2.1.4. Memory span. Memory is another cognitive ability that impacts search and evaluation (Oh & Kim, 2004; Woodman & Chun, 2006). In this dissertation study, the form of memory studied is working memory, which is operationalized using tests for memory span. This form of memory has been defined as the ability to recall a number of distinct elements for immediate reproduction (Ekstrom, French, Harman, & Dermen, 1976b, p. 101). Several studies have investigated working memory in IIR. MacFarlane, et al., (2012) designed a study to understand the impact of impaired memory on information search behaviors by comparing participants with normal memory abilities to participants with dyslexia, a reading disorder caused in part by impairment in phonological processing and working memory. They found that participants with reduced working memory abilities judged fewer documents non-relevant than participants with higher working memory abilities. Participants conducted searches using the TIPSTER collection from TREC 7 and 8. The researchers identified phonological processing working memory as the ability that enables the person to retain words in memory for the several seconds it takes to process and map the meanings of the words to the printed text. Difficulties created by a deficiency in this ability hinder the person's reading, spelling, and comprehension skills. The researchers tested the eight dyslexic and eight non-dyslexic participants for their reading, comprehension, and spelling abilities; working memory using the Wechsler Adult Intelligence "Digit Span" Scale; as well as for dyslexia using several established assessment instruments. Logged measures included total documents read, judged relevant, judged non-relevant, and examined. They also measured changes of judgments from relevant to non-relevant, level of agreement with TREC relevance judgments, session length, number of searches, hit-lists examined and pool views per iteration.

Both groups judged the same number of documents as relevant, but non-dyslexic users judged an average of 14 more documents non-relevant than did the dyslexic users. In bivariate correlations between the number of documents judged non-relevant and the cognitive measures, the researchers found that number of documents judged non-relevant was positively correlated with scores on the digit span test (i.e., the measure of phonological working memory). Participants with higher digit span scores (i.e., the non-dyslexic participants) judged more documents non-relevant. They suggested that deciding a document is relevant requires less cognitive effort than determining that a document is not relevant. They based this on the notion that in order to determine a document non-relevant, one would have to read and maintain in working memory the contents of the entire document in order to determine if the document matched or did not match the topic, whereas in order to determine that a document is relevant, all one has to do is identify the satisfaction of the first instance of relevance without having to necessarily read the entire document. Thus, the demands on short-term working memory storage would be greater in the case of determining non-relevant documents.

Other studies focused on individual differences in memory ability outside of those with impairments. Gwizdka (2013) explored the effects of task complexity and memory span ability on participants searching a collection of social bookmarks related to travel, sightseeing, and shopping, under two different interface conditions. Tasks required participants to find information about travel and sightseeing topics to recommend to a friend. The simple task required finding a simple fact, such as the location of an airport, whereas the complex tasks required participants to gather information about multiple topics and sites, such as gathering information about three kinds of art collections that could only be found by searching three different museum websites. The search mechanism was a tag set that automatically re-generated with each results set, as the user added (clicked) and deleted tags to form queries for information. The number of tags that were added and deleted was a proxy for number of query refinements. Tag deletions were considered ‘cognitive moves’, or query reformulations, about half the time and ‘physical moves’ or navigation actions to go back to a previous results list the other half. The interfaces were either a textual list of results with just tags and URLs or a list of results that also included a tag cloud overview. Once participants found their answers, they were instructed to write them in a message area on the screen, as if they were making a recommendation to their friend.

Participants were tested for working memory and were divided into high and low groups. The high memory span group interacted more with the search system by performing more actions to find more information, while the low memory span group was the opposite – they looked at fewer websites and engaged less with the system. Thus, cognitive ability of working memory span interacted with task complexity by affecting the search interaction behaviors of the participants. Gwizdka suspected this to be a type of satisficing behavior on the part of the low working memory span participants. Similar to his earlier work in (Gwizdka, 2009), he found that the extra actions of the high ability working memory span users did not result in necessarily immediate better search task outcomes.

2.2. Domain Knowledge and the Domain of Personal Finance

This section covers domain knowledge in general and its role in search and evaluation behaviors. It also covers the specific domain of personal finance and search and evaluation behaviors in the domain of personal finance.

2.2.1. Domain knowledge, search, and evaluation behaviors. In this section, studies are reviewed that provided evidence of effects of domain knowledge on online search and evaluation behaviors. Domain knowledge or expertise is defined differently in different studies. In some cases, especially where a student population forms the study participants, level of education is used as a proxy for expertise, such as graduate students being seen as having greater expertise in a subject area than undergraduates (Zhang, Anghelescu, & Yuan, 2005) or graduate students being seen in general as having high levels of domain knowledge (Wildemuth, 2004). In other cases, domain knowledge is self-identified, such as through rating of familiarity with terms in a subject domain thesaurus (Zhang et al., 2005) or by ranking one's level of expertise on given search topics (Hembrooke, Granka, Gay, & Liddy, 2005).

Domain knowledge has been found to affect querying behaviors. Users with higher domain knowledge have been found to issue longer queries (Hembrooke et al., 2005; Zhang et al., 2005) or issue longer queries when searching their area of expertise versus other topics (Freund & Toms, 2006). There is also evidence that domain-specific vocabulary is more likely used in the queries of users with high domain knowledge (Vakkari, Pennanen, & Serola, 2003; Wildemuth, 2004; Zhang et al., 2005), while low domain knowledge has been associated with lower ability to select the best search terms for retrieving relevant documents (Wildemuth, 2004).

Higher domain knowledge has not always lead to greater measures of search effectiveness. Zhang et al. (2005) found that users with higher domain knowledge were not more effective in their searches, as measured by Mean Average Precision (MAP). Higher domain knowledge as reflected in use of domain-specific search terms and tactics was also not associated with finding a greater amount of relevant documents (Vakkari et al., 2003), though it has been suggested that the combination of higher domain knowledge and sufficient knowledge of the search system positively impacts search performance (Vakkari et al., 2003).

Search strategies have been found to differ based on domain knowledge. Bhavnani (2001, 2002) found that users with high domain knowledge were more likely to access websites directly while users with lower knowledge found websites only through using search engines.

Wildemuth (2004) found that users in low states of domain knowledge required more search moves to find relevant information for solving problems posted to them. Hembrooke et al. (2005) defined nine domain knowledge-related search strategies based on the notion that domain knowledge is reflected in querying characteristics related to elaboration (for higher domain knowledge) or redundancy (for lower domain knowledge). The nine querying strategies were: elaboration, redundancy, broadening, refining, backtracking, plural making/taking, kitchen sink, poke-n-hope, and topic terms. Users with low levels of domain knowledge used querying strategies deemed less effective, such as plural making/taking, redundancy, poke-n-hope, and backtracking, while users with high domain knowledge issued queries in the strategy of elaboration and their queries were also more complex.

It has been found that cognitive components of domain knowledge may contribute to domain-specific search knowledge that leads to better search performance. Bhavnani (2001, 2002) found that when domain experts in healthcare and shopping searched within their areas of expertise, they were able to engage procedural knowledge and declarative knowledge in their specific areas of expertise and this enabled them to search more effectively. When searching outside their domains, they used general commercial search engines, modified queries, returned to earlier websites until they found the answers. Expert searchers, on the other hand, went to specialized websites for their tasks (e.g., a shopping expert did not use a commercial search engine but went directly to a consumer product review website to read reviews on the product from the information task) and followed more sequenced approaches to fulfilling the tasks they were given (e.g., the same shopping expert read the reviews on the review website, then compared prices using a price-comparison shopping site, then searched for coupons at the stores which had shown the lowest prices on the previous price comparison site). While these finding may indicate that users with high domain knowledge hold certain kinds of advantages when searching in their areas of expertise, the small sample size of this study (five experts in one category, four experts in the second category) points to the need for more research to be done with larger sample sizes appropriate for statistical inferences to be made.

2.2.2. The domain of personal finance topics. This subsection reviews studies from the literature in ILS and consumer finance that specifically investigated effects of domain knowledge in personal finance on online search and evaluation behaviors. Domain knowledge for the purpose of this proposal and the dissertation study is the domain of personal finance and is represented as the construct, *financial knowledge*.

2.2.2.1. Search and evaluation behaviors in the personal finance domain. Domain expertise in the context of financial knowledge has been examined in a few studies in ILS. Kuhlthau (1999) conducted a case study of a Wall Street investment analyst carried out over a five year period, to understand the ways in which the analyst's perceptions of uncertainty and complexity, and his construction and use of information sources changed as he progressed from being an industry novice with only three years' experience in his field to being an industry expert, recognized by the top trade publications as a leading authority in his research area. The analyst described having a low tolerance for the kind of information uncertainty requiring extensive information-gathering projects, such as that associated with having to cover analysis for an entirely new industry. He distinguished between routine and complex tasks during the novice period and the expert period, though he viewed far fewer tasks as complex as an expert. As a novice, his information needs centered around establishing his own knowledge base for understanding his industry, the need to be right in his work, and the need to establish a stable viewpoint about the markets and companies within the markets. As an expert, his information needs changed to needing to provide intelligent insights to his clients, without the emphasis on needing to be right or be viewed as right. He also indicated as an expert that he thrived on the wrong information put out by other industry analysts, because it allowed him opportunities to promote discussions with clients with better information. Sources of information for the analyst at both novice and expert stages were broken down into internal sources such as journals and trade publications, external sources such as annual reports and company visits, and institutional sources such as libraries. As an expert, the analyst placed much greater importance on certain information sources, like annual reports and company visits, and he also added a new source of information, his clients.

In a survey of 100 securities analysts from 40 large investment banking firms in the United States and the United Kingdom, Baldwin and Rice (1997) investigated the influence of individual characteristics and institutional resources on the information sources and channels used by analysts. They found that individual characteristics such as age, gender, marital status, education, membership in professional associations, number of hours worked, years of experience had little influence on the individuals' use of information and communication channels, which also meant that they did not directly or indirectly influence the outcomes of analysts' information activities (income, analyst ranking, number of industry reports written, and job satisfaction). However, institutional resources (e.g., staff size, budget, type and size of firm, location of firm, use of internal and external libraries and databases) were found to have a significant influence on analysts' information outcomes.

In a large-scale log analysis of information searchers on the Internet, White, Dumais, and Teevan (2009) found that users with greater financial expertise issued longer queries, used more domain-specific vocabulary in their queries, visited more webpages overall, visited more unique top-level domains, and visited webpages that had greater amounts of technical content. Domain knowledge has been found to impact the query vocabulary of users. For example, in the area of financial expertise, White et al. (2009) found that users with higher domain knowledge used more domain-specific vocabulary in their queries.

In a study by Berger and Messerschmidt (2009), evidence was found that among a representative sample of German citizens, financial knowledge influenced information searching in the personal finance domain. In this study, people with more knowledge of personal finance searched more extensively for information about personal finance.

2.2.2.2. Influence of cognitive abilities in the personal finance domain. In the consumer finance literature, researchers have sought to analyze the impact of cognitive abilities by using traditional IQ-type measures or proxies for general intelligence. For example, Li, Baldassi, Johnson, and Weber (2013) conducted a study on cognitive abilities to understand how abilities influenced decision-making in older adults. The researchers proposed a "complementary capabilities hypothesis," which means that increased levels of crystallized intelligence in older adults compensate for age-related declines in fluid intelligence, and that these stronger capabilities provide an alternative pathway for older adults when it comes to decision-making. They tested 173 younger adults aged 18-29 and 163 older adults aged 60-82 using eight measures of crystallized and fluid intelligence. Economic decision-making measures were used to measure six traits associated with personal economics: temporal discounting, loss aversion, financial literacy, debt literacy, susceptibility to anchoring, and resistance to framing. Financial literacy questions were taken from Lusardi and Mitchell (2006) and debt literacy questions were taken from Lusardi and Tufano (2015). They found that older participants generally made better decisions than younger participants in the form of more accurate responses to the financial and debt literacy questions and exhibited more patience. They also found that individual differences in cognitive abilities explained age relationships for temporal discounting, financial literacy, and debt literacy, but not for loss aversion. The researchers suggest that there are additional underlying aspects of the debt literacy trait which their measures did not capture such as domain knowledge or expertise.

Some studies have looked at cognitive abilities and decision-making mistakes, such as Agarwal and Mazumder (2013) and Gerardi, Goette, and Meier (2013). Agarwal and Mazumder (2013) conducted a data analysis of credit card and home loan data of U.S. military to identify ways in which cognitive abilities can be linked to making financial mistakes. They used two known financial mistakes as their dependent variables. The first mistake was related to credit card balance transfers. The second mistake involved home equity lines of credit. Scores on the Armed Forces Qualifying Test (AFQT) were used for cognitive abilities. The AFQT has two quantitative and two verbal components: arithmetic reasoning, math knowledge, paragraph comprehension, and word knowledge. In the credit card dataset (N=480), the researchers found that a one standard deviation increase in composite AFQT was associated with a 24% greater likelihood that the consumer would discover the best balance transfer strategy. Verbal scores were not associated with the balance transfer mistake. In the home loan dataset, one standard deviation increase in AFQT was associated with an 11% decrease in the likelihood that the consumer would make the rate-changing mistake. Verbal scores were much less strongly associated with this mistake. The math scores were strongly associated with avoiding both mistakes. Especially in the case of the home equity loans, where the rate changing mistake increased the annual percentage rate (APR) of the loan by 269 basis points or about an extra

$4,000 over the 5-year life of the loan, the mistakes have a high cost. Word knowledge was somewhat important in the home equity decision.

A similar study conducted by Gerardi et al. (2013), linked numerical ability to mortgage default. Using administrative data and the results of their telephone survey (in which they measured numerical ability over the phone), Gerardi et al. (2013) were able to distinguish between product choice and human behavior in the subprime market. That is, they were able to show whether subprime defaults were the result of mortgage product choice by the consumer or through the consumer's behavior once they have the loan. The dataset was a proprietary dataset of subprime mortgage contracts from New England and total sample was N=339. Borrowers with the lowest numerical ability spent 25% of the first five years of their loans in delinquency while those with the highest ability only spent 12% of their first five years in delinquency. After re-estimating the relationship against other independent variables like age, sex, education, FICO score, labor market, etc., and finding that no other variables changed the outcome, the researchers concluded that mortgage delinquency "seems specifically associated with numerical ability, not with general IQ levels or economic literacy" (Gerardi et al., 2013, p. 11269). However, verbal IQ did show a statistically significant negative correlation with the incidence of actual foreclosure. In other words, higher IQ did not help people from getting behind on their payments but it helped them avoid foreclosing on their homes. Importantly, researchers used additional details about the mortgage and borrower characteristics to see if the mortgage attributes of participants' loans were correlated with numerical ability. When control variables were added (this is called the "conditional" correlation), interest rate was found to be negatively correlated with numerical ability, thus, those with higher numerical abilities had lower interest rates on their loans on average.

2.3. Mental Workload

There is no well-known, formal definition for the psychological construct called *mental workload* that is accepted in all disciplines that study it. One that is useful in the realm of interactive searching proposes that "mental workload is the operator's evaluation of the attentional load margin (between their motivated capacity and the current task demands) while achieving adequate task performance in a mission-relevant context" (Jex, 1988, p. 11). There are many aspects of interactive searching that produce mental workload for users. In this section, studies are reviewed that describe how mental workload is induced by different kinds of search interfaces, querying activities, document evaluation, task characteristics, and individual differences.

2.3.1. The mental workload of querying. Several studies have investigated mental workload involved in querying search systems. Two studies found that when considering the stages of the user's search process, the one that creates the greatest amount of mental workload is the query formulation stage (Gwizdka, 2010; Shovon, Nandagopal, Du, Vijayalakshmi, & Cocks, 2015). Gwizdka (2010) conducted a search study of 48 users and using a dual-task method for measuring mental workload (which he called *cognitive load*). He found that mean reaction times in the dual task were significantly longer for users during the query stage than during stages for viewing search results or page content. Shovon et al. (2015) used neuroimaging techniques to construct time series observations of users' brains while conducting search tasks. They found that brain activity signifying mental effort was greatest during the query formulation stages of users' search, versus the search results list viewing or content viewing stages.

Some studies have found that mental workload can vary based on the method of inputting queries. Azzopardi, Kelly, and Brennan (2013) manipulated the layout of the query input area of search interface so that instead of typing query words and phrases into a blank white rectangular box, user would have to type one query word at a time into separate query word boxes. As was expected, the structured query interface influenced the search behaviors of users. They submitted fewer queries, spent more time on SERPs, examined more documents per query, and viewed deeper levels of the SERP than users assigned to the standard query interface condition (i.e., the single, rectangular box). In this study the cost of search was operationalized as the time required to submit a query. Using the GOMS model, it was determined that the greatest *cost* or time to query would be associated with the structured interface or a query suggestion interface. Cost, as in time cost, of querying was greatest for the structured query interface; however, this cost did not translate into an experience of greater mental workload for users. The results of the analysis of the self-report mental workload measure, the NASA-TLX (Hart & Staveland, 1988) were not statistically significant, however, they did exhibit a consistent trend in mental workload. The NASA-TLX (TLX) measures the mental, physical, and temporal demands imposed on individuals by tasks along with individuals evaluations of their performance, effort, and experienced frustration. In the study by Azzopardi et al. (2013), users who experienced the greatest mental demand were users of the standard query interface, followed by the structured query interface and the query suggestion interface. At the same time, users of the standard query interface posed the most queries, viewed the fewest documents per query, and investigated the SERPs at the shallowest levels. This may indicate that search cost measured in terms of time spent querying does not translate into perceived mental workload for users and interaction cost may not be equivalent to experiences of mental demand or mental workload.

Edwards, Kelly, and Azzopardi (2015) used the same experimental set up as Azzopardi et al. (2013) and in addition to investigating mental workload, also studied the construct of stress by measuring users' electrodermal activity and responses to a stress questionnaire. Users in the standard query interface condition reported greater mental demand, temporal demand, and effort than those in the structured query interface condition. While these findings were not significant, they supported similar findings in Azzopardi et al. (2013). The standard query interface users also reported feeling less successful in their searching than users in the structured condition, and this finding was significant. Also interesting is a comparison of some of the search behaviors between the two studies. In Azzopardi et al. (2013), users in the standard query interface condition submitted the most queries of all conditions, viewed the fewest documents per query, and were the shallowest in terms of the depth of their investigations of SERPs. Users in Edwards et al. (2015) also submitted the most queries of all conditions but hovered and went to greater depths on SERPs, a finding that is opposite to Azzopardi et al. (2013). This is a good example showing how individual measures of search behavior – in this case viewing depth on the SERP – can vary even in similar search situations, and so it is very important to use a variety of methods for measuring the complex construct of mental workload.

2.3.2. Mental workload created by features of the SERP. Several studies investigated different manipulations of SERPs to understand their impact on mental workload. Kelly and Azzopardi (2015) compared users' search behaviors and experiences with SERPs of three different sizes – three results per page (RPP), six RPP, and ten RPP. Users in the ten RPP condition viewed more documents and selected more relevant documents than the other two conditions (and these differences were significant from six RPP condition). Users in this condition also reported the highest scores for all NASA-TLX mental workload items except physical demand, with the largest difference showing up in the mental demand scores. Though not significant, the findings on mental workload are in the direction of a larger trend in which users searching on standard, rectangular search box/10-link search interfaces report greater mental workload than alternative search interfaces for entering queries and viewing SERPs.

In another study of SERP feature manipulations, Bota, Zhou, and Jose (2016) investigated the effects of SERP entity cards on search behaviors and mental workload. Entity cards were defined as on- or off-topic and diverse or not diverse, where diversity was defined as a card showing information sections containing Wikipedia text, images, and links to similar search results. Cards were either present or not present on SERPs, depending on the experimental condition. In conditions comparing SERPs showing entity cards versus SERPs not showing entity cards, researchers found that users experienced greater mental workload when viewing SERPs that showed entity cards if the cards were off-topic, than when they viewed SERPs that did no show any cards. This was not the case when entity cards were on-topic, in that mental workload was not greater for the card versus no card condition. In other words, information from the on-topic entity cards was integrated into users' processing without additional mental effort. In terms of information diversity of entity cards, researchers found that diverse cards created less mental workload for users than non-diverse cards and this was especially the case when the cards were also on-topic. Overall, the findings emphasized the importance that summarized information on SERPs, in the form of entity cards, is on-topic and diverse, to ensure lower mental workload for users.

2.3.3. Mental workload of evaluating documents. Some studies have investigated the mental workload created when users evaluate documents for relevance. Villa and Halvey (2013) asked 49 users to assess documents of different lengths from the TREC collection (AQUANT) for three levels of relevance – not relevant, relevant, and highly relevant. They found that the middle level of judgment, *relevant*, was perceived as requiring the most mental workload by participants. They also found that mental workload tended to increase as document length increased.

Gwizdka (2014) investigated users' binary relevance assessments of short, online news documents from the AQUAINT corpus in a laboratory experiment. Gwizdka's interest was to understand the degrees to which relevance affected how people read documents, the amount of cognitive effort invested in reading documents, and the extent to which eye tracking data could be used to infer relevance. Using cognitive effort measures derived from eye fixation measures such as reaction time, length of reading sequences, duration reading, duration scanning, number of reading sequences, and total number of fixations, Gwizdka found that judging topically relevant documents required the most cognitive effort followed by relevant and not relevant ("irrelevant") documents. He also measured regressions, which are eye movements that backtrack from their current location during the normal course of reading and found that the number of regressions for not relevant documents was 10%, followed by 20% for topically relevant documents, and 25% for relevant documents. This meant that participants did more re-reading of words than would be expected when reading topically relevant and relevant documents (typical regressions during reading are about 10-15%). It was found that participants re-read the relevant word or phrase just prior to making the relevance judgment. In terms of pupil dilation measures, participants had the largest pupil dilation measures for relevant documents and the smallest for not relevant documents, with pronounced effects during the one second prior to relevance judgment.

2.3.4. Mental workload, task characteristics and individual differences. Other aspects of interactive searching that have been studied in the context of mental workload are task characteristics and individual differences. Brennan et al. (2014) studied the effect of cognitive abilities on mental workload of adults searching tasks of varied complexity and found that participants with lower levels of the cognitive ability *perceptual speed* reported greater workload than those with higher levels of that ability, especially for the task considered most complex. In terms of the six dimensions of the NASA-TLX, the effect size was the greatest for physical demand and frustration dimensions, with participants who had lower levels of perceptual speed reporting more than twice the physical demand while completing *analyze* and *create* tasks (though this 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 the *create* tasks (with the experience again rated only near the midpoint of the scale).

2.4. Information Searching and Evaluation

The literature on information searching and evaluation is vast and this brief section is not intended to review it all. Instead, this section focuses on a specific subset of studies that motivated the design of the dissertation study.

2.4.1. Information searching.

Information searching, as defined by Wilson (1999), describes the set of behaviors used by a person for interacting with a computerized search system to search for textual data.

2.4.1.1. Search behaviors. Search behaviors are typically studied by examining users' interactions with the search system, analyzing users' self-report data from questionnaires, and interpreting signals from physiological measures. This section examines literature about variables investigated in the dissertation study – queries, mouse clicks, search engines results pages (SERPs) and webpages.

Query formulation can be seen as an essential part of successful information searching. Users express their information needs to textual information systems by submitting text queries. Early studies of information searching on the Internet indicated that people searched the Internet differently than people searched bibliographic and other types of defined-corpus systems. In their querying behavior, users submitted shorter queries and rarely modified them (Silverstein, Marais, Henzinger, & Moricz, 1999). In large-scale search engine log studies, users submitted less than three terms per query (Jansen, Spink, & Saracevic, 2000). Longer queries have been proposed as having advantages in search systems (Aula & Käki, 2003) and efforts were made in the past (Belkin et al., 2003) to elicit longer queries from users. However, others have found that shorter queries work well with the exact-match term techniques of Internet search engines (Downey, Dumais, Liebling, & Horvitz, 2008). Query length has been found to be affected by search expertise (Hölscher & Strube, 2000; Vakkari et al., 2003), topic familiarity and domain expertise (Hölscher & Strube, 2000; Vakkari, 2000), and task type (Toms et al., 2008).

Mouse clicks are used as implicit measures of engagement and interest. What people click on has been used as an implicit signal for what is important to them, to indicate what people think which results will meet their information needs. There is much evidence that users pay more attention to the top part of SERPs and exhibit "click-bias," in that they tend to click on the higher-ranked results on the SERP to the detriment (or complete exclusion) of lower-ranked results (Buscher, Dumais, & Cutrell, 2010; Dumais, Buscher, & Cutrell, 2010; Joachims, Granka, Pan, Hembrooke, & Gay, 2005; Pan, Hembrooke, Joachims, Lorigo, Gay, & Granka, 2007). In a more recent study Thomas, Scholer, and Moffat (2013) used eye tracking and discovered that although users generally proceeded from the top of the SERP to the bottom as has long been assumed, they did so in a more sophisticated fashion in which they read several snippets in a "zone of interest," moving forwards and backwards within that zone as they compared its results and before clicking through to a webpage. Users maintained "active bands" of relevant SERP snippets in their memory as they shifted their attention progressively downward on the page. More than 60% of the time, participants in the study (N=34) did not begin their investigations of the SERP with the first result. SERP results in the second and third rank on the page were actually the most common results viewed first by users. In addition to the finding itself being interesting (e.g., did users skip the first result because they thought it was an advertisement – i.e., maybe this is the discovery of a new form of ad banner blindness?), it also suggests an area where the study of individual differences could provide valuable information. For instance, individual differences in abilities may drive the size of the viewing range such that some users may maintain much broader or narrower “zones of interest.” This information could enable search engines to optimize SERPs for specific populations of their users to enhance selection of relevant results and to improve their overall search experience. More generally speaking, understanding behavior on SERPs is important to bear in mind as it might be the case that strategies may be driven by differences at the individual level, such as speed-related processing factors, memory capacity, or other factors.

Click behaviors on SERPs have been found to reveal user characteristics. SERP clicks have been interpreted as indicators of how active users are in their searching (Arguello, 2015) and this has been done by measuring the frequency of clicks on SERPs (Brennan et al., 2014) as well as the number of clicked results (Jiang, He, & Allan, 2014). The amount of time taken before clicking on the first relevant result has been found to be related to cognitive ability (Al-Maskari & Sanderson, 2011).

Behavior on webpages is also very important to understand. Buscher, Cutrell, and Morris (2009) found that individual differences influenced how participants attended to different aspects of webpages. Women gazed longer and at more of the page; users older than 30 gazed longer;

and those who had familiarity with the web site looked longer at top left, top right and bottom left. More experienced users looked longer at pages during page recognition tasks.

2.4.1.2. Search strategies. Typically, search strategies are seen as composite structures made up of actions described as moves (Fidel, 1985) and tactics (Bates, 1979). For example, Xie and Joo (2010) say that a search strategy "highlights a working plan and interactive reaction for a given situation. A search strategy consists of a series of sequential tactics that take into account both planned and situational elements" (pp. 259-260). Vakkari (2003) tells us that search strategies are some combination of terms, operators, and tactics of the information searcher.

Key studies of Web-based search strategies are summarized in Table 1. The earliest studies focused on basic browsing strategies (Catledge & Pitkow, 1995; Navarro-Prieto, Scaife, & Rogers, 1999), classifying different types of search behaviors (Hawk & Wang, 1999), or exploring differences between types of searchers (Hölscher & Strube, 2000). An early study of consumer health information searching strategies by Eysenbach and Köhler (2002) found that focus group participants articulated strategies for searching that involved assessing websites for credibility by looking at the source of the webpage or looking for a site that had a professional design or scientific tone to the writing. In the second part of the study, log data indicated that users continued to search after they found the answer to their questions, because they did not understand what they had found. Searching for debt-related information may evoke similar kinds of behaviors in searchers

As the Web grew more popular, researchers focused their studies on SERP strategies. Klöckner, Wirschum, and Jameson (2004) identified the depth-first and breadth-first strategies using eye tracking, while Aula, Majaranta, and Raiha (2005) used eye tracking and defined users' search strategies as *economic search* or *exhaustive search*, in which economic searchers would look at a result and make their decision immediately, while exhaustive searchers would carefully evaluate the entire SERP before making a decision about which item to click. Dumais et al. (2010) also used the *economic* and *exhaustive* searcher categories in their eye tracking, but differentiated between those searchers who paid attention to ads and those who did not. Thatcher (2006) identified 12 cognitive search strategies that included various forms of “safe player” strategies, parallel window searching, link-dependent strategy, to-the-point strategy, known address strategy, sequential player strategy, deductive reasoning strategy, virtual tourist strategy, and parallel hub-and-spoke strategy. Xie and Joo (2010) conducted a study of 31 adults and through the analysis of the log data and concurrent think aloud interview transcripts, they identified eight search strategies: iterative result evaluation, iterative exploration, whole site exploration, multiple query reformulation, simultaneous multiple resource search, item comparison, query initiation, and known-item initiation. Ondrusek, Ren, and Yang (2017) conducted a study of 35 MLIS students and were able to categorize search strategies in four ways: conceive, combine, design, and evaluate. Savolainen conducted two conceptual analyses about search strategies. In the first, he posited strategy as a plan and strategy as a pattern of actions (Savolainen, 2016) and in the second, he posited three main heuristic elements to strategies: the familiarity heuristic, the recognition heuristic, and the representativeness heuristic.

Table 1. *Summary of Search Strategies for Web Searching found in ILS Research*

Authors	Study Methods	Sample	System	Search Strategies
1. Catledge & Pitkow (1995)	Search log analysis – 3 weeks' data	N = 107 GIT computing staff, students, faculty	WWW, using NCSA Mosaic	Serendipitous browser, general purpose browser, searcher
2. Navarro-Prieto et al (1999)	Retrospective think aloud	N = 23 CompSci & Psych students	WWW, using Netscape Communicator 4.5	Top-down, bottom-up, mixed
3. Hawk & Wang (1999)	Questionnaires Cognitive Styles Transaction logs Concurrent Think Aloud	N = 24 Graduate students	WWW	Surveying, double-checking, exploring, link following, back and forward going, shortcut seeking, engine using, loyal engine using, engine seeking, metasearching
4. Eysenbach & Köhler (2002)	• Focus groups • Observation of live web searching with Think Aloud • Retrospective Interviews	• Focus Group N=21 • Search Group N=17 • Interviews N=17	Live Search: WWW	• Look for website credibility, source, professional design, a scientific or official touch, language, ease of use. • Live Searchers – 763 different webpages. Did not use medical portals, used only one search term, only one used Boolean, most chose results 1-3 on SERP then reformulated query without going to 2nd page, continued searching even after finding correct answer because they did not understand the information they found. • Interviews – Very few participants remembered which websites they had retrieved information from.
5. Klöckner, Wirschum, & Jameson (2004)	Eye tracking Mouse tracking	Exp. 1 N = 41 Exp. 2 N = 27	Closed system of Pre-created SERP list of 25 results from Google	Depth-first and breadth-first
6. Aula, Majaranta, & Raiha (2005)	Eye tracking	N=28	Pre-defined queries, result pages saved locally.	Economic search and exhaustive search

Authors	Study Methods	Sample	System	Search Strategies
7. Thatcher, A. (2006)	• Retrospective verbal protocol • Logfile analysis • Observation	N=80	WWW	Safe player broad first strategy; safe player search engine narrowing down strategy; safe player search engine player strategy; safe player known address search domain strategy; parallel player strategy; link-dependent strategy; to-the-point strategy; known address strategy; sequential player strategy; deductive reasoning strategy; virtual tourist strategy; parallel hub-and-spoke strategy.
8. Dumais et al (2010)	Eye tracking	N = 38	WWW (controlled SERP)	Exhaustive searchers, economic-results searchers, economic-ads searchers
9. Xie & Joo (2010)	Search logs Concurrent think aloud	N = 31	WWW	Iterative result evaluation, iterative exploration, whole site exploration, multiply query reformulation, simultaneous multiple resource search, item comparison, query initiation, known-item initiation
10. Savolainen (2016)	Conceptual analysis	N=57 research studies	N/A	Strategy as Plan and Strategy as Pattern of Actions Intended versus Realized Strategies and Deliberate versus Emergent Strategies
11. Savolainen (2017)	Conceptual analysis – heuristic elements	N=31 research studies	N/A	Familiarity heuristic, recognition heuristic, representativeness heuristic
12. Ondrusek, Ren, & Yang (2017)	Content analysis	N = 35 MLIS students	WWW plus library databases	Conceive, combine, design, evaluate

2.4.2. Information evaluation. Information evaluation is also known as relevance assessment. The earliest debates about relevance in information science occurred at the 1958 International Conference for Scientific Information (ICSI) (Rees & Saracevic, 1966). The resulting consensus fell along four ideas about relevance: 1) relevance is more than comparing internal performance within systems; 2) relevance is not the exclusive property of document content; 3) relevance is not dichotomous; and 4) there is such as thing as "user relevance" that can be judged (Rees & Saracevic, 1966, p. 6). Since that time, scholars have provided extensive reviews of relevance literature during the past forty years that record and summarize the main directions of thinking and conceptualizations of relevance in information science (e.g., (Borlund, 2003; Mizzaro, 1997; Saracevic, 1975, 2007a, 2007b; Schamber, 1994; Schamber, Eisenberg, & Nilan, 1990).

2.4.2.1. Evaluation behaviors. User are known to be less than diligent when evaluating online information. As far back as the early Usenet groups, researchers realized that users often did not read online information documents in their entirety (Morita & Shinoda, 1994). Users scan the textual content of webpages. Weinreich, Obendorf, Herder, and Mayer (2008) logged the web browsing behaviors of 25 participants for periods from two to four months and found that of the approximately 60,000 first-page visits to non-SERP webpages, nearly half of the pages were visited for less than 12 seconds. Only 11.6% of first-page visits lasted longer than two minutes. Rokhlenko, Golbandi, Lempel, and Leibovich (2013) found that 30% of their 204 study participants read zero to one paragraph of the four-paragraph webpages in the task. About 25% read two to three paragraphs, while about 46% read all four paragraphs. Possible reasons for the lack of thorough reading behavior on webpages may be related to the mental challenges associated with reading hypertext (DeStefano & LeFevre, 2007), but further research on this proposed answer has not been conducted. Yilmaz, Verma, Craswell, Radlinski, and Bailey (2014) suggested that users do not read entire pages sequentially and analysis by Guo and Agichtein (2012) pointed to several patterns of reading and scanning in post-click searcher behaviors. When documents were relevant, users slowed down and moved the mouse horizontally while reading text. They also tended to focus their attention in just a few places, whereas in non-relevant documents their attention was distributed more evenly across the page. All of these studies have advanced our understanding of users' behaviors when they are evaluating information but clear links to how these behaviors affect relevance outcomes are still unknown.

2.4.2.2. Evaluation strategies. Several studies have investigated evaluation strategies. Tombros, Ruthven, and Jose (2005) used think aloud techniques to elicit information from searchers about what features of the Web pages helped them determine the usefulness of the pages for each of their tasks. Participants assessed the utility of 862 web pages and gave 1,602 mentions of page features. Features reported were webpage content, structure, and quality. Xu and Chen (2006) used reported the results of a pilot (N=72 document evaluations) and main study (N=132 student participants, 264 document evaluations) of judgment criteria participants used to assess the relevance of documents related to four search tasks. Their results also indicated that topicality and novelty were the most significantly associated with relevance judgment.

Crystal and Greenberg (2006) reported on a user study (N=12) of motivated seekers of health information research on the Web. Participants were asked to assess the relevance of health information. The researchers identified eight categories of document criteria: affiliation of information, authority/person authoring the information, data provided, influence of the work, methods employed by the researchers, scope of the work, topic of the work, and characteristics of the document. In an exploratory study conducted in 2000-2001 in Finland, Savolainen and Kari (2006) investigated how participants (N=18) judged the relevance of hyperlinks and Web pages while searching for self-generated topics related to everyday life. The researchers identified 18 different user-defined relevance criteria total, with four main criteria used most often: specificity, topicality, familiarity, and variety. Xie, Benoit, and Zhang (2010) reported the results of an empirical study of 31 Milwaukee-area adult participants using self-generated search tasks (one work-related and one personal-related). Participants employed 18 types of relevance criteria classified into four categories: content coverage, document quality, design, and accessibility. Xie and Benoit (2013) found when evaluating lists, users made “snap” decisions, taking only a few seconds to decide about the relevance of the item, whereas in document evaluation, users spent much more time evaluating. Balatsoukas and Ruthven (2012) found that users employed 12 relevance criteria for judging surrogate information: topicality, scope, user background, quality, tangibility, resource type, affectiveness, recency, ranking, serendipity, format, and document characteristics. Participants spent more time fixating on surrogate information related to topicality, scope, user background, and quality than the other criteria.

Compilation of user-identified relevance criteria. Ultimately, across all of the studies reviewed, it was found that researchers asserted it is not enough to focus solely on putting comprehensive, quality information on a website, but that issues related to design, accessibility, and item characteristics also need to be taken into consideration in relation to the SERP organization and layout. A compilation of common user-identified relevance criteria across many of the studies reviewed is shown in Figure 2.

Category	Criteria	Schamber, 1991	Park, 1992	Cool et al. 1993	Barry, 1993, 94	Bateman, 1999	Spink et al. 1999	Maglaughlin et al., 2002	Tombros et al., 2005	Crystal & Greenberg, 2006	Savolainen & Kari, 2006	Xu & Chen, 2006	Xie, Benoit, & Zhang 2010	Xie & Benoit 2013
Author	Credibility/status	-	✓	✓	✓	✓	-	✓	-	✓	-	-	-	✓
Content	Subject matter - topicality, aboutness	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	-	-
	Breadth - completeness, depth, level, scope, specificity	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	-	✓
	Quality - accuracy, credibility, validity, verifiability	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓
	Clarity - presentation quality, readability, understandability	✓	✓	✓	✓	✓	-	-	-	-	✓	✓	✓	-
	Novelty - new information	-	✓	✓	✓	✓	✓	✓	✓	-	-	✓	✓	✓
	Connections - lists, links to other information		✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
	Background information or data	-	✓	✓	✓	✓	-	✓	-	✓	-	-	-	-
	Methodological information	-	✓	✓	✓	✓	-	-	-	✓	-	-	-	-
	Stimulus - thought catalyst, novel information	-	-	✓	✓	✓	-	✓	-	-	✓	-	-	-
	Geographic focus - proximity, coverage	✓	-	-	-	-	-	-	-	✓	-	-	-	-
Full text	Currency - recency, timeliness	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓
	Document or article type	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	-
	Accessibility - availability, obtainability	✓	✓	-	✓	✓	✓	-	-	-	✓	-	✓	-
	Novelty	-	✓	-	✓	-	✓	✓	-	-	-	-	-	-
	Utility	-	✓	✓	✓	✓	✓	✓	-	-	-	-	-	✓
Source	Authority - quality, reliability, reputation, value, visibility	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Novelty	-	✓	✓	✓	-	✓	✓	-	-	-	-	-	-
User	Affectiveness - appeal, competition	✓	✓	✓	✓	✓	✓	✓	-	-	✓	-	-	-
	Experience - understanding, familiarity, prior knowledge	-	✓	✓	✓	✓	✓	✓	-	-	✓	✓	-	-
	Requirements - time constraints	-	✓	-	✓	-	✓	✓	-	-	✓	-	-	-

Figure 2. Compilation of user-defined relevance criteria from the ILS literature.

CHAPTER 3: RESEARCH QUESTIONS AND THEORETICAL MODELS

The objective of this dissertation research is to investigate the influence of cognitive abilities and financial knowledge on adults searching the Internet for information about different kinds of financial loans. This section covers the research questions, theoretical models, and hypotheses for the study.

3.1. Research Questions

The following are the research questions of this dissertation:

1. How do cognitive abilities and financial knowledge influence the *search performance*, *relevance assessments*, and *mental workload* of adults searching the Internet for information about different kinds of financial loans?
2. What are users' *strategies for finding and evaluating information* on the Internet about different kinds of financial loans? How do users' cognitive abilities and financial knowledge influence these strategies?

The research models tested in the dissertation study propose that two independent variables – cognitive ability and financial knowledge – influence users' behaviors in searching for and evaluating information, and their experiences of mental workload in specific ways such that individual differences can be detected across users. The two cognitive abilities investigated, perceptual speed and working memory, are part of Three Stratum Theory of cognitive abilities (Carroll, 1993) and the financial knowledge variable is drawn from readings in cognitive economics and behavioral finance literature.

3.2. Research Question #1 – Theoretical Model, Hypotheses, and Statistical Models

The first research question asks "*How do cognitive abilities and financial knowledge influence the search performance, relevance assessments, and mental workload of adults searching the Internet for information about different kinds of financial loans?*" The three outcomes – search performance, relevance assessments, and mental workload -- are the dependent variables. They are constructs measured through operationalizations of user behaviors. The *search performance* construct is operationalized using search behaviors such as mouse clicks, search queries, and webpage views. The *relevance* construct is operationalized with participants' explicit relevance judgments of webpages. The construct of *mental workload* is operationalized using self-report measures from questionnaires and signals interpreted from eye gaze behaviors.

The proposed model of the relationships between the independent variables and the dependent variables is shown in Figure 3. It was developed with reference to Three Stratum Theory of cognitive abilities (Carroll, 1993), empirical evidence from research in IIR, and evidence from the literature in cognitive economics and behavioral finance. In the next five subsections, hypotheses and exploratory relationships are proposed for the relationships between the independent and dependent variables (DVs) as follows: influence of perceptual speed on the DVs, influence of working memory on the DVs, influence of financial knowledge on the DVs, and two interaction models. The first model is the interaction of perceptual speed and financial knowledge on the DVs and the second is the interaction of working memory and financial knowledge on the DVs.

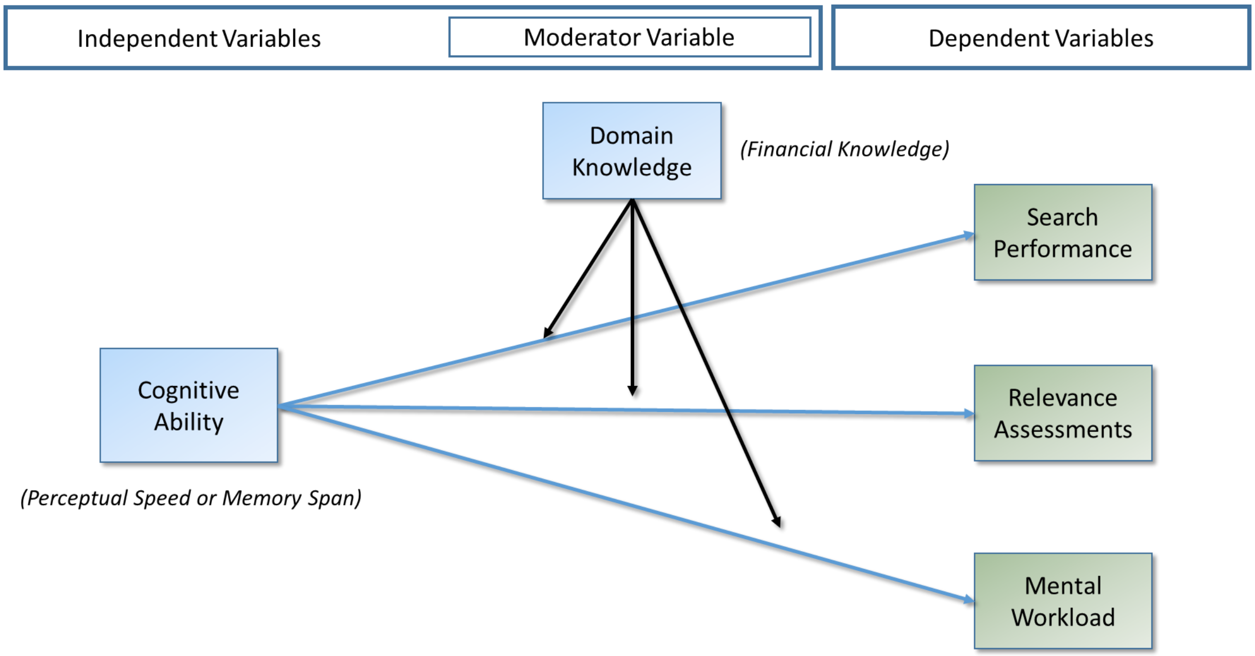


Figure 3. Theoretical model showing the influence of two independent variables – cognitive ability and domain knowledge (as the moderating independent variable) – on three dependent variables: search performance, relevance assessments, and mental workload.

3.2.1. Perceptual speed model. Evidence from the literature indicates that the cognitive ability *perceptual speed* impacts the search and evaluation behaviors and the perceived mental workload of information searchers. Findings from the literature support the development of three hypotheses which are explained in the next three subsections.

3.2.1.1. Perceptual speed - influence on search behaviors. Users with high perceptual speed ability have been found to interact more with search systems. For example, in Brennan et al. (2014), participants with higher perceptual speed ability issued longer queries on average per task, had more SERP clicks, viewed more URLs per query, and viewed more URLs per task. Therefore, it is hypothesized:

> ***Hypothesis 1*:** Participants with higher perceptual speed ability will interact more while searching which will manifest by issuing longer queries, having more clicks on SERPs, and viewing more URLs per query and per task than participants with lower perceptual speed ability.

3.2.1.2. Perceptual speed - influence on relevance assessments. Perceptual speed ability has been found to impact behaviors related to evaluating information. Al-Maskari and Sanderson (2011) discovered that higher ability participants found their first relevant documents faster than participants with lower ability. Therefore, it is hypothesized:

> ***Hypothesis 2a*****:** Participants with higher perceptual speed ability will bookmark their first relevant webpages faster than those with lower perceptual speed ability.

Allen (1994) found that participants with higher perceptual speed achieved greater precision and recall. In addition, in several studies involving information tasks requiring users to analyze and compare academic performance data, Steichen, Conati, and Carenini (2014) and Toker, Conati, Carenini, and Haraty (2012) found that users with low perceptual speed performed slower and with less accuracy. These comparison tasks seem similar to ones involving comparison of other kinds of numeric data, such as debt-related products that charge differing amounts of interest across different categories. Thus, the translation of reduced speed and accuracy is argued to be in line with the ability of a person to select relevant documents in a given amount of time. This can be measured as *interactive user precision* (Veerasamy & Heikes, 1997)[4]. Therefore, it is hypothesized:

> ***Hypothesis 2b*****:** Participants with higher perceptual speed ability will achieve greater interactive user precision than those with lower perceptual speed ability.

3.2.1.3. Perceptual speed - influence on mental workload. Perceptual speed ability has also been found to impact users' experiences of mental workload. In Brennan et al. (2014), participants with lower perceptual speed reported greater mental workload than those with higher perceptual speed. Steichen et al. (2013) found that participants with higher perceptual speed had

[4] Interactive user precision is described in Kelly (2009, pp. 112-114) as a TREC precision measure that compared TREC relevant documents with users' saved document. In the case of the dissertation study, expert assessors' judgments would take the place of TREC-relevance assessments.

lower mean fixation durations and standard deviations of fixation durations. Therefore, it is hypothesized:

> ***Hypothesis 3:*** Participants with higher perceptual speed ability will experience less mental workload, manifested in lower eye gaze measures for mean fixation duration and standard deviation of fixation durations, and lower scores on the self-report mental workload questionnaire, than those with lower perceptual speed ability.

3.2.1.4. Statistical model showing the influence of perceptual speed on the dependent variables. Figure 4 shows a statistical model of Hypotheses 1, 2, and 3, in which perceptual speed influences each of the three dependent variables.

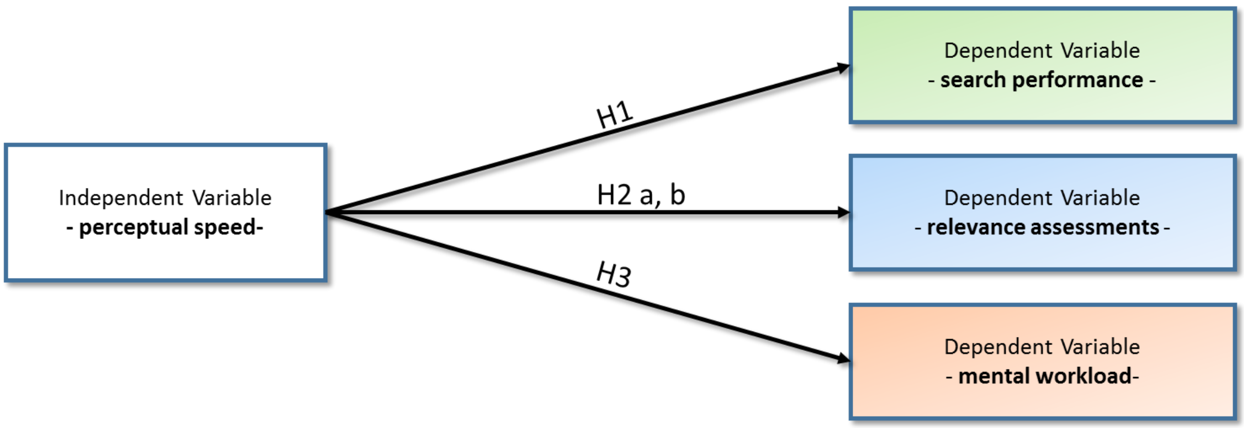


Figure 4. Statistical model showing the influence of perceptual speed on the dependent variables, with arrows labelling the hypotheses.

3.2.2. Working memory model. Findings from the literature support the development of two hypotheses relevant to this part of the model.

3.2.2.1. Working memory – influence on search behaviors. Working memory has been found to play a role in certain search behaviors. Gwizdka (2013) found that in an experimental system, users with lower working memory selected fewer word tags and opened fewer documents than those with higher working memory (the act of selecting and de-selecting tags was considered a proxy for querying behaviors). Therefore, it is hypothesized:

> ***Hypothesis 4a*:** Participants with higher working memory ability will issue more unique queries and open more webpages than those with lower working memory ability.

Eye movement metrics that have been used as indicators of memory load include fixation count and fixation durations (Rayner, 1998). In general, research in reading, scene perception, usability, and visual search has found that fixation durations increase in length as cognitive processing becomes more effortful (Holmqvist & Nystrom, 2011). Visual search research explicitly links this cognitive processing to increases in memory load (Meghanathan, van Leeuwen, & Nikolaev, 2015; Peterson, Beck, & Wong, 2008). Based on these findings, it seems that participants with lower working memory ability would have longer average fixation durations while searching because their memory load would be greater than those with higher working memory. Therefore, it is hypothesized:

> ***Hypothesis 4b:*** Participants with lower working memory will have more fixations on average on SERPs and webpages and will also have longer fixation duration measures.

3.2.2.2. Working memory – influence on relevance assessments. Working memory has been found to affect relevance judgments of documents. MacFarlane et al. (2012) found that users with lower working memory judged fewer documents as not relevant. Therefore, it is hypothesized:

> ***Hypothesis 5*****:** Participants with lower working memory will be less selective in their evaluation behaviors such that, of the webpages they view, these participants will have a lower proportion of not relevant webpages than those participants with higher working memory.

3.2.2.3. Working memory – influence on mental workload. While intuitively it seems that users with lower working memory would experience greater mental workload in complex tasks than their peers, evidence from the literature does not fully support this (Brennan et al., 2014). Thus, there is no hypothesis for the effect of working memory on mental workload. Instead, several measures will be explored to uncover possible relationships between the variables.

3.2.2.4. Statistical model showing the influence of working memory on the dependent variables. Figure 5 shows the influence of working memory on the three outcome variables. A possible relationship between working memory and mental workload is represented by the arrow labeled "Exploratory Measure #1."

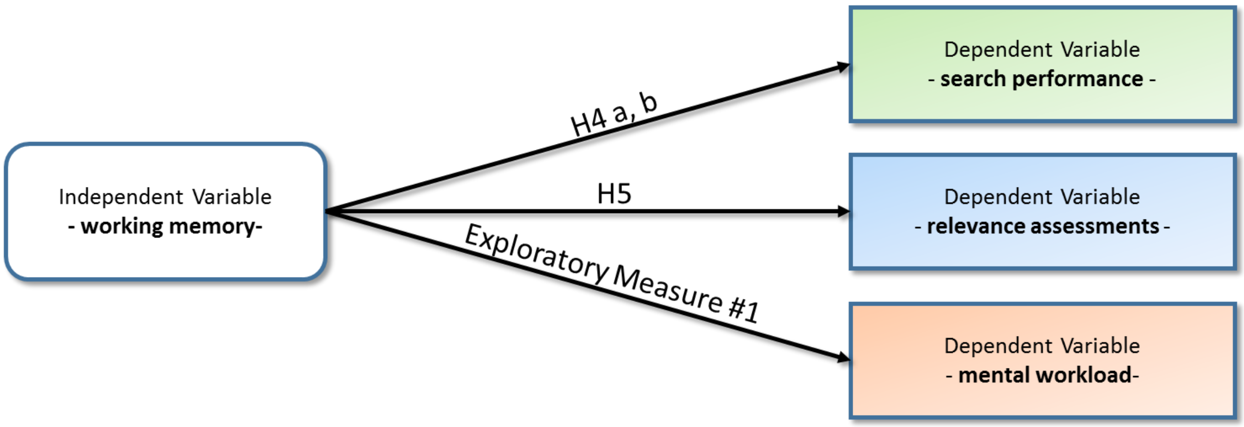


Figure 5. Statistical model showing the influence of working memory on the dependent variables, with hypotheses and the exploratory relationship labelled.

3.2.3. Financial knowledge model. Knowledge of the domain of personal finance in this dissertation is represented as the construct of *financial knowledge*. Findings from the literature support two hypotheses.

3.2.3.1. Financial knowledge - influence on search behaviors. Evidence from the literature indicates that financial knowledge influences the searching behaviors of consumers, such as in Berger and Messerschmidt (2009), where people with more financial knowledge searched more extensively for information about personal finance. In a large-scale log analysis of information searchers on the Internet, White et al. (2009) found that users with greater financial expertise issued longer queries, used more domain-specific vocabulary in their queries, visited more webpages overall, visited more unique top-level domains, and visited webpages that had greater amounts of technical content. Findings from research about other domains support this as well. Specifically, users with higher domain knowledge have been found to issue longer queries (Hembrooke et al., 2005; Zhang et al., 2005) or issue longer queries when searching in their area of expertise versus other topics (Freund & Toms, 2006). Zhang et al. (2005) found that users with higher domain knowledge in the areas of engineering and science issued more queries per task. Therefore, it is hypothesized:

> ***Hypothesis 6*****:** Participants with higher levels of financial knowledge will issue longer queries and more queries than participants with lower levels of financial knowledge.

3.2.3.2. Financial knowledge - influence on relevance assessments. Domain knowledge has been found to impact evaluation behaviors. Zhang et al. (2005) found that users with higher domain knowledge found a greater number of relevant documents overall. Therefore, it is hypothesized:

> ***Hypothesis 7***: Participants with higher levels of financial knowledge will bookmark a greater number of webpages than participants with lower levels of financial knowledge.

3.2.3.3. Financial knowledge - influence on mental workload. There is no literature indicating a relationship between domain knowledge and users' experiences of mental workload. Therefore, no hypothesis is proposed for a relationship between these two variables. A possible relationship between the two variables, however, will be explored.

3.2.3.4. Statistical model showing the influence of financial knowledge on the dependent variables. Figure 6 shows a statistical model of the influence of financial knowledge on search performance, relevance assessments, and mental workload. A possible relationship between financial knowledge and mental workload is represented by the arrow labeled "Exploratory Measure #2."

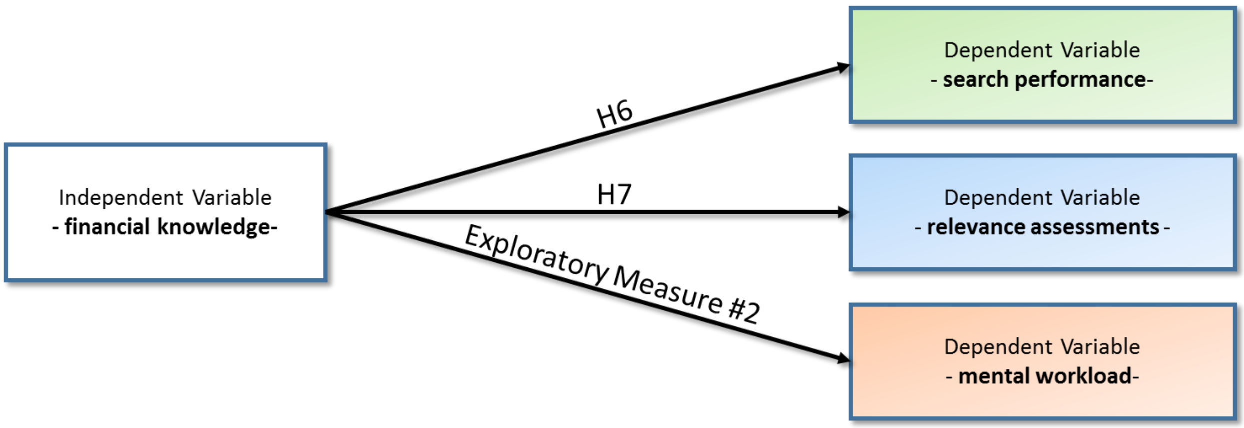


Figure 6. Statistical model showing the influence of financial knowledge on the dependent variables, with arrows labelling the first two hypotheses and a possibly third relationship labelled *Exploratory*.

3.2.4. Interaction of perceptual speed and financial knowledge on the dependent variables. There is evidence from consumer finance research that financial knowledge influences information searching in the personal finance domain (Berger & Messerschmidt, 2009) and it is my theory that financial knowledge acts as a moderating variable on the influence that cognitive abilities have on each of the three outcome variables. The relationships indicate an interaction of financial knowledge with perceptual speed that lead to the development of several hypotheses as well as additional opportunities for inquiry. Based on the research supporting development of hypotheses one, two, three, four, six, and seven, the interaction of financial knowledge and perceptual speed would be ordered as follows:

Where:
F = Financial knowledge
P = Perceptual speed
H = Higher
L = Lower

Order of effects of interaction between financial knowledge and perceptual speed:

HF, HP > HF, LP > LF, HP > LF, LP

In addition to more obvious conclusions about the order of the effects of the interaction between the two variable (e.g., people with higher financial knowledge and higher perceptual speed will outperform all other groups), the order suggests that of the two factors, financial knowledge (i.e., domain knowledge) plays a stronger role for individuals than perceptual speed (i.e., cognitive abilities) in search tasks related to personal finance topics. The following three subsections describe the statistical models and hypotheses for these relationships.

3.2.4.1. Statistical model for interaction of perceptual speed and financial knowledge on search performance. The hypotheses for the interactions regarding search performance are the following:

> ***Hypothesis 8a*****:** Participants who have both higher levels of financial knowledge and perceptual speed ability will issue more queries than any other group of participants.
>
> ***Hypothesis 8b***: Participants with higher levels of financial knowledge but lower perceptual speed ability will issue more queries per task than participants with higher perceptual speed ability but lower levels of financial knowledge.
>
> ***Hypothesis 8c***: Participants who have higher perceptual speed ability but lower levels of financial knowledge will issue more queries than participants with lower perceptual speed ability and lower levels of financial knowledge.

Figure 7 shows the statistical model for the interaction relationship with the interaction of the IVs shown in bold type. The previously described hypotheses regarding search performance and relevance assessments are included to make the model complete, but are grayed out.

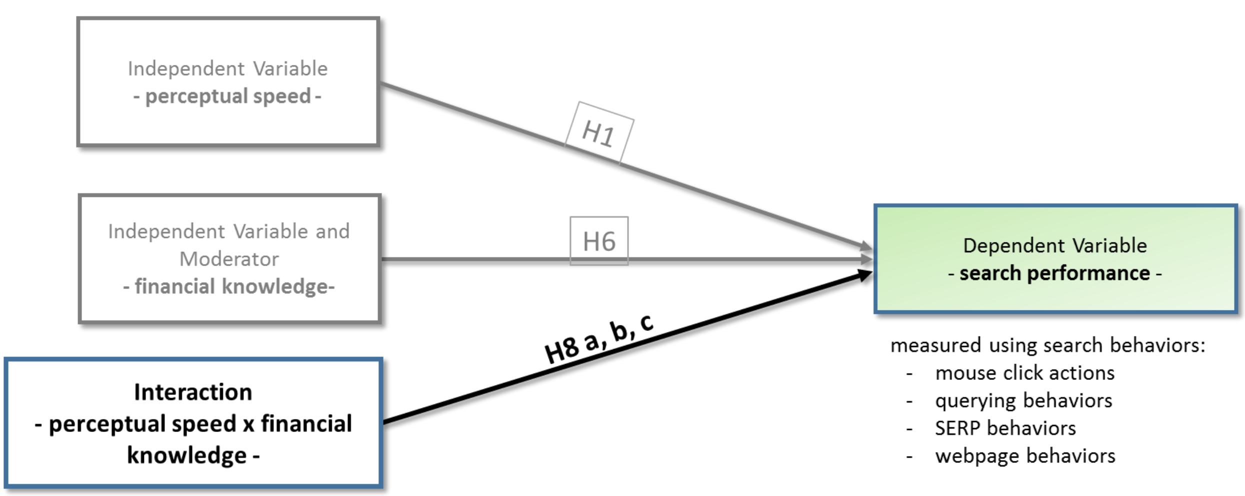


Figure 7. Statistical model showing the interaction of perceptual speed and financial knowledge (as the moderator) on the influence of search performance, with hypotheses labelled.

3.2.4.2. Statistical model for interaction of perceptual speed and financial knowledge on relevance assessments. Hypotheses for the interaction of the independent variables on relevance assessments are the following:

> ***Hypothesis 9a*****:** Participants who have both higher levels of financial knowledge and perceptual speed ability will bookmark a greater number of webpages than any other group of participants.
>
> ***Hypothesis 9b***: Participants with higher levels of financial knowledge but lower perceptual speed ability will bookmark a greater number of webpages than participants with higher perceptual speed ability but lower levels of financial knowledge.

Figure 8 shows the statistical model for the interaction relationship. The interaction of the IV's is shown in bold type. The previously described hypotheses regarding relevance assessments are included to make the diagram complete, but are shown in gray.

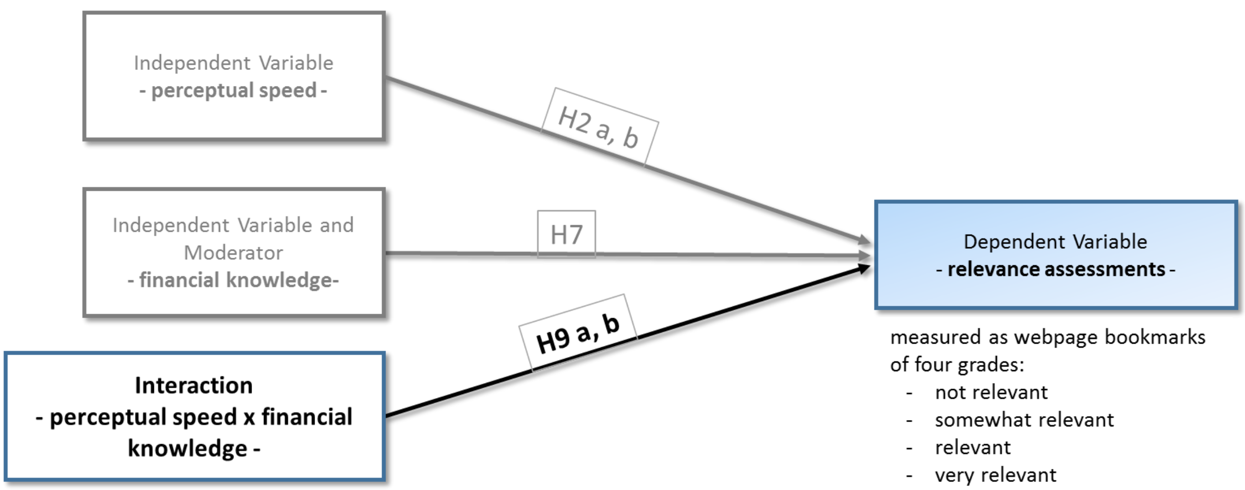


Figure 8. Statistical model showing the interaction of perceptual speed and financial knowledge (as the moderator) on the influence of relevance assessments, with hypotheses labelled.

3.2.4.3. Statistical model for interaction of perceptual speed and financial knowledge on mental workload. Since there is no existing literature that provides enough evidence to form a hypothesis about the potential interaction of perceptual speed and financial knowledge on workload, this statistical model is exploratory. Figure 9 shows the model, with the previously discussed mental workload-related hypotheses included to make the model complete but grayed out. The exploratory interaction of the IV's is shown in bold.

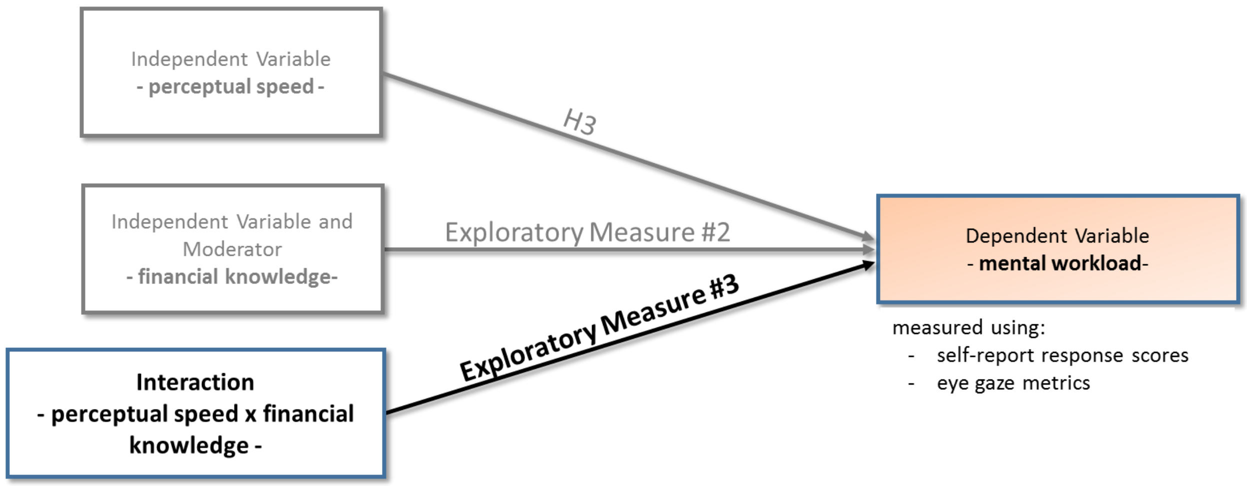


Figure 9. Statistical model showing the possible interaction of perceptual speed and financial knowledge (as the moderator) on the influence of mental workload, with the exploratory relationship labelled.

3.2.5. Interaction of working memory and financial knowledge on the dependent variables. Based on the research supporting development of hypotheses four, five, six, and seven, the interaction of financial knowledge and working memory would be ordered as follows:

Where:	Order of effects of interaction between financial knowledge and working memory:
F = Financial knowledge M = Working memory H = Higher L = Lower	HF, HM > HF, LM > LF, HM > LF, LM

Similar to the perceptual speed model, this model suggests that the effect of financial knowledge plays a more important role in certain circumstances. However, there is no existing research to support the development of hypotheses for this set of relationships, so the data will be explored to uncover possible relationships.

3.2.5.1. Statistical model for interaction of working memory and financial knowledge on search performance. There are no hypotheses for this relationship. Therefore, an exploratory model is shown in Figure 10 in bold type. The previously discussed working memory hypotheses are included to make the model complete, but are grayed out.

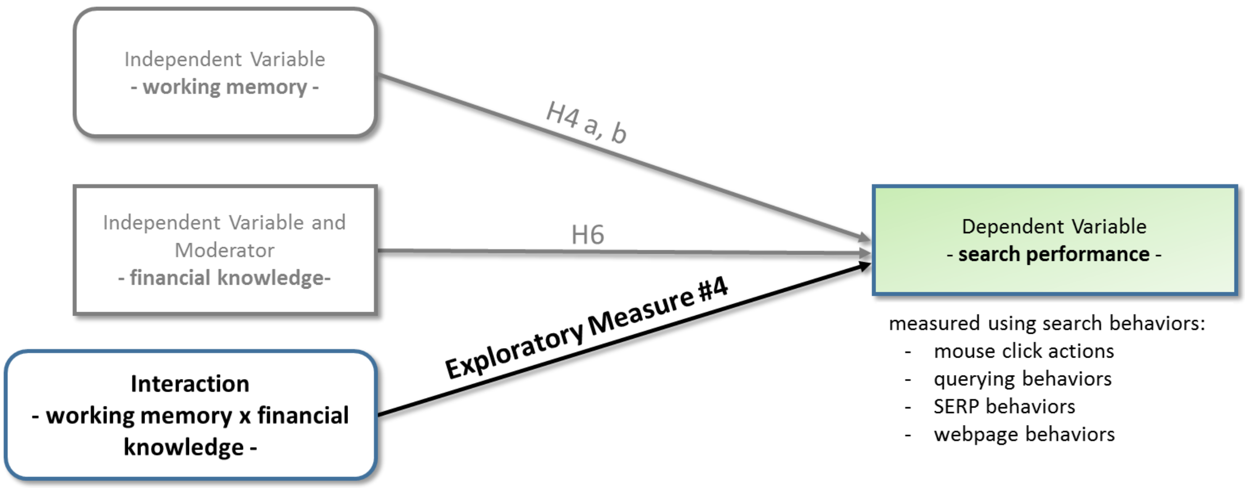


Figure 10. Statistical model showing the possible interaction of working memory with financial knowledge (as the moderator) on the influence of search performance, labelled *Exploratory.*

3.2.5.2. Statistical model for interaction of working memory and financial knowledge on relevance assessments. There are no hypotheses for this relationship. Therefore, an exploratory model is shown in Figure 11 in bold type. The previously discussed working memory hypotheses are included to make the model complete, but are grayed out.

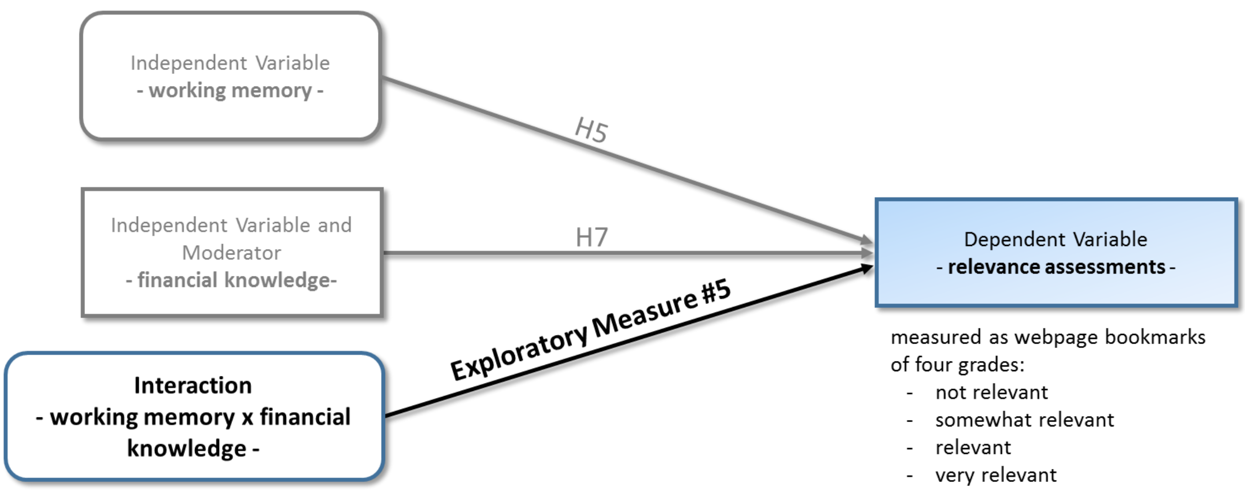


Figure 11. Statistical model showing the possible interaction of working memory with financial knowledge (as the moderator) on the influence of relevance assessments, labelled *Exploratory.*

3.2.5.3. Statistical model for interaction of working memory and financial knowledge on mental workload. There are no hypotheses for this relationship. Therefore, an exploratory model is shown in Figure 12 in bold type. The previously discussed working memory hypotheses are included to make the model complete, but are grayed out.

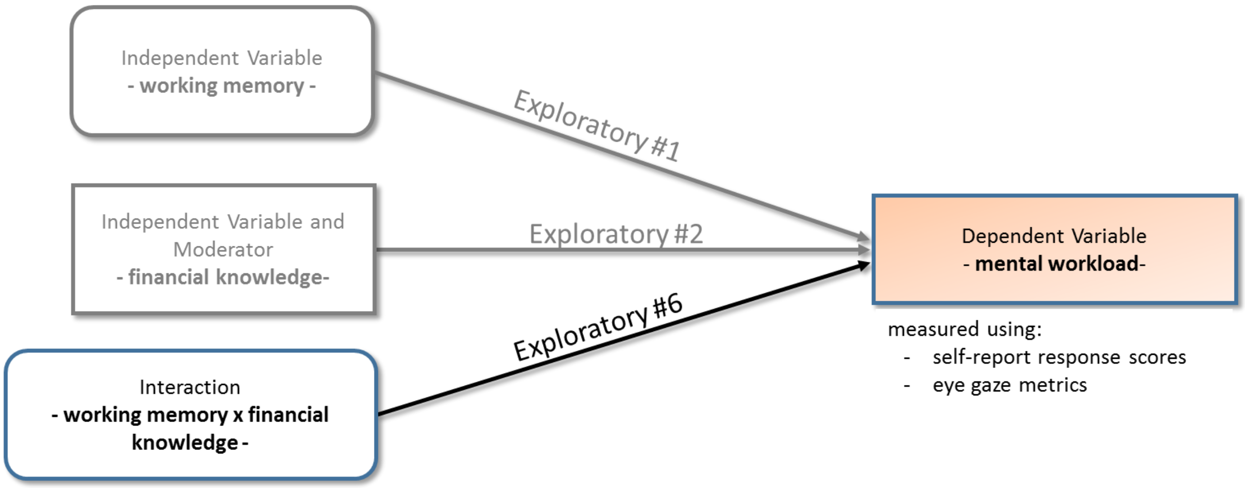


Figure 12. Statistical model showing the possible interaction of working memory with financial knowledge (as the moderator) on the influence of mental workload, labelled *Exploratory.*

3.3. Research Question #2 – Strategies for Finding and Evaluating Information

The second research questions asks, *"What are users' strategies for finding and evaluating information on the Internet about different kinds of financial loans? How do users' cognitive abilities and financial knowledge influence these strategies?"* In this dissertation, *finding information* is synonymous with information searching. *Evaluating information* refers to the relevance judgments of users and is compared with the relevance judgments of expert assessors. The examination of participants' search strategies in the dissertation study will be undertaken as exploratory investigation and therefore, there are no hypotheses. Further explanation of the approach for studying participants' search strategies is provided in the next section, *Research Methods.*

CHAPTER 4: METHODS

This chapter begins with a detailed explanation of the dissertation study design, which includes definitions for the constructs that were measured and descriptions of instruments used for measuring them. This is followed by an explanation of the tasks that were developed for the study. After this, the study procedure steps are outlined in sequence. Finally, the chapter ends with a description of the recruitment of the participants and a description of their characteristics.

4.1. Study Design

To pursue the objective of the dissertation study, a quasi-experimental design was used (Littlejohn, 1996) in a laboratory-based IIR study. The study consisted of measuring three independent variables and three dependent variables using measurement instruments and data collection techniques chosen from an extensive review of the literature and also from first-hand use. In this section each variable, its operationalization, and data collection technique is described in detail.

4.1.1. Cognitive abilities. Cognitive ability is defined as *a person's inherent and acquired intellectual capacity to comprehend the requirements of a cognitive task within its context and successfully achieve the task's desired outcome*. The theoretical basis for the measurement of cognitive abilities in this study is Three-Stratum Theory (Carroll, 1993).

4.1.1.1. Perceptual speed measurement. The cognitive ability, *perceptual speed,* is defined as "the ability to find a given configuration borne in a person's mind during a search process, amidst a mass of distracting material" (Carroll, 1993, p. 308). For the dissertation study, perceptual speed was operationalized using the P-1 *Finding A's* psychometric test taken from the Ekstrom Kit of Factor-Referenced Tests (Ekstrom, French, Harman, & Dermen, 1976a). In this test, participants are shown lists of words and asked to cross out every word that contains the letter "a" as quickly and accurately as possible. It is in paper-and-pencil format and has two parts that are each two minutes long. The final score is the total number of words correctly crossed out after both 2-minute parts. Tests were hand-scored after the close of the experiment, using the Ekstrom Kit manual instructions, and the results were stored in a spreadsheet for further analysis.

4.1.1.2. Memory span measurement. The working memory construct was operationalized in the study using the cognitive ability *memory span*, which is defined as "the ability to recall a number of distinct elements for immediate reproduction and involves storage and retrieval of information in short-term memory" (Ekstrom et al., 1976b). Memory span was measured in the study using the CogLab 2.0 *Memory Span* psychometric test (Francis, Neath, & VanHorn, 2008). In the test, participants were presented timed trials in which a list of items is presented one at a time in random order and then the participant is asked to recall the items in the same order after they are all presented. If the participant recalls the items correctly, the list increases by a single item and if the participant does not recall them correctly, a single item is removed from the next list. Lists may consist of all digits, all letters that sounds dissimilar, all letters that sound similar, short words, and long words. The maximum memory span measurable with this test is 10. The test was administered using the CogLab 2.0 CD and the test data was captured in HTML files and downloaded into statistical software for further analysis. The test was scored by adding up the number of items correctly recalled and averaging them across the five components.

4.1.3. Financial knowledge measurement. *Financial knowledge* is comprised of knowledge of financial products, understanding of financial concepts, skills in math, and skills in numeracy (Hastings, Madrian, & Skimmyhorn, 2013). It is often used as an equivalent for the phrase, *financial literacy,* which has been defined as "the ability to use knowledge and skills to manage one's financial resources effectively for lifetime financial security" (Jump$tart, 1997). For the purpose of this dissertation, the phrase *financial knowledge* is used to mean the construct being measured and the phrase *financial literacy* is used to describe the actual test instrument that was used as a measure of the construct.

Three measures were used to gain an understanding of participants' financial knowledge. The three measures were a two-question self-assessment measure, a 14-question survey about participants' experience with basic financial products, and an 8-question financial literacy test. The first two measures were taken during the first session of the lab study and the literacy test was administered during the second session. By using a combination of three measures – an objective literacy test, a self-assessment measure, and a product experience measure – it was expected that a more robust view of each participant's overall financial knowledge could be gathered. The two self-report questions (Appendix B) and the financial product survey questions (Appendix C) were administered as part of the entry questionnaire, which also contained the demographic questions (Appendix A). Asking the financial self-report and financial product survey questions before the start of the search tasks was designed to eliminate the possibility that participants might learn new information during the search tasks that would influence their answers to these questions. The financial literacy test was administered during the second session, which was held within several days of the first session. Responses for all financial items were captured using Qualtrics survey software and downloaded into a spreadsheet for further analysis.

Questions for the three measures were obtained from a combination of the 2015 National Financial Capability Survey (NFCS) (Lin et al., 2016) and a national survey of debt literacy reported in Lusardi and Tufano (2015). The NFCS is a national study of the financial capability of American adults that is part of a large multi-year project funded by the Financial Industry Regulatory Authority (FINRA) Investor Education Foundation, in consultation with the U.S. Department of the Treasury and the President's Advisory Councils on Financial Literacy and Financial Capability. National and state-by-state findings are available from the online surveys of over 25,000 adults, conducted in 2009, 2012, and 2015. There were several benefits to using the financial literacy questions from these national U.S. surveys. It eliminated the need to create a financial knowledge instrument for the dissertation study and also brought with it the authority of instruments created for national projects that had been vetted by multiple financial professionals.

4.1.4. Search behavior measures. Participants' search behaviors were measured using behavioral signals of their activities related to mouse clicks, queries, SERPs and webpages. The measures were based on findings from the literature review and included interactions from Al-Maskari and Sanderson (2011); Hassan, White, Dumais, and Wang (2014); Jiang et al. (2014); and White and Morris (2007). Data was captured with the logging feature of the Tobii Studio. Table 2 shows a list of the measures and definitions for each.

Table 2. *Search Behavior Measures and Definitions*

Type	Measure	Definition
Clicks	Number of clicks	Total number of clicks per task
	Number SERP clicks	Number of clicks on SERPs
	Time to first click	Time from the beginning of the task to the first mouse click
Query	Queries	Number of total queries entered into the search box during the search task
	Unique queries	Number of unique queries entered
	Query length	The number of distinct terms in each query submitted, not including stopwords
SERPs	SERPs	Number of unique SERPs displayed during the search task
	SERP display time	Total time spent on SERPs
Webpages	Webpages	Number of total non-SERP webpages displayed during search task
	Unique webpages	Number of unique non-SERP webpages displayed during search task
	Webpage display time	Total time spent on non-SERP webpages

4.1.5. Relevance assessments. Evaluation behaviors were measured for webpages as bookmarks created by participants (Table 3) according to four possible grades: not relevant, somewhat relevant, relevant, and very relevant (the complete instructions to participants for the study, including how to use these grades is shown in Appendix J). Participants' assessments of relevance in the study were measured as the number of webpages saved as the different levels of bookmarks. Webpage links and their relevance grades were captured in a spreadsheet at the end of each participant's session.

At the end of the data collection, all of the webpage links were loaded into a single MS Excel workbook, with tabs sorted by task topic. Each webpage link was given a unique identification number and then participants' judgments were deleted so that the file could be used by the two expert assessors for their own relevance grades.

Two expert assessors were chosen based on their extensive backgrounds in finance. Expert #1 is a retired financial services professional who worked as a senior executive at a major mortgage finance company for more than twenty years and Expert #2 is a former certified financial planner who managed financial portfolios for private investors during her career. The two expert assessors judged all of the bookmarks made by participants, grading each using the four levels described above (the complete instructions to assessors is shown in Appendix L). The length of time from the end of the data collection to when assessors returned the spreadsheets ranged from two to five months, raising a slight concern that this time difference might result in changed content on some of the webpages. To address this concern about the time difference, a manual cross-check of the original screen capture videos and the live webpages was conducted after assessors' returned the graded pages. Any webpage that had changed materially (i.e., other than ad or layout changes) was returned to the assessor with a screen capture of the original page and a re-grade was requested. There were less than a dozen such instances.

To compare participants' judgments with assessors, *interactive user precision* was calculated as follows:

Measure 1 = raw (this is the raw number of webpages bookmarked by participants)

Measure 2 = rel (this is the number of webpages assessed as relevant by expert assessors)

Precision = (raw ∩ rel) / raw

Using this formula, it was possible to calculate precision for different levels of financial knowledge and cognitive abilities. These results are shown in the results section.

Table 3. *Evaluation Behavior Measures*

Type	Measure	Definition
Participant	Pages bookmarked	Total number of webpages bookmarked per task
	Very relevant	Webpages bookmarked *very relevant*
	Relevant	Webpages bookmarked *relevant*
	Somewhat relevant	Webpages bookmarked *somewhat relevant*
	Not relevant	Webpages bookmarked *not relevant*
Expert Assessor	Very relevant	Number of participants' webpages bookmarked by expert as *very relevant*
	Relevant	Number of participants' webpages bookmarked by expert as *relevant*
	Somewhat relevant	Number of participants' webpages bookmarked by expert as *somewhat relevant*
	Not relevant	Number of participants' webpages bookmarked by expert as *not relevant*

4.1.6. Mental workload measures. Mental workload was captured using a self-report questionnaire and eye gaze measures. The self-report questionnaire used was the NASA-TLX (Hart & Staveland, 1988). The NASA-Task Load Index (TLX) is a multi-dimensional subjective rating questionnaire that measures the mental, physical, and affective demands imposed on individuals by work tasks. Since its development in the late 1980s for use in aeronautics testing at NASA, it has been administered extensively inside and outside the aeronautics field. The TLX has been described as one of the most widely used mental workload measurement scales (Megaw, 2005). The TLX has six individual scale questions, shown in Appendix G[5]. The scale questions were asked along with three post-task questions at the end of each search task in Qualtrics survey software. The scores were averaged for the final mental workload measure.

[5] In addition to the scoring of the six scales, the test also asks participants to conduct 15 paired comparisons of the six scale items, however in the interest of time, this section of the measure was not administered. In addition, given that additional measures of mental workload were captured through eye tracking, elimination of the paired comparison part of the TLX seemed reasonable.

4.1.7. Eye gaze measures. Numerous studies have found that as mental processing becomes more challenging, the length of fixation durations increase (Holmqvist & Nystrom, 2011) and that the use of the fixation duration measure is preferable to pupil size measures because fixation durations are sensitive to both acute and continuous processing memory load while pupil size only changes during increased processing load (Meghanathan et al., 2015). Table 4 shows the eye gaze measures that were captured for the dissertation study.

Table 4. *Eye Gaze Measures*

Measure	Definition
Fixation count	Total number of fixations per task
Fixations per query	Number of fixations per query
Fixations per SERP	Number of fixations per SERP
Fixation duration	Length (ms) of a fixation

A Tobii X2-60 eye tracking system was used for the study with Tobii Studio software version 3.4.8. The eye tracker is an infrared camera that captures the movement of the fovea, which is the center of the pupil. The sampling frequency for the X2-60 camera is 60 Hz which means it captures 60 data points per second per eye, or about one data point every 16.7 milliseconds. As each data point is captured, it is assigned with a timestamp and the X,Y coordinates of its location on the monitor. This data is processed by the Tobii software into eye fixations and then overlaid onto a video recording of the screen so that it can be used for visualizing the data and calculating eye tracking metrics. The way in which data is processed into fixation counts, duration thresholds, and other measures is via the fixation filter used in the software. There are several types of fixation filters to choose from and the one I used was *I-VT Fixation Filter.* The main reason for using this fixation filter was because the default values have been set by the manufacturer "to provide accurate fixation classifications for the most common eye tracking use cases, e.g., web, market research, and standard reading studies” (Tobii, 2016, p. 54). One of the features of this filter includes that it classifies eye movements based on the velocity of directional shifts of the eye. It also features a gap fill-in interpolation function for filling in missing eye value samples that occur from short time periods such as a temporary reflection on the participant’s eye or eyeglasses and are typically less than 50 milliseconds long. Other features are included to reduce data noise, merge incorrectly split fixations, and discard short fixations. The minimum fixation duration for this filter is set to 60 milliseconds.

4.1.8. Search strategy measures. Stimulated recall data were captured using Morae screen recording software for the last task that each participant completed (the moderator’s guide is shown in Appendix K). Fifteen participants conducted their last search task on payday loans, 15 on the reverse mortgage task, and 14 on student loans. Stimulated recall was conducted only on the last search task to reduce memory burden on participants for thinking back through the task and also to keep the first session within a 75-minute timeframe. The instructions that were presented to participants were read aloud to introduce this portion of the study and were as follows:

> *During this next section of the study, I’m going to play back a screen recording of the actions you took during the last task you completed. While you watch this recording, I would like for you to state aloud why you took the actions shown on the screen and what you were thinking when you took those actions. I would like you to walk me through the decision-making processes you underwent as you searched. There are no right or wrong answers here. I am simply looking for your thoughts as you review the steps you took during the experiment. Even minor thoughts will be helpful to this study.*

The audio recordings from the stimulated recall were transcribed professionally and qualitatively coded by the researcher. Several steps were taken to ensure credibility of the coding process, including periodic discussions with my doctoral advisor and a peer de-briefing once the codebook was finalized. Guidance on how to conduct the peer de-brief was obtained from Schwandt, Lincoln, and Guba (2007) and Costello (2015).

4.2. Search Tasks

Designing tasks for studies of search and evaluation behaviors is challenging because there are many different factors to consider. This section explains the dissertation search task design process. It starts with a brief explanation of the main influences and conceptual grounding from the literature, then moves on to the goals for the task design, and then describes and justifies the selection of the task topics. The resulting three tasks that were designed and used in the study are shown in Appendix E.

4.2.1. Conceptual grounding. In the literature, search tasks are a subset of the broader concept of work tasks. A work task is a set of activities with a distinguishable beginning and end that are conducted with a goal and purpose in mind. In the context of information-intensive settings, work tasks are made up of subtasks, of which information search tasks are one kind (Byström & Hansen, 2005). In an information search task (or simply, "search task"), a person performs a sequence of activities toward the goal of finding some general or specified range of information (Ingwersen & Järvelin, 2005). Within the context of laboratory-based studies, search tasks are generated in three main ways. The tasks may originate from the information needs of study participants themselves and in this case are called *self-generated* or *natural tasks.* They may also be generated by the researchers in the form of *simulated situations* (Borlund, 2000) which are designed with careful consideration of the participants' background. Finally, search tasks may *assigned tasks* that are generally designed with the intent of evoking different kinds of search behaviors in the participants. In the dissertation study, search tasks were assigned to participants.

There are many other different dimensions and aspects of search tasks that have been investigated in the literature and a complete explanation of those many dimensions is outside the scope of this thesis. However, it may be useful for the reader to know that several works had a strong influence on the development of the tasks for this study including those already cited as well as Vakkari (2003), Wildemuth and Freund (2009), the NSF Task-Based Information Search System Workshop materials at: https://ils.unc.edu/taskbasedsearch/, Toms et al. (2008), and Marchionini (2006).

4.2.2. Design goals. The tasks were designed using ideas and concepts primarily drawn from the works of Borlund (2000) and Wildemuth and Freund (2012). It is important to clarify that the tasks in this dissertation study are not exact designs of Borlund's *simulated work task situations,* because they do not include the use of natural tasks based on participants' own information needs that are used to form a baseline against the simulated information needs (Borlund, 2016). As well, the tasks encompass many of the attributes of *exploratory search tasks* as indicated in Wildemuth and Freund (2012), but one attribute they do not have is that they were not designed to explore the search processes of users over a long period of time. The following is a list of the guidelines used in developing the tasks:

- create ***realistic*** situations that participants could imagine themselves encountering in their everyday lives;
- create ***open-ended*** tasks that could not be answered by a single view of a Google SERP entity card or results snippet;
- create tasks that encourage ***exploration*** such as described by Marchionini (2006), in which people engage in search activities related to investigating (e.g., analyzing information, including or excluding information, synthesizing, evaluating, etc.)

and learning (e.g., acquiring knowledge, comprehending and interpreting information, comparing, etc.).

- create tasks with high-enough ***complexity*** that it would be unlikely that a participant could fully address the information need specified by the task based on information from a single webpage. The idea behind this is to create tasks that would require participants to hold multiple ideas in their minds while searching and integrate old information with new information while they were searching and evaluating, which would increase the cognitive load of the participant (Rouet, 2003).
- create tasks that are ***interesting*** enough that people will be motivated to take them seriously; and
- finally, create an easy to follow task ***structure*** for participants. Similar to the simulated information needs situations of Borlund 2000, task situations were created which gave participants information about the source of the information need, the environment of the situation, and the problem which has to be solved, in order to help the participant understand the objective of the search.

4.2.3. Selection of task topics. The task topics were chosen based on a review of government and non-profit financial education resources (shown in Appendix I) as well as Household Debt and Credit Data reports published by the Federal Reserve Bank of New York[6].

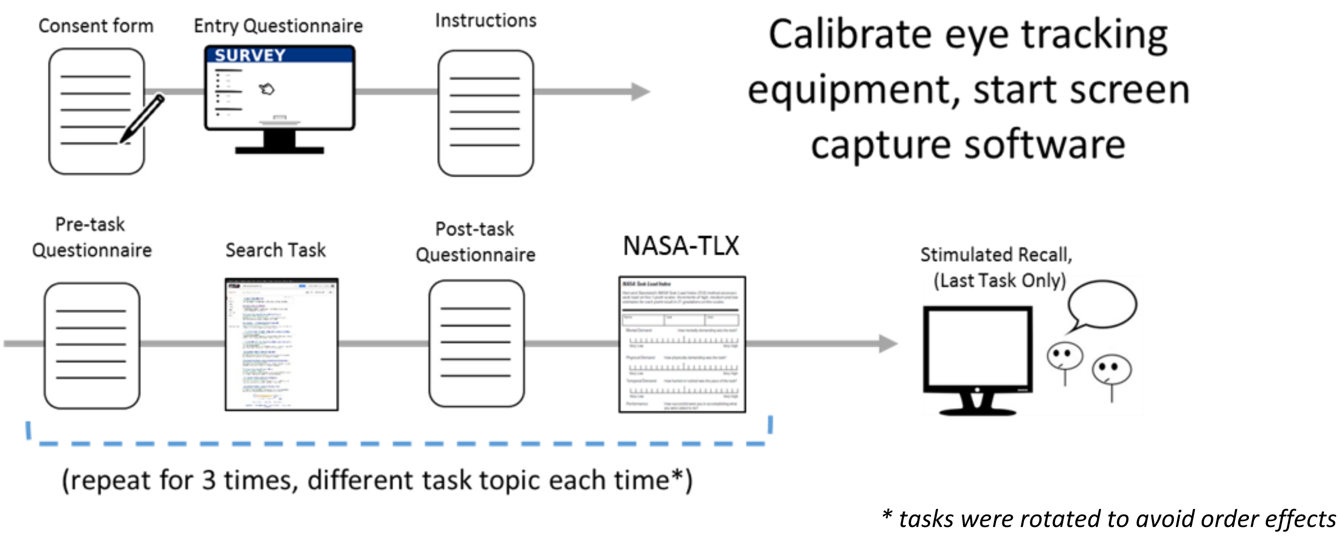
4.3. Procedure

Prior to launching the study, a pilot test was conducted using four doctoral student peers and two non-student friends. After this an application was filed with UNC's Institutional

[6] https://www.newyorkfed.org/microeconomics/hhdc

Review Board explaining the details of the dissertation study. The application was approved as IRB #17-0077 on March 9, 2017. The experiment was conducted in two sessions on two different days, in Room 09 of the Interactive Information Systems Lab in Manning Hall at UNC-Chapel Hill. The procedure is illustrated in Figure 13. Prior to the lab study, all aspects of the procedure were documented in a procedures binder by the author to ensure consistency across all participants when carrying out the experiment. Each person participated in two sessions and participants were studied one at a time.

SESSION 1 (~60 minutes).



SESSION 2 (~25 minutes).

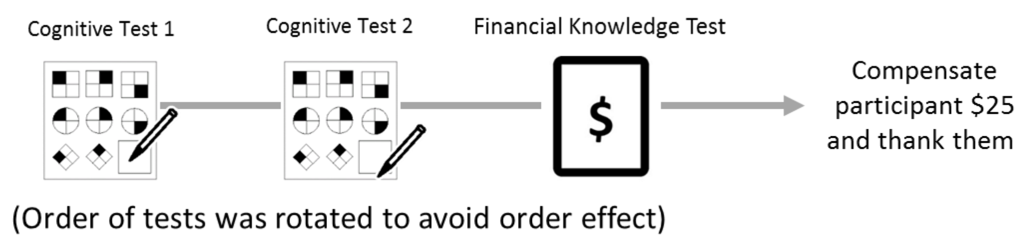


Figure 13: Illustration of experiment procedure for the dissertation study, shown in two sessions.

The first session of the lab study was the search and evaluation session and lasted about 75 minutes. The search session elements were set up as media elements in the Tobii eye tracking system so that participants could easily transition through the items in the study with as little disruption as possible from the researcher and also so that participants' eye gaze behaviors and search actions could be captured by the Tobii eye tracker and the screen recording using TechSmith's Morae software. As each lab study began, the participant was greeted by the researcher and asked to sit at the computer terminal so he or she could read the online participant consent form before proceeding further. Next, the participant was asked to fill out the entry questionnaire containing basic demographic questions (Appendix A - 5 items), the financial knowledge self-rating questions (Appendix B - 3 items), and the financial product experience questions (Appendix C - 14 items). The participant then read through on-screen instructions about the information search tasks of the study, which provided directions on how to save bookmarks for very relevant, relevant, somewhat relevant, and not relevant websites. After this, the participant was calibrated with the Tobii X2-60 eye tracking system. Next, the participant filled out a pre-task questionnaire (Appendix D) which also contained the description of the first search task. This pre-task questionnaire was previously used in Azzopardi et al. (2013), Edwards et al. (2015), Edwards (2015), and Kelly and Azzopardi (2015). After completing the pre-task questionnaire, the participant clicked "done" and was then taken to the start page for the search session, showing the Google search box (participants were told in the instructions that it was okay to use any search engine for searching; starting at Google was simply a convenience). On the lower right hand corner of the screen the search task was shown on a small screen note at all times. During the search task, the participants searched for relevant web pages and saved web pages as bookmarks into any of the four folders (not relevant, somewhat relevant, relevant, and highly relevant) the participant deemed appropriate for the web page and the task.

During the time of the above-described search session, the researcher sat out of sight of the participant, observing the participant's live actions on a second monitor, to ensure that the eye tracker was capturing eye movements and to observe the screen recording video. If the participant began to move his or her head out of the view of the eye tracker, the researcher gently reminded the participant to maintain a stable head position in front of the computer monitor.

The participant filled out the on-screen post-task (Appendix F) and mental workload questionnaire (Appendix G) once the search session was finished. Then the participant clicked "done" and was taken to the beginning of the next search task. This process of filling out the pre-task questionnaire, conducting the search session, and filling out the post-task questionnaire was repeated three times. Task topics were rotated to avoid order effects.

Upon completion of the third search session and post-task questionnaire, the "done" button took participants to a page indicating the search session was complete. The researcher then turned off the eye tracker and re-wound the last search task on the Morae video recorder to begin the stimulated recall session. Stimulated recall was conducted only for the last of the three tasks, to ensure that the information was fresh in the participant's mind. Conducting stimulated recall in this manner is supported by guidance from Ericsson and Simon (1996) who indicate that this approach helps keep the focus on the tasks themselves as the priority of the study. The stimulated recall instructions were described in Section 4.1.8. The procedure and questions are shown in Appendix K.

In the second session, which lasted about 20 minutes, participants took the two cognitive tests and the financial literacy test. Most of the second sessions took place within one to four days after the first session. Test order was rotated.

4.4. Participants

Forty-eight participants were recruited using the staff email list at the University of North Carolina (UNC) at Chapel Hill, via recruitment email (Appendix H). This sample size was based on several considerations. An initial power analysis was conducted using G*Power, a statistical software program freely available on the Internet. Effect size was set to 0.4 and power at 0.8, with 1 degree of freedom, 2 groups (high and low cognitive ability), and 1 covariate (financial knowledge). These settings were based on several considerations. First, there were large effect sizes in Brennan et al. (2014) with only a sample of N=21. Second, given the nature of this research as partially exploratory, there is a greater tolerance for Type II errors. In addition, power of 0.8 is a commonly accepted level of power. Based on these parameters, the estimated sample size needed was N = 52 for a single session study. However, for practical reasons the sample size was reduced to N = 40. An additional 20% of participants was recruited based on recommendations from the eye tracking literature and on recommendations from colleagues.

Recruitment within the general adult population was chosen because this population has been found to have sufficient distribution of cognitive ability scores in previous work (Brennan et al., 2014). Using a convenience sample of employed workers was chosen because it was expected that participants would have at least the basic level of personal finance experience necessary for this search study, as the result of having generated earned income on a regular basis. In addition, recruitment was for native-English speakers, as it has been found in retrospective think aloud that native English speakers are more suitable as participants to prevent interference by linguistic competence issues of non-native English speaking participants (Maidel, 2014). This requirement also reduced the chances that participants would be unfamiliar with the U.S. economic and banking environments for consumers. In addition, by recruiting employees working at UNC and using the SILS Interactive Information Systems Lab for the study, issues related to parking for participants or having to travel with expensive equipment were eliminated.

One participant was released early in the study because she was unable to follow the instructions during the first session and then because the eye tracker could not calibrate her eyes in her return session, making the overall sample size N = 47. Demographics, self-rating of financial knowledge, and financial product experience are reported for this sample. The sample of the quantitative results was N = 42 and the sample for the qualitative results was N = 44. Details about participant exclusions from those analyses are explained in their respective sections in Chapter 5: Results.

4.4.1. Demographics. Thirty-four participants were female and 13 were male. While a third category for identifying gender was available ("gender non-conforming"), no participants selected this option. The average age was 32.5 years (SD=12.97), with ages ranging from 18 to 62. The median age was 29. Twenty-nine participants identified as White or Caucasian, ten identified as Black or African-American, three identified as Asian, two identified as Hispanic or Latino, two as Other-Middle Eastern, and one as American Indian. In terms of the highest level of education completed, 16 participants said they held post-graduate degrees, 15 had Bachelor's degrees, ten had some college but no degree, three had Associate's degrees, and three were high school graduates. Twenty-eight participants were employed full-time, ten were employed part-time, seven were full-time students, and two were self-employed. A list of participants' occupations is shown in Table 5. Numbers in parentheses indicate how many participants reported that occupation title. Some of the titles have been generalized to protect the privacy of participants with unique titles that might be easily identified within UNC staff.

Table 5. *List of Study Participants' Occupations*

Occupation Title	**Occupation Title**
Administrative Support (4)	Librarian (5)
Admissions Counselor	Librarian Assistant
Asset Management Technician	Operations Support Technician
Associate Dean	Business Owner
Business Services Coordinator	Pharmacy Technician (2)
Clinical Instructor	Program Coordinator
Clinical Research Coordinator	Research Assistant (4)
Community Outreach Staff	Research Instructor
Conference Coordinator	Resident Advisor
Counselor	Scheduling Assistant
Director	Sports Program Supervisor
Engineer	Staff member
Facilities and Operations Staff	Student Affairs/Higher Education
Full-time Student	Teacher Assistant
Information Manager	Training Coordinator
Laboratory Assistant (2)	No job title given (4)

4.4.2. Self-rating of financial knowledge. Participants were asked to complete a multi-item entry questionnaire at the beginning of the first session which included two questions asking them to rate their levels of financial knowledge (Appendix B). As previously explained in section 4.1.3, these questions were taken from the 2015 NFCS study of American adults. Responses to the questions are shown in Table 6.

Table 6. *Means and Standard Deviations for Financial Self-Rating Scores*

Question	*n*	*M (SD)*
1. On a scale from 1 to 7, where 1 means very low and 7 means very high, how would you assess your overall financial knowledge?	47	4.21 (.175)
2. How strongly do you agree or disagree with the following statements? Please give your answer on a scale from 1 to 7, where 1 = "Strongly Disagree," 7 = "Strongly Agree," and 4 = "Neither Agree Nor Disagree." You can use any number from 1 to 7.		
a. I am good at dealing with day-to-day financial matters.	47	5.40 (.208)
b. I am pretty good at math.	47	4.87 (.210)

4.4.3. Financial product experience. Table 7 shows responses to the financial product experience questions. Participants' responses to these questions provide several opportunities for characterizing the overall financial experience of the group. First, the sample of participants had strong foundational financial experience in that most participants followed a budget (77%), all had a checking account (100%), and all but one had a savings account (98%). In addition, the participants in the sample can be considered somewhat financially conservative based on the fact that more than 75% of the sample held three or fewer credit cards. In fact, a full 23% (N = 11) had no credit cards at all. The sample had more limited experience when it came to advanced financial products and this is indicated by the fact that less than half of the participants held non-retirement account investments and less than half had ever owned a home. Finally, most participants in the sample did not view themselves as being the most knowledgeable person in their households when it came to money matters such as savings, investing, and debt. Appendix C shows responses compared to U.S. and N.C. averages from the NFCS.

Table 7. *Frequencies for Financial Product Experience Questions*

Question	Answer	Number	Percent
Who in your household is most knowledgeable about saving, investing, and debt?	You	19	40%
	Someone else	22	47%
	You and someone else, equally	6	13%
	Don't know	0	0%
Does your household have a budget? A household budget is used to decide what share of your household income will be used for spending, saving, or paying bills.	Yes	36	77%
	No	9	19%
	Don't know	2	4%
Do you have a checking account?	Yes	47	100%
	No	0	0%
	Don't know	0	0%
Do you have a savings account, money market account, or CDs?	Yes	46	98%
	No	1	2%
	Don't know	0	0%
Do you have any retirement plans through a current or previous employer, like a pension or 401(k)?	Yes	26	55%
	No	19	40%
	Don't know	1	2%
	Prefer not to answer	1	2%
Not including retirement accounts, do you have any investments in stocks, bonds, mutual funds, or other securities?	Yes	14	30%
	No	32	68%
	Don't know	1	2%
	Prefer not to answer	0	0%
Have you ever owned a home?	Yes	20	43%
	No	27	57%
	Don't know	0	0%
	Prefer not to answer	0	0%
Do you currently own your home?	Yes	18	38%
	No	28	60%
	Don't know	1	2%
	Prefer not to answer	0	0%
Do you currently have any mortgages on your home?	Yes	20	43%
	No	26	55%
	Don't know	0	0%
	Prefer not to answer	1	2%
Do you have any home equity loans?	Yes	6	13%
	No	36	77%
	Don't know	5	11%
	Prefer not to answer	0	0%

Question	Answer	Number	Percent
How many credit cards do you have?	1	11	23%
	2-3	15	32%
	4-8	9	19%
	9-12	0	0%
	13 or more	1	2%
	No credit cards	11	23%
	Don't know	0	0%
	Prefer not to answer	0	0%
Do you currently have an auto loan?	Yes	16	34%
	No	31	66%
	Don't know	0	0%
	Prefer not to answer	0	0%
Do you currently have any student loans? If so, for whose education was this/were these loan(s) taken out? (Select all that apply)*	Yourself	20	43%
	Spouse/ partner	4	9%
	Your child(ren)	0	0%
*Responses add up to >100%	No, do not have	23	49%
	Don't know	0	0%
	Prefer not to answer	1	2%

CHAPTER 5: RESULTS

The results section is organized as follows: first, a data overview is given along with an explanation of important assumptions underlying the data analysis. Next, descriptive statistics are summarized for the pre- and post-task questionnaires, psychometric and knowledge tests, search performance measures, relevance assessments, mental workload questionnaire, and eye movement data. Following that, inferential statistics are presented that examine the nature of the relationships between the independent and dependent variables and address the hypotheses for research question #1. The final part of the results section addresses research question #2 with the qualitative analysis of the stimulated recall interviews. Participants' strategies for finding and evaluating information are explained there.

5.1. Data Overview and Analytic Assumptions

Several aspects of the data are important to describe. This data overview describes the data exclusion of several participants from the study, data reconciliation issues that arose in the search interaction log data, the use of dichotomized variables in the analysis, and the underlying distributions of the data. Decisions about how to handle the data resulted in numerous assumptions and those are also discussed. Beyond the data-management-driven assumptions, there were other assumptions made as well that played a role in the data analysis and resulting findings – those too are discussed in this section.

5.1.1. Sample data excluded. A total of 48 participants were recruited for the study. During the study, one participant (P41) was disqualified for not following the instructions for the study and also because I was unable to calibrate her eyes with the Tobii X2-60 eye tracker on her second attempt at completing the study. She was compensated with $20 for her time and the small amount of data from her participation was deleted. Other participants' data were removed from the quantitative of the datasets as follows: P13 and P31 were removed because the logging software did not function properly during their search sessions. P38 was removed because logging files for the 2nd and 3rd search tasks were saved over accidentally. P44 and P46 were removed because the data for their Memory Span tests was accidentally deleted during a backup procedure. This left a final sample for the quantitative portion of N = 42. For the qualitative portion of the data, P44, P46, P13, and P31 were included but the data from P05 and P06 were excluded because the voice recording device was not turned on before their stimulated recall interviews so the interviews were never recorded. This left a final sample for the qualitative portion of N = 44.

5.1.2. Search logging issues. As mentioned previously, search data was logged using the Tobii X2-60 eye tracker software and the bookmark data was captured as HTML webpage links using the Favorites function of Internet Explorer and then stored in a Microsoft Excel spreadsheet. The webpage URLs from the eye tracking logs should have matched the number of bookmarks that participants saved because participants were instructed to bookmark every content page they viewed. However, the discrepancy between the eye tracking logs and the bookmark logs occurred for all participants and at a wide margin of differences. Further investigation revealed that the webpage URL data from the eye tracker included previous webpages that participants passed over as they were using the *Back* button on the browser, which lead to a large number of webpages being counted that were simply repeated webpages.

Therefore, the measure of *Unique URLs* has been used in hypotheses and data analysis related to what was originally a measure of all webpages.

This issue also occurred for the data that were collected for the *Number of SERPs* visited. SERPs were counted more than once because of the use of the *Back* button in navigation. A possible proxy to use for the Number of SERPs visited and viewed is the *Number of Unique Queries.*

5.1.3. Probability level for hypothesis testing. The probability level, or alpha (α), can be set at various levels, depending on the level of confidence researchers wish to have in their decisions to reject the null hypothesis. The convention is to set $\alpha = .05$. In certain cases, though, it is preferable to set α to a higher ($\alpha = .01$) or lower level ($\alpha = .10$) to manage the risk of making Type I or II errors. One of the circumstances of this dissertation study is that multiple tests were conducted on the same data, which increases the risk of Type I errors, that is, the risk of rejecting the null hypothesis (H_0) when it is true and should not be rejected. One way to reduce the risk of making Type I errors is to be more conservative about the probability level. Thus, for the hypothesis tests in this dissertation, $\alpha = .01$.

5.1.4. Nature of the data variables. The data for each of the three independent variables was dichotomized into high and low groups by splitting each variable at the median score. This technique has a well-known drawback of data loss (Pallant, 2007, pp. 106-107). However, in order to build on previous knowledge about cognitive abilities from past search studies, it was deemed useful to analyze the data in this study similarly to the ways it was done in previous studies. In addition, it was believed that the *high-low* categorization would facilitate the construction of meaning behind search behaviors of those groups that using a method such as regression (where effect is measured as a percent contribution, but does not distinguish one participant from another) would not allow for.

5.1.5. Data distributions. In order to determine which statistical tests to apply to study data it is important to first gather information about specific characteristics and properties of the data that affect the assumptions of the different kinds of tests. Some characteristics are straightforward to know, such as the types of variables that were used (e.g., nominal, ordinal, interval). Other characteristics require analysis of the data, such as the shapes of the data distributions and comparisons of data variances across groups. In this section, assumptions related to the normal distribution and homogeneity of variances are discussed.

Since parametric tests are typically more powerful than their non-parametric counterparts (Pallant, 2007), it makes sense that researchers would want to run parametric tests whenever possible on their data. Parametric tests are more powerful, but they are also more restrictive in the assumptions that need to be met in order for test statistics to be accurate. If data violate assumptions for parametric tests, the researcher has several options for how to proceed with data analysis: 1) use the parametric technique anyway if that particular technique is robust enough against whatever assumption is being violated; 2) transform the data so that the assumptions of the statistical test can be met; and 3) use non-parametric statistical tests instead. Transforming data is discouraged because the transformations themselves can create new problems for the researcher to deal with (Field, 2009; Tabachnick & Fidell, 2007), so this approach was not used in the analysis of data for this study that did not meet assumptions of parametric tests.

Parametric tests such as correlations, t-tests, and analyses of variances rely on the assumption that population data are normally distributed in a bell-shaped curve known as the Gaussian distribution. In addition, some of these tests also rely on the assumption of homogeneity of variances across groups. Thus, these were the two main characteristics looked at in my data.

Two methods were used to assess the normality of the data. First, visual inspection of the data distributions were conducted using frequency distribution (i.e., histograms). Since visual inspection methods are considered unreliable (Altman & Bland, 1995), a more formal approach was also taken, that is, conducting significance tests and creating normal plots. Skewness and kurtosis were measured and a threshold of +/- 1.96 was used for determining whether the skew or kurtosis was large enough to cause a problem (Field, 2005). Since the Lilliefors-corrected Kolmogorov-Smirnov (K-S) test is no longer recommended (Ghasemi & Zahediasl, 2012), the Shapiro-Wilk test was run on data sets. To test for homogeneity of variances, the Levene test was used and the threshold for significance (i.e., the finding that the variances violate the assumption for homogeneity) was any value of $p < .05$.

Across the various data sets that comprise the study's whole data collection, there were only a few cases of assumption violations. For example, the Shapiro-Wilk test of the financial knowledge test indicated that the data was not normally distributed ($W = .939$), $p = .026$). Also, Levene's test of Homogeneity of Variances indicated that data from several of the pre-task questionnaire response violated this homogeneity assumption.

Not all of the assumption violations are important for several reasons. First, based on readings from the statistics literature (Field, 2005; Ghasemi & Zahediasl, 2012; Pallant, 2007; Rasch & Guiard, 2004), it was determined that the sample size of N = 42 is large enough to generally assume that violations of the normal assumption would not be an issue in parametric tests. In addition, two of the main tests used, the Students t-test and Analysis of Variance

(ANOVA) are known to be robust against departures from normality (Altman & Bland, 1995; Rasch & Guiard, 2004).

The assumption of homogeneity of variances is important when dealing with the ANOVA *F*-statistic. The *F*-statistic is not robust against violations of the homogeneity of variances assumption and so in cases where variances were heterogeneous (such as the pre-task questionnaire data), there were two options to use for ANOVA calculations, which were the Brown-Forsythe *F* or the Welch *F*. Both techniques work except when there is an extreme mean that has a large variance – in this case, the Welch test is the preferred statistic (Field, 2005). Since that was not the case with the data in this study, the Brown-Forsythe *F* was used in ANOVAs where variances violated the homogeneity assumption and it is indicated in the text of the analysis that the Brown-Forsythe *F* was used.

In this dissertation, in cases where test statistics deviate from the standard parametric test, this is pointed out in the text or noted in data tables. For all quantitative results reported the sample size is N = 42.

5.1.6. Assumption about scores on cognitive abilities and financial knowledge tests. I assumed that all the data collected from the sample for cognitive abilities and financial knowledge was heterogeneous enough to evoke meaningful individual differences between participants for these three independent variables.

5.1.7. Assumption about search task differences. The tasks in the study were designed as equally complex tasks and so it was expected that there would be no task effects on the independent variables. To check this assumption, task effects were tested for and in cases where task effects were found, Bonferroni post hoc tests were run to determine what the effects were, and are reported in this document.

5.2. Descriptive Statistics

This section provides descriptive statistics for the pre- and post-task questionnaires, cognitive ability and knowledge tests, search performance measures, relevance measures, mental workload questionnaire, and eye tracking data.

5.2.1. Pre- and post-task questionnaires. At the beginning of each search task, participants filled out a five-item online questionnaire (Appendix D) about the task topic. The questionnaire measured items such as participants' knowledge of the task topic and their perceptions of how difficult it would be to search for information about the topic. Responses were measured on a five-point Likert-type scale in which the score of *1* indicated the lowest or most negative responses (e.g., "nothing," "not at all," etc.) and scores of *5* indicated the highest or most positive responses (e.g., "very much," very often," etc.). Table 8 shows the means and standard deviations of participants' responses.

Table 8. *Means and Standard Deviations for Pre-Task Questionnaire Responses*

	Task Topic			
Pre-task Question	Reverse Mortgage	Payday Loan	Student Loan	Total
Knowledge of Topic	1.88 (.97)	2.12 (1.27)	2.83 (1.42)	2.28 (1.28)
Relevance of Topic	1.88 (1.15)	2.10 (1.28)	3.62 (1.58)	2.53 (1.55)
Interest in Learning	2.86 (1.14)	2.98 (1.22)	3.64 (1.30)	3.16 (1.26)
Frequency of Searching	1.21 (.57)	1.45 (.83)	2.40 (1.45)	1.69 (1.14)
How Difficult to Search	2.48 (.83)	2.83 (1.01)	2.19 (.83)	2.50 (.93)

A one-way analysis of variance (ANOVA) was calculated on the pre-task question item to determine if there were task effects. The one-way ANOVA results shown in Table 9 uses the more robust Brown-Forsythe *F*-ratio because several of the pre-task questionnaire response data sets violated the homogeneity of variance assumption (Field, 2005, p. 347). There was a main effect for task topic on participants' knowledge of the task topic, $F(2, 114) = 6.87$, $p < .001$, $\eta^2 = .101$. Bonferroni post-hoc analysis[7] indicated that participants' rated their knowledge of the student loan task topic higher than both the reverse mortgage and payday loan task topics. There was a main effect for task topic on participants' perception of the relevance of the topic to their own lives, $F(2, 114.75) = 20.72$, $p < .001$, $\eta^2 = .252$. Bonferroni post-hoc analysis indicated that participants rated the student loan task topic as most relevant to their lives, over both reverse mortgage and payday loan task topics. There was a main effect for task topic on participants' interest in learning about the task topic, $F(2, 121.53) = 5.04$, $p = .008$, $\eta^2 = .076$. Bonferroni post-hoc analysis indicated that participants rated their interest in learning about the student loans topic as greater than both the reverse mortgage and the payday loan task topics. There was a main effect for task topic on how frequently participants reported having ever searched online for the task topic, $F(2,79.52) = 16.07$, $p < .001$, $\eta^2 = .207$. There was a main effect for task topic on participants' perceptions of how difficult it would be to search on the task topics, $F(2, 118.66) = 5.43$, $p = .006$, $\eta^2 = .081$. Bonferroni post-hoc analysis indicated that participants expected the payday loan task topic to be more difficult to search for than the student loan task topic.

[7] When using the Brown-Forsythe *F*-ratio, the degrees of freedom for the post-hoc analysis are taken from the Robust Tests of Equality of Means.

Table 9. *Results of ANOVA for Pre-Task Questionnaire, Using Brown-Forsythe F-ratio*

Pre-task Question	*SS*	df_m	df_r	*F*	η^2
Knowledge of Topic	20.64	2	114.0	6.87**	.101
Relevance of Topic	75.44	2	114.7	20.72***	.252
Interest in Learning Topic	15.06	2	121.5	5.04**	.076
Frequency of Searching	33.33	2	79.5	16.07***	.207
Difficulty in Searching Topic	8.71	2	118.7	5.43**	.081

Note. ** $p < .01$; *** $p < .001$

After completing each search task, participants filled out an online questionnaire that contained three post-task questions (Appendix F) and the six-question NASA-TLX mental workload index (Hart & Staveland, 1988) (Appendix G). The questionnaire items were measured on a 10-point Likert-type scale. During the data analysis, in order to compare the pre-task "difficulty" question with the post-task "difficulty question," the 10-point scale of the post-task responses was transformed to a 5-point scale, using SPSS to calculate the linear transformation. Linear transformation approximates the interval scale to the same degree after the transformation such that the interval and other properties of the distributed are unaffected[8]. This enabled comparison for paired sample *t*-tests of the two items. Table 10 shows the results of the 5-point transformation of the post-task questionnaire items. Questions ask how difficult it was for participants to find relevant items (*Difficulty Finding Documents*, 1=Very easy, 5=Very difficult), how participants rated their own skills at finding relevant items (*Your Ability to Find Relevant*, 1=Not good, 5=Very good), and how participants rated the system's ability at retrieving relevant items (*Rate System's Ability*, 1=Not good, 5=Very good). Across the three

[8] http://psychstat3.missouristate.edu/Documents/IntroBook3/sbk13.htm

tasks, participants rated the difficulty in finding relevant documents an average of 3.85 on the scale from 1 to 5, their ability to find relevant documents an average of 3.9, and the system's ability to retrieve relevant documents an average of 4.29.

Table 10. *Means and Standard Deviations for Post-Task Questionnaire Items*

	Task Topic			
Post-task Question	Reverse Mortgage	Payday Loan	Student Loan	Total
Difficulty Finding Documents	3.71 (1.10)	3.85 (1.13)	4.00 (1.07)	3.85 (1.10)
Ability to Find Relevant Documents	3.80 (1.04)	3.81 (1.19)	4.07 (.92)	3.90 (1.06)
System's Ability to Retrieve	4.25 (.77)	4.39 (.56)	4.24 (.79)	4.29 (.71)

A one-way ANOVA was run to determine if there was a task effect on any of the post-task questions. The ANOVA used the Brown-Forsythe *F*-ratio and indicated no significant differences across the three tasks on the post-task questions, which meant that participants did not find any of the tasks more difficult than others, nor did they believe they performed differently on any of the tasks versus the others, nor did they think the system performed differently for any of the tasks versus the others.

On the pre-task and post-task questionnaires, a question was asked about the difficulty of the tasks. For the pre-task questionnaire, this item was worded, *"How difficult do you think it will be to search for information about this topic?"* with 1 = very easy and 5 = very hard. For the post-task questionnaire, the item was worded, "*How difficult was it to find relevant documents?"* with 1 = very difficult and 5 = very easy. To understand if there was a difference between the amount of difficulty participants expected searching to be before the task and the how difficult they felt the task was after they completed the task, a paired-samples t-test was conducted to compare the pre- and post-task responses. There was a significant difference in the scores for difficulty in the pre-task (*M* = 2.48, *SD* = .833) and post-task (*M* = 3.71, *SD* = 1.100) questions for the reverse mortgage task topic, $t(41) = -5.047$, $p < .001$. There was a significant difference in the scores for difficulty in the pre-task (*M* = 2.83, *SD* = 1.010) and post-task (*M* = 3.85, *SD* = 1.134) questions for the payday loan task topic, $t(41) = -3.937$, $p < .000$. There was also a significant difference in the scores for difficulty in the pre-task (*M* = 2.19, *SD* = .833) and post-task (*M* = 4.00, *SD* = 1.065) questions for the student loan task topic, $t(41) = -7.705$, $p < .001$. In all cases, participants experienced searching for the task topics as easier than they originally expected.

5.2.2. Cognitive abilities. There were two cognitive abilities measured: perceptual speed and memory span. Perceptual speed was measured using the *Finding A's* test from the Ekstrom Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976a). Table 11 shows descriptive statistics for perceptual speed. The average score on the perceptual speed test was 63.1 (*SD* = 18.2) out of a possible score of 200. The median score for perceptual speed was 59 which was used as the cut-point score for the median split for low and high perceptual speed groups. Participants who scored less than or equal to 59 were grouped in the low perceptual speed group (N = 23) and those who scored higher than 59 were group in the high perceptual speed group (N = 19).

Table 11. *Means and Standard Deviations of Perceptual Speed Test (N=42)*

Cognitive Test	n	Min	Max	*M*	*SD*
Perceptual Speed	42	30	109	63.10	18.17

The distribution of perceptual speed scores is shown in Figure 14.

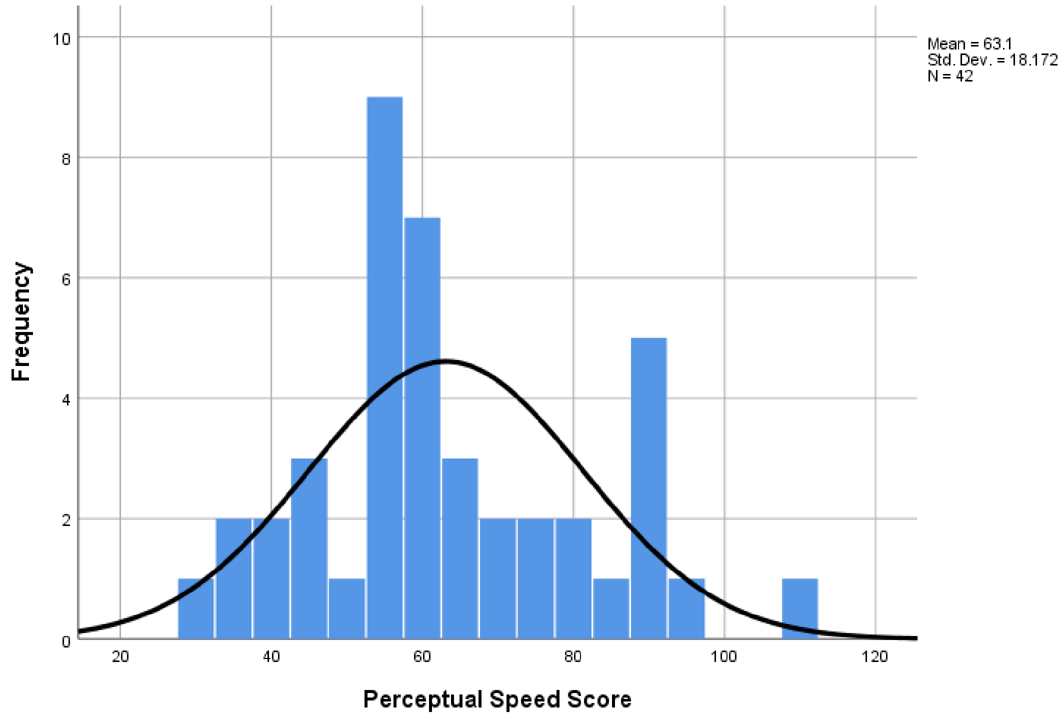


Figure 14. Score distribution for participants on the perceptual speed test.

Working memory was measured using a memory span test from CogLab 2.0 (Francis et al., 2008). Table 12 shows the descriptive statistics for the memory span test. The average score on the memory test was 5.65 (*SD* = .644) out of a possible score of 10. The median score was 5.70 which was used as the cut-point score for the median split for low and high working memory groups. Participants who scored less than 5.70 were grouped in the low working memory group (N = 21) and those who scored higher than 5.70 were grouped in the high working memory group (N = 21).

Table 12. *Means and Standard Deviations of Memory Span Test (N=42)*

Cognitive Test	n	Min	Max	*M*	*SD*
Memory Span	42	4.2	6.8	5.65	.64

The distribution of memory span scores is shown in Figure 15.

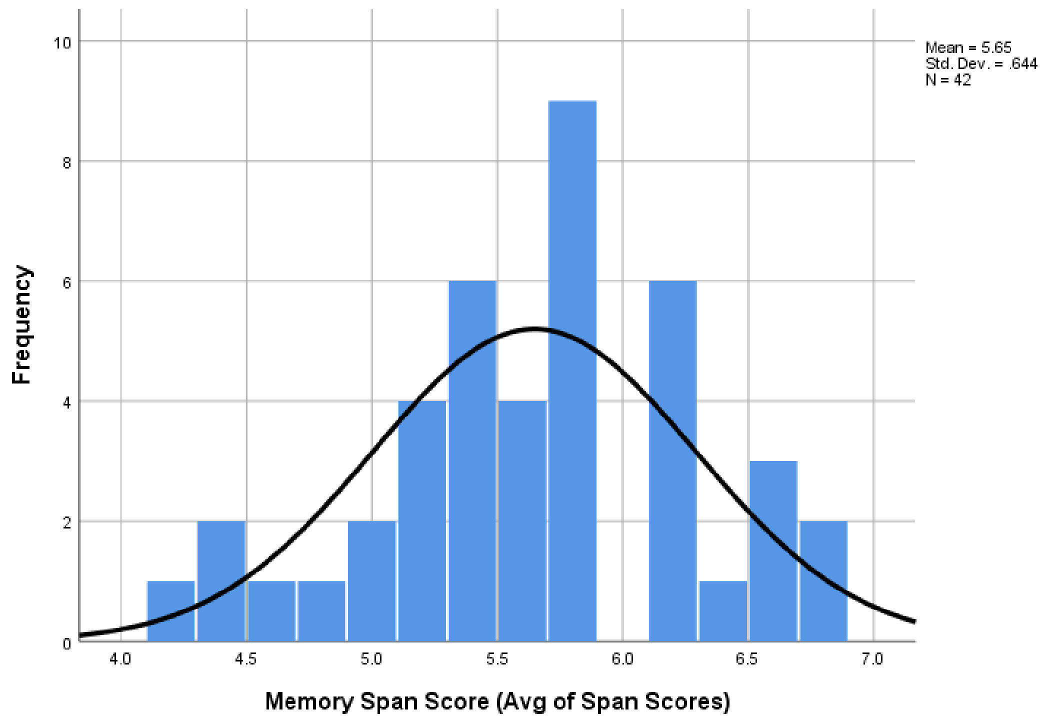


Figure 15. Score distribution for participants on the memory span test.

5.2.3. Financial knowledge. Table 13 shows participants' scores on the financial knowledge test. The highest possible score on the financial knowledge test was an eight, if all eight questions were answered correctly. No participant received a perfect score. The average score on the financial knowledge test was 4.64 (*SD* = 1.340). The median score was 5 and this was used as the cut-point for the median split for low and high financial knowledge. Participants who scored lower than 5 were grouped in the low financial knowledge group (N = 19) and those who score 5 or higher were grouped in the high financial knowledge group (N = 23). Figure 16 shows the distribution of scores for the test.

Table 13. *Descriptive Statistics, Financial Knowledge Scores (N = 42)*

	N	Min	Max	*M*	*SD*
Financial Knowledge test	42	2	7	4.64	1.34

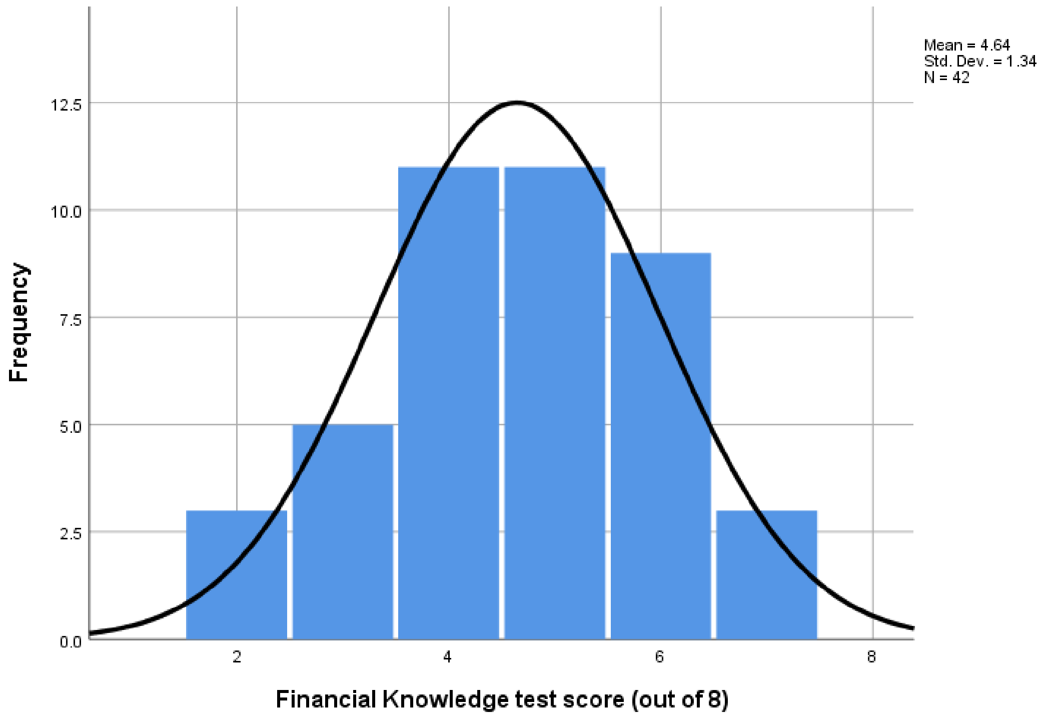


Figure 16. Score distribution for participants on the financial knowledge test.

5.2.4. Search behaviors. Search performance was operationalized using search behavior data collected using the system logging software of the Tobii eye tracker. Four categories of data were gathered – *Clicks* (e.g., clicks on SERPs and webpages as well as time to first click), *Queries* (e.g., queries entered, unique queries entered, and length of queries), *SERPs* (e.g., SERPs displayed and SERP display time), and *Webpages* (e.g., webpages visited, unique webpages visited, and webpage display time). Descriptive statistics for these measures by task and total are described in the next subsections.

5.2.4.1. Clicks. Clicks per task and total were collected. Table 14 shows the average number of clicks per task and total, the average number of clicks on SERPs per task and total, and the average time to first click in seconds per task and total. The average number of clicks per task was 82.8 (*SD 38.54),* the average number of clicks per SERP was 19.6 (*SD 14.01)*, and the average time to first click was 15.2 seconds *(SD 27.08)*. One-way ANOVAs for the number of clicks and time to first click and a one-way ANOVA using the Brown-Forsythe *F* for the number of SERP clicks indicated no significant differences across the three tasks for any of these measures.

Table 14. *Means and Standard Deviations for Click Data*

	Task Topic			
Search Measure	Reverse Mortgage	Payday Loan	Student Loan	Total
Number of Clicks	89.6 (38.77)	73.9 (35.86)	85.0 (40.10)	82.8 (38.54)
Number of SERP Clicks	21.2 (15.72)	17.0 (10.53)	20.1 (15.15)	19.6 (14.01)
Time to 1st Click (Seconds)	14.6 (16.77)	15.4 (30.82)	15.5 (31.69)	15.2 (27.08)

5.2.4.2. Queries. Data for the mean number of queries, mean number of unique queries, and mean length of queries were collected. Table 15 shows the means and standard deviations for each measure. On average across the three tasks, participants issued 7.4 queries per task *(SD* = 5.24). The number of unique queries issued was an average of 5.9 per task *(SD* = 3.16*)*. The average query length per task was 4.5 words *(SD* = 1.38). One-way ANOVAs indicated there were no significant differences in the number of queries and number of unique queries across the three tasks.

Table 15. *Means and Standard Deviations for Query Behaviors*

Search Measure	Task Topic: Reverse Mortgage	Payday Loan	Student Loan	Total
Number of Queries	8.0 (5.50)	7.6 (5.29)	6.7 (4.95)	7.4 (5.24)
Number Unique Queries	6.3 (3.00)	6.1 (3.68)	5.1 (2.68)	5.9 (3.16)
Average Query Length	5.0 (1.30)	4.3 (1.62)	4.3 (1.08)	4.5 (1.38)

5.2.4.3. SERPs and webpages. Data was collected about the number of SERPs that participants viewed and the number of webpages they viewed. Means and standard deviations for this data are shown in Table 16. On average, participants visited 19.6 SERPs (*SD* = 11.44) per task. They also visited an average of 17.8 webpages (*SD* = 10.99). However, as noted earlier in Section 5.1.2., the data for both the number of SERPs and the number of webpages visited was inflated because users had to back over webpage and SERPs in order to navigate back to the search engine. A better representation of the participants' visits to webpages is the average number of *Unique Webpages* visited, which was an average of 10.8 per task (*SD* = 4.71). The number of unique webpages visited per query on average was 2.2 pages (*SD* = 1.94) and the number of unique website domains visited was 7.3 (*SD* = 3.01). A one-way ANOVA indicated no significant task differences for any of these measures.

Table 16. *Means and Standard Deviations for SERPs and Webpages Viewed*

Search Measure	Task Topic: Reverse Mortgage	Payday Loan	Student Loan	Total
SERPs and Webpages	41.1 (14.64)	33.6 (13.55)	37.3 (14.37)	37.3 (14.42)
SERPs	22.7 (12.93)	18.5 (9.91)	17.6 (11.00)	19.6 (11.44)
Webpages	18.5 (10.53)	15.1 (10.21)	19.7 (11.89)	17.8 (10.99)
Unique Webpages	11.6 (5.00)	9.3 (3.49)	11.6 (5.20)	10.8 (4.71)
Unique Webpages / Query	2.3 (2.08)	1.9 (2.02)	2.5 (1.67)	2.2 (1.94)
Unique Domains	8.3 (3.47)	7.4 (2.82)	6.3 (2.41)	7.3 (3.01)

5.2.5. Relevance assessments. Relevance assessments were collected from participants during each of the three tasks via the bookmarking function in Internet Explorer (IE). Folders were set up ahead of time in IE for the four grades of relevance: not relevant, somewhat relevant, relevant, and very relevant. After each participant's first session of the study (the searching part), the bookmarks were downloaded from IE into a spreadsheet that was organized by participant ID and task ID. Table 17 shows means and standard deviations for the relevance bookmarks by grade for each task. On average, participants graded less than one document per task as *not relevant* (M = .52, SD = .90), 1.60 (SD = 1.49) documents per task as *somewhat relevant,* 2.70 (SD = 2.05) documents per task as *relevant,* and 3.91 (SD = 2.60) documents per task as *very relevant.* A one-way ANOVA indicated that there were no significant differences across the three tasks for the proportion of relevance grades assigned to webpages.

Table 17. *Means and Standard Deviations for Bookmarks by Relevance Grade*

	Task Topic			
Relevance Grade	Reverse Mortgage	Payday Loan	Student Loan	Total
Not Relevant	.48 (.80)	.62 (.99)	.45 (.92)	.52 (.90)
Somewhat Relevant	1.76 (1.77)	1.52 (1.44)	1.50 (1.24)	1.60 (1.49)
Relevant	2.50 (1.64)	2.45 (1.78)	3.14 (2.57)	2.70 (2.05)
Very Relevant	4.40 (2.62)	3.29 (2.13)	4.05 (2.91)	3.91 (2.60)

In addition to user relevance assessment, as described earlier in the literature review, this dissertation study is also interested in the construct of domain relevance as explored by Hjørland (2010). Domain relevance in the dissertation study was captured using experts' relevance judgments of the corpus of webpages created from the union of all participant's webpage selections. Table 18 shows the cross-tabulated results of the two assessor's relevance judgments.

Table 18. *Cross-tabulated Results of Expert Assessor's Relevance Judgments*

		Expert 1 Judgments				
		Not Relevant	Somewhat Relevant	Relevant	Very Relevant	Total
Expert 2 Judgments	Not Relevant	191	153	45	18	407
	Somewhat Relevant	17	72	54	15	158
	Relevant	2	10	25	10	47
	Very Relevant	1	16	14	4	35
Total		211	251	138	47	647

Several approaches were taken for measuring assessor agreement. First was to use non-binary judgment groupings and calculate two versions of Cohen's Kappa: the Linear Kappa measured .275 and the Quadratic Kappa measured .356. In the second approach, the judgment categories were collapsed to create binary judgments. This binary data was calculated several ways. First the *Not Relevant* judgments comprised the *Not Relevant* category and the *Somewhat Relevant, Relevant,* and *Very Relevant* judgments were combined into the *Relevant* category. The Cohen's Kappa for this was .331.

Two other binary configurations were also calculated. The first was to combine the *Not Relevant* and *Somewhat* Relevant judgments into the *Not Relevant* category while combining the *Relevant* and *Very Relevant* judgments into the *Relevant* category. The second was to combine the *Not Relevant, Somewhat Relevant,* and *Relevant* judgments into *Not Relevant* and use the *Very Relevant* judgments for the *Relevant* category. Neither of these approaches were worthwhile, however. Thus, the final weighted kappa was .356 and the non-weighted was .331.

5.2.6. Mental workload questionnaire. Participants were asked to complete the NASA-TLX mental workload questionnaire (Hart & Staveland, 1988) at the end of each task. Table 19 shows the descriptive statistics for the questionnaire. Responses were given on a 10-point scale. The average mental workload score for all participants across all tasks was 4.03 (*SD = 1.095*). One-way ANOVAs indicated no significant differences across the three tasks for any of the components of the TLX index nor for the final average scores.

Table 19. *Means and Standard Deviations for NASA-TLX Mental Workload Questionnaire*

	Task Topic			
Questionnaire Item	Reverse Mortgage	Payday Loan	Student Loan	Total
How mentally demanding was it to complete the search task?	4.3 (2.06)	4.2 (2.06)	3.8 (1.91)	4.1 (2.00)
How physically demanding was it to complete the search task?	1.7 (1.47)	1.7 (1.37)	1.5 (1.04)	1.6 (1.30)
How hurried or rushed did you feel while completing the search task?	3.4 (2.13)	3.2 (2.13)	3.6 (2.31)	3.4 (2.18)
How successful were you in completing the search task?	7.3 (2.14)	8.0 (1.67)	7.9 (1.70)	7.7 (1.85)
How hard did you work to accomplish your level of performance	4.7 (2.37)	4.2 (2.15)	4.2 (2.24)	4.4 (2.25)
How insecure, discourage, irritated, stressed, or annoyed did you feel while completing the search task?	3.1 (2.30)	2.9 (2.11)	2.9 (2.05)	3.00 (2.14)
Average TLX Score	4.1 (1.11)	4.03 (1.13)	4.0 (1.07)	4.0 (1.10)

In order to explore the factor structure underlying the TLX, a Principal Component Analysis was conducted to compute composite scores for the factors underlying the index questions in the TLX. The value of conducting the analysis is to see if there is valid evidence supporting the conclusion that the scores from the TLX are a valid assessment of participants' mental workload. Initial Eigen values (Table 20) indicated that the first three factors explained 53.10%, 15.49%, and 12.63% of the variance respectively.

Table 20. *Total Variance Explained, NASA-TLX Component Analysis*

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.186	53.105	53.105
2	.929	15.490	68.594
3	.757	12.625	81.219
4	.486	8.099	89.317
5	.393	6.551	95.868
6	.248	4.132	100.000

The communalities (Table 21) show the proportion of each variable's variance that can be explained by the underlying latent variables. Most of the extraction values are high, indicating that the extracted components represent the variables well. The lowest value is for the *physically demanding* question, which may not be well represented in this study that has very little physical effort required.

Table 21. *Communalities for the NASA-TLX Mental Workload Questionnaire*

Questionnaire Item	Extraction[a]
How mentally demanding was it to complete the search task?	.748
How physically demanding was it to complete the search task?	.206
How hurried or rushed did you feel while completing the search task?	.409
How successful were you in completing the search task?	.525
How hard did you work to accomplish your level of performance	.680
How insecure, discourage, irritated, stressed, or annoyed did you feel while completing the search task?	.618

[a] Extraction method: Principal Component Analysis

The scree plot is shown in Figure 17. The scree plot graphs the Eigenvalues against each factor (or component) number. In this graph, from the third factor on, the line becomes flatter and flatter, indicating that each factor is adding smaller and smaller amounts to the total variance.

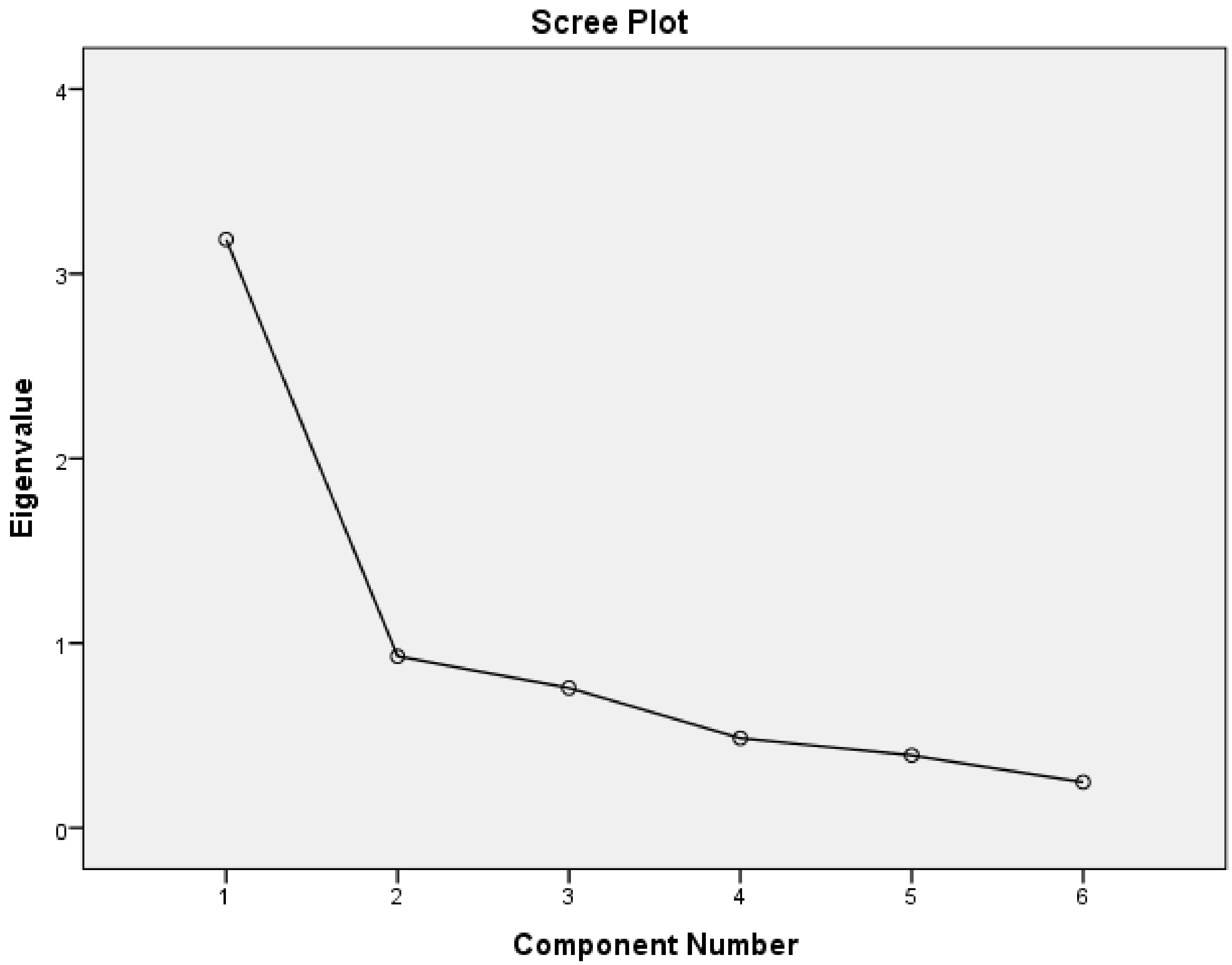


Figure 17. Scree plot for the principal component analysis of the NASA-TLX Mental Workload Questionnaire.

5.2.7. Eye movement data. Eye gaze data was collected from participants using a Tobii X2-60 eye tracker system and the Tobii eye tracking software package. Fixation counts were captured along with fixation durations at 60 Hz. Table 22 shows the means and standard deviations for the fixation data. A one-way ANOVA indicated that there were no significant differences for these measures across the three tasks.

Table 22. *Means and Standard Deviations for Eye Fixation Data*

Fixation Measure	Task Topic: Reverse Mortgage	Payday Loan	Student Loan	Total
Fixation count	1,685 (628)	1,509 (682)	1,728 (634)	1,641 (650)
Fixations on SERPs	459 (274)	418 (316)	404 (272)	427 (287)
Fixations on Webpages	1,115 (529)	1,000 (510)	1,165 (511)	1,094 (517)

Fixation duration data was also captured with the eye tracker. Table 23 shows the means and standard deviations of the sums of the total fixations durations, of the fixation durations on SERPs and of fixation durations on webpages. A one-way ANOVA indicated that there were no significant differences for these measures across the three tasks.

Table 23. *Means and Standard Deviations for Fixation Duration Measures*

Fixation Duration Measure	Task Topic: Reverse Mortgage	Payday Loan	Student Loan	Total
Sum of Fixation Durations	326,554 (146,408)	291,145 (159,791)	330,739 (149,690)	316,146 (151,898)
Fixation Durations on SERPs	90,070 (57,500)	83,967 (70,322)	80,893 (61,863)	84,976 (63,058)
Fixation Durations on Webpages	216,738 (117,568)	189,666 (110,310)	218,186 (107,047)	208,197 (111,610)

5.3. Hypothesis and Exploratory Results with Corresponding Statistical Models

This section presents the results of the hypotheses tests and exploratory results. Each measure and result is explained in detail, however, none of the hypotheses nor exploratory measures were supported.

5.3.1. Perceptual speed model.

5.3.1.1. Hypothesis #1. The first hypothesis stated that participants with higher perceptual speed ability would interact more while searching, as manifested by issuing longer queries, having more clicks on SERPs, and viewing more URLs per query, and viewing more (unique) URLs per task than participants with lower perceptual speed ability.

Participants with higher perceptual speed issued shorter queries on average per task (M = 4.29, SD = .55) than those with lower perceptual speed (M = 4.7, SD = 1.27), $t(40)$ = 1.34, p = .186, d = .431. Participants with higher perceptual speed had fewer clicks on SERPs in total (M = 50.4, SD = 37.19) than those with lower perceptual speed (M = 65.6, SD = 35.84), $t(40)$ = 1.34, p = .189, d = .414. Participants with higher perceptual speed viewed about the same number of unique URLs per query on average per task (M = 1.9, SD = 1.12) than those with lower perceptual speed (M = 2.1, SD = 1.46), $t(40)$ = .512, p = .612, d = .161. Participants with higher perceptual speed viewed fewer unique URLS on average per task (M = 10.2, SD = 3.49) than those with lower perceptual speed (M = 11.3, SD = 3.38), $t(40)$ = 1.09, p = .282, d = .337.

Independent samples *t*-tests indicated there were no significant differences between participants with high versus low perceptual speed ability for any of these measures. The null hypothesis could not be rejected, therefore, Hypothesis #1 was not supported.

5.3.1.2. Hypothesis #2a and b. The second hypothesis had two parts. Part (a) stated that participants with higher perceptual speed ability would bookmark their first relevant webpage faster than those with lower perceptual speed ability. The search measure, *Average Time to 1st Click*, was measured as the time it took the participant to click on the first result in the SERP. This measure is used as a proxy for selecting the first relevant webpage, because there was no reliable way to accurately calculate the time it took participants to bookmark their first pages. Independent samples *t*-tests were calculated to test the effects of perceptual speed on the *Average Time to 1st click* measure, but test results indicated there were no significant differences between participants with high perceptual speed ability (M = 21.57, SD = 37.68) versus low perceptual speed ability (M = 9.85, SD = 4.85), $t(40) = -1.480$, $p = .147$, $d = .436$.

Even though test results were not statistically significant, there were differences between the two groups on the individual tasks that are worth noting. The data in Table 24 shows the comparison of means and standard deviations for the average time to first click for participants with higher versus lower perceptual speed ability, and the totals averaged across the three tasks. For the Reverse Mortgage task topic, participants with higher perceptual speed took more than 7.5 seconds longer to click on the first SERP result than participants with lower perceptual speed. This pattern was consistent across all tasks. High perceptual speed participants took more than twice as much time on average, 11 seconds longer, to click on the first results and in the Student Loan task topic the difference was even larger, with high perceptual speed participants taking an average of 16.5 seconds longer to click on their first result than the low perceptual speed participants. Across all tasks, the average amount of additional time taken before clicking on the first result by participants with higher perceptual speed was 11.7 seconds. The trend of these results is in the opposite direction of the stated hypothesis. The null hypothesis could not be rejected therefore, Hypothesis #2a was not supported.

Table 24. *Means and Standard Deviations for Time to 1st Click by Perceptual Speed Level*

Measure	PS Level	Task Topic: Reverse Mortgage	Task Topic: Payday Loan	Task Topic: Student Loan	Total
Time to 1st Click	L	11.12 (8.24)	10.42 (5.70)	8.02 (3.12)	9.85 (4.85)
	H	18.76 (22.88)	21.40 (45.33)	24.56 (46.02)	21.57 (37.68)

The second part of Hypothesis #2 (part b), stated that participants with higher perceptual speed ability would achieve greater *Interactive User Precision (IUP)* than those with lower perceptual speed ability. IUP was calculated as follows:

1. Each relevance grade was given a score. *Not Relevant* was equal to 1, *Somewhat Relevant* was equal to 2, *Relevant* was equal to 3, and *Very Relevant* was equal to a score of 4.
2. Each assessor's grades were then converted into the scores from Step 1.
3. For each webpage, the two assessor's scores were averaged together. The average score became the "Expert Relevant" score. This is analogous to TREC studies in which relevance assessments from experts are called "TREC Relevant" scores.
4. Webpages that received an Expert Relevant score of 2 or less were categorized as *Not Relevant* and each webpage that scored greater than 2 was categorized as *Relevant.*
5. Participants relevance grades were compared to the Expert Relevant score for the pages they selected. All pages that a participant scored as relevant that was also scored Expert Relevant was counted toward the participant's interactive user precision score.

6. To calculate IUP for each participant, the number of “Expert Relevant” webpages they bookmarked was divided by the total number of webpages they had judged. The resulting scores ranged between 0 to 1.

Although the direction of the difference between the two groups was in line with the stated hypothesis, the independent samples *t*-test resulted in no significant difference in IUP for participants with higher perceptual speed ability (M = .362, SD = .141) versus those with lower perceptual speed ability (M = .349, SD = .111), $t(40)$ = -.318, p = .752, d = .094. The null hypothesis could not be rejected and therefore, Hypothesis #2b was not supported.

5.3.1.3. Hypothesis #3. The third hypothesis stated that participants with higher perceptual speed ability would experience less mental workload, manifested in lower eye gaze measures for mean fixation duration and standard deviation of fixation durations, and lower scores on the self-report mental workload questionnaire, than those with lower perceptual speed ability. Participants with higher perceptual speed reported lower mental workload (M = 3.9, SD = .84) on the NASA-TLX than those with lower perceptual speed (M = 4.2, SD = 1.0), $t(40)$ = 1.017, p = .315, d = .318. Participants with higher perceptual speed had lower mean fixation durations (M = 174.0, SD = 36.18) than those with lower perceptual speed (M = 193.0, SD = 31.24), t(40) = 1.86, p = .071, d = .572. Participants with higher perceptual speed had lower standard deviations of fixation durations (M = 118.0, SD = 31.24) than those with lower perceptual speed (M = 138.0, SD = 28.78), t(40) = 2.16, p = .037, d = .666. None of the differences in these measures were significant, thus, Hypothesis #3 was not supported.

5.3.2. Working memory model.

5.3.2.1. Hypothesis #4a and b. The fourth hypothesis had two parts. Part (a) stated that participants with higher working memory ability would issue more unique queries and open more unique webpages than those with lower working memory ability. Participants with higher working memory ability issued fewer unique queries on average per task (*M* = 5.3, *SD* = 2.2) than those with lower working memory ability (*M* = 6.4, *SD* = 3.0), *t*(40) = 1.277, *p* = .209, *d* = .394, but this difference was not significant. Participants with higher working memory also opened fewer unique webpages (*M* = 10.7, *SD* = 3.9) than those with lower abilities (*M* = 10.9, *SD* = 3.0), *t*(40) = .177, *p* = .860, *d* = .055, and this difference was also not significant. The null hypothesis could not be rejected and Hypothesis #4a was not supported.

Part (b) of Hypothesis #4 stated that participants with lower working memory would have more fixations on average on SERPs and webpages and would also have longer fixation duration measures than those with higher working memory. In terms of the fixation count measures, participants with lower working memory had fewer average fixations per SERP (*M* = 11.3, *SD* = 7.1) than those with higher working memory (*M* = 12.1, *SD* = 6.5), *t*(40) = -.400, *p* = .692, *d* = .123. Participants with lower working memory also had fewer average fixations per webpage (*M* = 28.8, *SD* = 16.7) than those with higher working memory (*M* = 35.6, *SD* = 15.4), *t*(40) = -1.367, *p* = .179, *d* = .422. None of these differences were statistically significant. Participants with lower working memory had shorter average length of fixation durations per task (*M* = 182.7, *SD* = 39.6) than those with higher working memory (*M* = 186.3, *SD* = 29.6), *t*(40) = -.334, *p* = .740, *d* = .017, but this difference was not statistically significant. The null hypothesis could not be rejected and Hypothesis #4b was not supported.

5.3.2.2. Hypothesis #5. The fifth hypothesis stated that participants with lower working memory would be less selective in their evaluation behaviors such that, of the webpages they viewed, these participants would have a lower proportion of not relevant webpages than those with higher working memory ability. Participants with lower working memory did have a lower number of webpages graded *not relevant* in total (*M* = 1.29, *SD* = 1.271) than those with higher working memory (*M* = 1.95, *SD* = 3.057), *t*(40) = -.923, *p* = .362, *d* = .282, however, the difference was less than one bookmark and the independent *t*-test indicated the difference was not statistically significant. The null hypothesis could not be rejected, therefore, Hypothesis #5 was not supported.

5.3.2.3. Exploratory Measure #1. Several measures were explored to attempt to uncover possible relationships between working memory and participants' mental workload. The most straightforward was to investigate the difference between participants' NASA-TLX mental workload scores based on their membership in the high and low memory span ability groups. Participants with lower working memory reported greater overall mental workload (*M* = 4.2, *SD* = 1.011) than those with higher working memory (*M* = 3.8, *SD* = .838), *t*(40) = 1.318, *p* = .195, *d* = .153, however, this difference was not statistically significant. In a similar vein of thinking, the difference between the high and low groups was calculated on the post-task difficulty question "*How difficult was it to find relevant documents?*". The difference here was also small and not statistically significant, with participants who had lower working memory scoring slightly lower (*M* = 3.8, *SD* = .70) than those with higher working memory (*M* = 3.9, *SD* = .60), *t*(40) = -.497, *p* = .622, *d* = .153. The final exploratory measure was the amount of time participants spent on SERPs. This was based on the finding from Brennan et al. (2014), in which participants with lower working memory spent more time on SERPs across all tasks. In addition, Gwizdka (2017) compared *reading fixation durations* for high and low working memory groups searching the YASFIIRE system and found that while lower working memory searchers spent less overall time on tasks, they tended to increase their reading time on SERPs in the last phase of their search tasks. In the current study, similar to Brennan et al. (2014), participants with lower working memory spent more time on SERPs (*M* = 9 minutes, 33 seconds, *SD* = 4 minutes, 31 seconds) than those with higher working memory (*M* = 7 minutes, 55 seconds, *SD* = 4 minutes, 37 seconds), $t(40) = 1.159$, $p = .253$, $d = .358$, though this difference was not statistically significant.

5.3.3. Financial knowledge model.

5.3.3.1. Hypothesis #6. The sixth hypothesis stated that participants with higher levels of financial knowledge would issue longer queries and more queries than participants with lower levels of financial knowledge. Both measures trended in the opposite direction of the stated hypotheses. Participants with higher levels of financial knowledge issued shorter queries on average per task (*M* = 4.4, *SD* = 1.1) than those with lower financial knowledge (*M* = 4.6, *SD* = .87), $t(40) = .638$, $p = .527$, $d = .200$. They also issued fewer queries per task (*M* = 6.9, *SD* = 4.3) than those with lower financial knowledge (*M* = 8.0, *SD* = 5.2), $t(40) = .745$, $p = .461$, $d = .229$. There were no significant differences between participants with low versus high financial knowledge for either measure. The null hypothesis could not be rejected, therefore, Hypothesis #6 was not supported.

5.3.3.2. Hypothesis #7. The seventh hypothesis stated that participants with higher levels of financial knowledge would bookmark a greater number of webpages than participants with lower levels of financial knowledge. While the direction of the results was in-line with the hypothesis, that is, participants with higher financial knowledge did bookmark more webpages on average per task (*M* = 9.4, *S D* =3.7) than those with lower financial knowledge (*M* = 8.0, *SD* = 3.5), $t(40) = -1.336$, $p = .189$, $d = .413$, the difference was not significant. The null hypothesis could not be rejected, therefore, Hypothesis #7 was not supported.

5.3.3.3. Exploratory measure #2. Several exploratory measures were investigated to uncover a possible relationship between financial knowledge level and mental workload. The measure entailed independent sample *t*-tests to compare high and low levels of financial knowledge with each of the following: total score on the NASA-TLX mental workload questionnaire, $t(40) = .435$, $p = .666$, $d = .1355$, score on the post-task question "*How difficult was it to find relevant documents?*", $t(40) = .128$, $p = .899$, $d = .040$, and scores for IUP, $t(40) = .445$, $p = .659$, $d = .139$. Differences across all measures were negligible or non-existent and not statistically significant.

5.3.4. Perceptual speed and financial knowledge interaction model.

5.3.4.1. Hypothesis #8a, b, and c. The eighth hypothesis had three parts, each comparing the average number of queries issued by one interaction group to the others. The means and standard deviations of each group are shown in Table 25. Part (a) stated that participants with higher levels of financial knowledge and perceptual speed ability would issue more queries per task than any other group of participants. This was not the case; this group issued the second largest number of average queries per task. Part (b) stated that the group with higher financial knowledge and lower perceptual speed would issue more queries per task than those with lower financial knowledge and higher perceptual speed, but this group issued an almost equal number of queries per task on average as the group with lower financial knowledge and higher perceptual speed. Part (c) stated that the group with lower financial knowledge and higher perceptual speed would issue more queries than those with lower financial knowledge and lower perceptual speed, but this group issued fewer queries than those with lower financial knowledge and lower perceptual speed.

Overall, the results do not indicate a direction of influence of financial knowledge and perceptual speed on the average number of queries per task. The between-groups ANOVA indicated that Hypothesis #8a, b, and c were not supported, $F(3, 38) = 2.407$, $p = .082$, $\eta_p^2 = .160$.

Table 25. *Means and Standard Deviations of Average Queries per Task by Financial Knowledge and Perceptual Speed Groupings*

Group	*n*	Average Queries per Task
High Financial Knowledge, High Perceptual Speed	11	8.4 (4.45)
High Financial Knowledge, Low Perceptual Speed	12	5.6 (3.96)
Low Financial Knowledge, High Perceptual Speed	8	5.5 (2.92)
Low Financial Knowledge, Low Perceptual Speed	11	9.9 (5.86)

5.3.4.2. Hypothesis #9a and b. The ninth hypothesis had two parts, each comparing the total number of webpages bookmarked by one interaction group to the others. The means and standard deviations for each group are shown in Table 26. Part (a) stated that participants with higher financial knowledge and higher perceptual speed would bookmark the most webpages. This was not the case; this group bookmarked the second largest number of webpages. Part (b) stated that participants higher financial knowledge and lower perceptual speed ability would bookmark a greater number of webpages than those with lower financial knowledge and higher perceptual speed. This was the case; the group with higher financial knowledge and lower perceptual speed saved an average of 7 more bookmarks per task than the group with lower financial knowledge and higher perceptual speed, however the Bonferroni post hoc tests of the between-groups ANOVA indicated this difference was not significant (p > .999).

The between-groups ANOVA indicated that Hypothesis #9a and b were not supported, $F(3, 38) = .740$, $p = .535$, $\eta_p^2 = .055$.

Table 26. *Means and Standard Deviations of Total Bookmarks by Financial Knowledge and Perceptual Speed Groupings*

Group	*n*	Total Bookmarks
High Financial Knowledge, High Perceptual Speed	11	26.3 (7.49)
High Financial Knowledge, Low Perceptual Speed	12	29.8 (13.72)
Low Financial Knowledge, High Perceptual Speed	8	22.8 (14.94)
Low Financial Knowledge, Low Perceptual Speed	11	26.0 (11.40)

The null hypothesis for Hypothesis #9a and b could not be rejected, therefore, Hypothesis #9 was not supported.

5.3.4.3. Exploratory measure #3. An exploratory measure was not explored for this model because the relationships of the two IVs were not significant in earlier models. Thus, there was no compelling evidence suggesting that an exploratory measure would generate meaningful findings.

5.3.5. Working memory and financial knowledge interaction model.

5.3.5.1. Exploratory measure #4. An exploratory measure was not explored for this model because the relationships of the two IVs were not significant in earlier models. Thus, there was no compelling evidence suggesting that an exploratory measure would generate meaningful findings.

5.3.5.2. Exploratory measure #5. An exploratory measure was not explored for this model because the relationships of the two IVs were not significant in earlier models. Thus, there was no compelling evidence suggesting that an exploratory measure would generate meaningful findings.

5.3.5.3. Exploratory measure #6. An exploratory measure was not explored for this model because the relationships of the two IVs were not significant in earlier models. Thus, there was no compelling evidence suggesting that an exploratory measure would generate meaningful findings.

5.4. Qualitative Data Analysis and Results

Data for the qualitative research portion of the dissertation was collected for the last search task that each participant conducted using stimulated recall techniques. The 44 recorded sessions were transcribed by a professional transcription service into text format. Procedures for coding and analyzing the data followed guidance from the literature (Saldana, 2009). A 20% sample of the data was used to create a codebook and then those codes were applied to the remainder of the sample. A hierarchical coding system was developed with top-level nodes that categorized data related to the elements of the research question -- finding and evaluating information, financial knowledge, and cognitive abilities. Subordinate nodes emerged from the data at varying levels of granularity.

The approach for coding was to break down the components of Research Question #2 into subcomponents and then code each subcomponent. Research Question #2 states, *"RQ#2: What are users' strategies for finding and evaluating information on the Internet about different kinds of financial loans? How do users' cognitive abilities and financial knowledge influence these strategies?"*

These two questions were broken down into the individual questions, which were:

- What are users' strategies for finding information?
- What are users' strategies for evaluating information?
- How do users' cognitive abilities (memory span and perceptual speed) influence these strategies?
- How does users' financial knowledge influence these strategies?

This breakdown was used for determining the five top-level nodes for the coding activity: finding information, evaluating information, financial knowledge, cognitive ability – memory, and cognitive ability – perceptual speed. However, the coding analysis was not useful for answering the questions about users' cognitive abilities and financial knowledge (further discussion on this is included in the discussion section), so these nodes are not included in the table showing the top-level nodes (Table 27).

Table 27. *Main Coding Nodes for Research Question #2.*

Coding Node	No. Participants	No. References
01. Finding Information	44	420
02. Evaluating Information	44	359

All participants mentioned multiple tactics and strategies for both finding and evaluating information. It was not possible to effectively code for perceptual speed, memory, and financial knowledge, because only a small number of instances were found that could possibly interpreted as relating to participants' abilities and knowledge. Details about the attempts to address this part of Research Question #2 are covered in the discussion section. The current section covers the first part of Research Question #2, which asks about finding and evaluating information.

During and after the qualitative data analysis, I took several steps to increase the trustworthiness of the data and its analysis. First, I discussed my coding process and findings with my doctoral advisor regularly. Second, I asked for feedback on specific parts of the data analysis from several expert researchers in my department at UNC and from other academic institutions at a research conference I attended. Third, I conducted a peer de-brief with a doctoral student peer, Leslie Thomson, who is well-versed in qualitative methods and was also in the final phases of her doctoral studies work.

5.4.1. Strategies and tactics for finding information. The top-level node *Finding Information* was comprised of nine major subnodes, as shown in Table 28. In some cases, the subnodes contained both tactics and strategies while in other cases, the subnodes contained only tactics. The nine subnodes are tasking starting tactics and strategies, querying tactics, finding and excluding tactics and strategies, SERP tactics and strategies, exploration tactics, resources and sources tactics and strategies, advancing or accomplishing task tactics, monitoring the task tactics and strategies, and searching for self tactics and strategies.

Table 28. *Primary Code Categories for "Finding Information," Number of Participants, Number of Uttered References*

Code Categories	No. Participants	No. References
A. Task Start Tactics and Strategies	44	63
B. Resources and Sources Tactics and Strategies	38	134
C. Finding and Excluding Tactics and Strategies	33	97
D. SERP Tactics and Strategies	15	29
E. Exploration Tactics	10	18
F. Querying Tactics	5	6

5.4.1.1. Task start tactics and strategies. At the start of each stimulated recall session, a set of instructions was read aloud to each participant:

> *During this next section of the study, I'm going to play back a screen recording of the actions you took during the last task you completed. While you watch this recording, I would like for you to state aloud why you took the actions shown on the screen and what*

> *you were thinking when you took those actions. I would like you to walk me through the decision-making processes you underwent as you searched. There are no right or wrong answers here. I am simply looking for your thoughts as you review the steps you took during the experiment. Even minor thoughts will be helpful to this study.*

After reading the instructions, the screen recording was started of the last task the participant completed. This made it possible to collect information about the tactics and strategies participants used to begin their last search tasks.

Participants talked about starting their search tasks in a number of ways. By the order of the number of participants who were coded with these instances, the tactics participants used for starting their search tasks were: by formulating the first query (code: formulate a query), looking for general information about the task topic (code: gather general information), looking for specific kinds of websites to find information (code: look for specific kinds of websites), looking for specific information about the task topic (code: find specific information), actively seeking or avoiding certain criteria of information (code: employ inclusion or exclusion criteria), or re-orienting the task scenario as if it were applying to their personal situation (code: apply task to self). These codes and the number of participants who were coded with that instance, along with the number of instances of it mentioned by participants (No. References) are shown in Table 29.

Table 29. *Nodes and Subnodes for Starting the Search Task*

Nodes and Subnodes	No. Participants	No. References
A. Task Start Tactics and Strategies	44	63
Formulate a Query	16	16
Look for Specific Kinds of Websites	13	13
Gather General Information	12	12
Find Specific Information	8	9
Employ Inclusion Criteria	7	8
Apply Task to Self	2	2

The highest number of participants in the sample (N = 16 out of 44) said that the first thing they did when starting the search task was procedural, that is, to create the first search query (code: *Formulate a Query*) (Table 30). To formulate their first queries, participants used a number of approaches. The most popular approach was to use words from the task scenario as keywords (N = 11 of 16). An example of this was the explanation from P46: *"I wasn't sure what the most appropriate term to use was. I mean I've seen commercials on it and so I just looked. I used the prompt as my first attempt at terminology."* In terms of the task topics for this code, 10 of the 11 participants who used this approach for querying were completing the payday loan task. The large number of participants on this task for this code makes sense, given that of the three tasks, the payday loan task was the only task that did not provide the name of the financial product in the task scenario language. Other approaches were to simply type the financial product name into the search box (N = 3 of 16), think about the specific query terms to use (N = 1 of 16), or use the auto-complete query that came up in Google (N = 1 of 16).

Table 30. *Codes for Formulating a Query*

Subnodes	No. Participants	No. References
Formulate a Query	16	16
Use task words for query	11	11
Use product name for query	3	3
Use auto-complete for query	1	1
Think about what query terms to use	1	1

The second most frequently mentioned way (N = 13 participants) for starting the search task was to start with a narrow focus by looking for specific websites (code: *Look for Specific Kinds of Websites*) (Table 31). The most often mentioned type of specific website mentioned by participants were government websites (N = 9 participants). These participants indicated that they intentionally looked for government websites or government-sponsored information at the beginning of their search processes. Four of these participants were completing the reverse mortgage task, three were completing the student loans task, and two were completing the payday loans task. One participant explained his strategy for starting his search about the student loan task in this way: “One of the first things I wanted to do instead of going to the .coms was to go to ed.gov because what better way to start searching then (on) something that’s sponsored by the government?” (P21). Other participants chose to start their searches on government websites once they saw those websites appear on the SERP: “a government site came up in search results, so I figured that was a good place to start” (P02). Other kinds of sites participants mentioned they looked for were non-profit websites with the “.org” domain, educational websites with the “.edu” domain, and commercial website with the “.net” domain. Two participants indicated they went to websites they knew about.

Table 31. *Codes for Looking for Specific Kinds of Websites*

Subnodes	No. Participants	No. References
Look for Specific Kinds of Websites	13	13
.gov or government websites	9	9
.org and .edu first	1	1
.org and .net first	1	1
Go to site I have used before	1	1
I know exactly which site to go to	1	1

Some participants (N = 12) took a more general approach to starting their search tasks by looking for general information about the financial product in the task (code: *Gather General Information*) (Table 32). Five participants who were all completing the reverse mortgage task, said they started the task by looking for a definition for the financial product. Three participants, all completing the student loans task, said they wanted to search for basic information. Three participants said they wanted to start by looking for general or main information about the financial product. One participant (P40), said she was interested in finding the kinds of basic information about payday loans she could use to explain the financial product to someone who knew little about it.

Table 32. *Codes for Gathering General Information*

Subnodes	No. Participants	No. References
Gather General Information	12	12
Definition for the Product	5	5
Basic Information	3	3
General or Main Information	3	3
Ways to Explain it to Someone Else	1	1

Another group of participants (N = 8) took a narrow approach to their search tasks by seeking specific kinds of information (code: *Find Specific Information)* (Table 33). These participants either explicitly stated they started the search by looking for answers to the guiding questions from the task scenarios (N = 4 of 7) or implicitly by saying they started by looking for information on how loan payments worked for student loans (N = 3 of 7). One participant (P01) started the search task on students loans by searching for the specific types of student loans. Another participant (P08) said he started his search task on student loans by looking for ways to refinance student loans into “cheaper options,” even though there was no indication in the task scenario that the student loans were costly.

Table 33. *Codes for Finding Specific Information*

Subnodes	No. Participants	No. References
Find Specific Information	8	9
Answer the task questions	7	7
Find cheaper options	1	1
Find loan types	1	1

A small group of participants' (N = 7) first statements about starting their search tasks were about the kinds of information they wished to avoid or exclude from their searching (code: *Employ Exclusion Criteria*) (Table 34). Four participants indicated they wanted to avoid lenders at this early stage of their search tasks. Each of the three search tasks was represented in that group. Other kinds of sites participants wished to exclude from their searching were commercial loan sites (N = 2), advertisements (N = 1), and ".coms" or commercial websites (N = 1).

Table 34. *Codes for Employing Exclusion Criteria*

Subnodes	No. Participants	No. References
Employ Exclusion Criteria	7	8
Avoid lenders at this point	4	4
Be wary of commercial loan sites	2	2
Avoid .coms at beginning of search	1	1
Avoid ads	1	1

The final subnode for how people began their search tasks has to do with participants (N = 2) who applied the task scenario to their own personal contexts (code: *Apply Task to Self*) (Table 35) as a way to interpret the task or determine what kind of advice to give their friend (i.e., for the payday loan scenario).

Table 35. *Codes for Applying Task to Self*

Subnodes	No. Participants	No. References
Apply Task to Self	2	2
Apply task to myself so I can interpret it	1	1
Put the task into my own context	1	1

5.4.1.2. Resources and sources tactics and strategies. The second largest subnode of coding for the *Finding Information* was the subnode coded *Resources and Sources Tactics and Strategies* (N = 38) (Table 36). This subnode described all of the websites participants mentioned by name as they talked about finding information, as well as the three main sources of information that they talked about – federal government, education, and non-profit.

Table 36. *Codes for Resources and Sources Tactics and Strategies*

Subnodes	No. Participants	No. References
B. Resources and Sources Tactics and Strategies	38	134
Websites mentioned	34	96
Federal Government Websites (.gov)	22	31
Education Websites (.edu)	4	4
Non-profit Websites (.org)	3	3

Across all stimulated recall interviews, 34 participants mentioned 44 different websites by name (code: *Websites mentioned)*. The top 10 websites by frequency of mention are shown in Table 37.

Participants who mentioned the Federal Trade Commission (FTC) website (N = 9) were searching on the payday loan task (N = 5 of 9) or the reverse mortgage task (N = 4 of 9). Some participants said they chose the FTC's website from search engine results because they expected to find certain kinds of information, such as legal information (P17 and P36), information about payday loan scams (P04), information about consumer protections for reverse mortgages (P36), or additional resources (P24). Some said they chose the FTC's website because they believed it was a "*reputable*", "*reliable*", or trustworthy resource. Another participant (P24) viewed the FTC as a kind of information quality benchmark for finding other sources: *"I was hoping that I would find more sites that were similar to the Federal Trade [Commission], something that had some weight, some recognition, some standing."*

The next most often mentioned websites were Wikipedia (N = 5), Nerdwallet (N = 5), and the Consumer Financial Protection Bureau (CFPB) (N = 5). Participants chose Wikipedia as a means to find other resources, either by finding them in the article references or in the editors' comments (P47); for its factual information (P04); to get ideas for searching (P03); or for figuring out the name for payday loans (P12 and P46).

Nerdwallet was a website that participants mentioned as one they had never heard of before (P13 and P48) or about which they were uncertain in terms of its trustworthiness (P14). One participant searching on the student loans task (P07) expressed a clear preference for information from a government site over the information on Nerdwallet: *"And this one, it's an official government site, so I wanted to look at that rather than, you know, Nerdwallet."* Other participants, however, found the Nerdwallet site useful, for information about student loans (P19) and payday loans (P48), such as when P48 said, *"I found this very relevant . . . I hadn't looked at Nerdwallet before."*

The CFPB website was used by several participants for finding information about student loan repayment scams (P07, P29). Another participant (P13) mentioned she did not like the format of the CFPB's website section about student loans because it required users to answer specific questions about their own student loans before providing general information. A participant searching on the reverse mortgage task (P10) mentioned that the CFPB had taken a reverse mortgage company to court.

The third most-often mentioned websites were of the AARP, the Yellow Pages, and the U.S. Department of Education (Dept of Ed). All of the participants who mentioned AARP (formerly known as the American Association of Retired Persons) were searching on the reverse mortgage task. One participant (P27) incorrectly associated Tom Selleck as the spokesperson for AARP and reverse mortgage (he is the spokesperson in reverse mortgage commercials sponsored by American Advisors Group, AAG). P33 said she chose to go to AARP's website "*because that's for older people, but then this website completely sucked. It had nothing*," while another participant (P36) said she went to the AARP site because AARP is "*looking out for people of a certain age*," and another (P43) also thought it would be a good resource but decided there were too many articles to go through and abandoned her effort. The Yellow Pages was mentioned by four participants who were also searching on the reverse mortgage task (P23, P26, P32, and P36). P26 decided to go to the Yellow Page website because she said *"I feel like my grandparents would trust the Yellow Pages."* P23 chose the Yellow Pages because of the title on the SERP even though she did not believe it was the best resource to answer the aspect of the task scenario about local reverse mortgage lenders in Nags Head, NC: *"So now I'm contemplating did I see anything that I thought might be more reliable than the Yellow Pages. But it had Nags Head in it's title so I felt like it would give me specific sources, like for something in Nags Head."* P32 and P36 both expressed the desire to see better or more detailed reviews on the Yellow Pages listings. The Department of Education website was selected by four participants searching on the student loans task (P07, P14, P19, P44). P07 said the Department of Education's website was more official than Nerdwallet's and P14 and P19 said it was "useful" and “reliable.” P44 went to the Department of Education’s student aid website at the very beginning of her task: “*So I read the task, which was about student loans, which I have a little of and my partner has a lot of . . . and I really don't like them . . . I know exactly which website to go to because we visit it frequently. That’s this Federal Student Aid website (https://studentaid.ed.gov/sa/).*” She also noted that the websites that show up on Google can be “tricky” because their titles are very similar to the one by the Department of Education, sometimes having .org domains in them.

The remaining websites in the top 10 mentioned were Wells Fargo, US News, and the U.S. Department of Housing and Urban Development (HUD). P10 went to the Wells Fargo site as part of a larger tactic to find information about reverse mortgages from large, national lenders. P31 participant went to the Wells Fargo site to search about payday loans because her sister banks there. P43 read reviews of lenders in the reverse mortgage space and spent some time reading about Wells Fargo.

Table 37. *Top 10 Websites Mentioned During Information Searching*

Subnodes	No. Participants	No. References
Websites Mentioned	34	96
FTC	9	11
Wikipedia	5	7
Nerdwallet	5	5
CFPB	5	5
Yellow Pages	4	4
Dept. of Education	4	4
AARP	4	5
Wells Fargo	3	3
US News	3	4
HUD	3	3

5.4.1.3. Finding and excluding tactics and strategies. Throughout the course of the search tasks, participants talked about a variety of criteria they used for finding information to complete the tasks as well as kind of information they sought to avoid during their search process (code: Finding and Excluding Tactics and Strategies) (Table 38).

Table 38. *Codes for Finding and Excluding Tactics and Strategies*

Subnodes	No. Participants	No. References
C. Finding and Excluding Tactics and Strategies	33	97
Criteria for Finding	28	64
Criteria for Excluding	23	33

In terms of participants' criteria for finding information, there were four main subnodes of coding (Table 39): criteria related to webpage content attributes, the type or kind of webpage content, the webpage source attributes, and the source types.

Table 39. *Codes for Finding Criteria*

Subnodes	No. Participants	No. References
Criteria for Finding	28	64
Content Attributes	6	6
Content Type or Kind	17	29
Source Attributes	4	6
Source Type	17	23

Participants (N = 6) mentioned attributes of webpage content that they were interested in finding (Table 40), including content in list format (N = 2), content that was current (N = 2), content that was neutral in terms of presenting the financial product in a positive or negative light (N = 1), and content the presented both positive and negative information about the financial product (N = 1).

Table 40. *Codes for Webpage Content Attributes*

Subnodes	No. Participants	No. References
Content Attributes	6	6
Content Format - list	2	2
Date or Currency Information	2	2
Neutral Information	1	1
Both sides of the argument	1	1

In terms of the type or kind of content on webpages, participants (N = 17) offered different kinds of examples (Table 41). Eight participants said they wanted to find pages with the pro's and con's of the financial product in their task, eight said they wanted to find basic or general information about the financial product, two were looking for alternatives to the financial product from their task, and one individual each said they were looking for the following: websites for seniors, the typical interest rate for the task financial product, information about loan payoff grants, legal information about the product, a person to speak with about the product, information that answered the question from the task scenarios, information on the page that would confirm the product was a scam, content on the page that would enable the person to eliminate that page as a source of information, a single answer that did not have a lot of options, and information the person knew from firsthand knowledge.

Table 41. *Codes for Types and Kinds of Webpage Content*

Subnodes	No. Participants	No. References
Content Type or Kind	17	29
Pro's and Con's	8	8
Find basic or general information about this product	8	9
Find product alternatives	2	2
Find websites for seniors	1	1
Find what normal interest rates are for this product	1	1
Look for loan payoff grants	1	1
Find legal information about the product	1	1
Find a human resource to talk to	1	1
Guiding questions	1	1
Confirmation this product is a scam	1	1
Content on page that will tell me I can eliminate it	1	1
Don't want options, want to talk grandfather out of it	1	1
Find what I know exists from firsthand experience	1	1

Another set of criteria for finding information for the search task had to do with attributes of the information sources, which typically meant the websites or the organizations that supported the websites (Table 42). Four participants mentioned the following attributes they were seeking in the sources of information they found: reputable, legitimate, trustworthy, and professional.

Table 42. *Codes for Source Attributes*

Subnodes	No. Participants	No. References
Source Attributes	4	6
Reputable	2	2
Legitimate	2	2
Trustworthy	1	1
Professional	1	1

Participants also said they were looking for specific types of sources (N = 17) (Table 43). This may have meant that the participant inspected the URL address to learn the source of the website. In other cases, participants would explicitly state they were looking for a specific website source, such as government websites (N = 9), non-profits (N = 2), or websites by educational organizations (N = 2). Participants also looked for website sources such as Wikipedia (N = 1), their own bank's website (N = 1), or their employer's website (N = 1). In one instance, a participant (P33) indicated that she was interested in finding only one source for the reverse mortgage task, not many, because she did not want to give the grandfather in the task too many options (i.e., she wished to talk him out of getting a reverse mortgage).

Table 43. *Codes for Source Types*

Subnodes	No. Participants	No. References
Source Type	17	23
Government website	9	10
Look at URL address	4	4
Domain - .org	2	2
Domain - .edu	2	2
Go to my bank's website	2	2
Look at my employer's website	1	1
Go to Wikipedia	1	1
Find only one source, not many	1	1

In terms of criteria for excluding information, participants (N = 23) identified types of webpage content and website sources they wished to avoid (Table 44). Many participants (N = 15 out of 23) expressed the desire to avoid advertisements (code: *Avoid or Ignore Ads*). An example of this sentiment came from P32 while she was searching on the reverse mortgage task, *"I definitely don't look at the ads because I know that they pay Google to put their name first [on the SERP]."* Another participant (P43) also searching on the reverse mortgage task, explained her reasoning for skipping ads on the SERP, *"Often times, I'll skip the first ones that have the advertisement with it just because it feels like it's people trying to sell things or get customers and so they may not give you both sides of the story because they would only give you what's beneficial to them. So, I typically scroll through the first couple that have ads."*

Some participants intentionally avoided bank and lender websites (N = 6) (code: *Avoid or Eliminate Banks or Lenders*). P03 implied this when she was explaining her searches on the payday loan task when she said, *"I was looking for a reputable source, or what I thought might be a reputable source. 'Cause a lot of them were banks."* Another participant (P16), searching on the reverse mortgage task, explained the reason she sought out government sources but avoided bank sources was because of paid content bias: *"So I'm looking for something from a government organization or someone who is presumably not paid by the lender."*

In some cases, participants lumped together their mental set of undesirable sites and information they wished to avoid, such as advertisements and lender websites. P36, searching on the reverse mortgage task, put it this way: *"I'm looking for what I consider legitimate sites. I'm looking for things that aren't ads obviously or that look like they're from a bank or from someplace that's going to redirect me to a mortgage service company or a broker."*

A few participants said they avoided "sketchy" websites (N = 3), such as P29 who said, "*I was just looking at some of the results and I didn't want an advertisement. I didn't want student hero. Student loan hero sounded a little sketchy maybe."* Other kinds of websites or information participants actively avoided included blogs (N = 1), content of "simple tricks" (N = 1), websites that were not familiar to the participant (N = 1), and Wikipedia (N = 1).

Table 44. *Codes for Excluding Information*

Subnodes	No. Participants	No. References
Criteria for Excluding	23	33
Avoid or ignore ads	15	20
Avoid or eliminate banks or lenders	6	6
Avoid sites that look sketchy or stupid	3	3
Avoid blogs	1	1
Avoid content showing simple tricks	1	1
Avoid sites I don't recognize the source	1	1
Avoid Wikipedia	1	1

5.4.1.4. SERP tactics and strategies. Participants shared about various tactics and strategies on the SERP (N = 15) (Table 45). The tactic with the most mentions was coded *Click the First One*, and stood for instances in which participants (N = 5) defaulted their search selections on the SERP to clicking on the first result. P14, searching on student loans, captured the vagueness of this tactic best: "*I clicked the first one 'cause I was like, I don't know what*." Other tactics were similar to some of the previously mentioned tactics that were used at the beginning of participants' search tasks, such as using language from the task scenario for the query language (N = 2), searching on the type of domain (N = 2), scanning the SERP for the pro's and con's of the financial products (N = 2), and others. The unique tactics in this category included instances when participants said they looked at the "related searches" area of the SERP for ideas on what to search (N = 2), broadening or narrowing the search queries (N = 1 each), scanning the titles (N = 2), snippets (N = 2), or SERP entity card (N = 1). One participant (P07) stated that he used only Google as his search engine and never used other search engines. This participant also mentioned hovering over results on the SERP and then looking in the lower left corner of the SERP to see where the page was going to re-direct to. Another (P27) talked about searching on Google to buy a boat with a reverse mortgage and this person was the only person of the 47 participants who actually searched using that aspect of the task scenario.

Table 45. *Codes for SERP Tactics and Strategies*

Subnodes	No. Participants	No. References
D. SERP Tactics and Strategies	15	29
Click the first result	5	6
Look at related searches on the SERP for ideas	2	2
Get query terms from the task language	2	2
Type domain into search box	2	2
Scan SERP for Pro's and Con's	2	2
Broaden the search	2	2
Look at snippet descriptions	2	2
Look at result titles	2	3
Click on each site and get information	1	1
Use Google search engine, not others	1	1
Narrow search	1	2
Look at bottom of page for URL re-direct	1	1
Look at entity card	1	1
Search to buy a boat via this product	1	2

5.4.1.5. Exploration tactics. The subnode *Exploration Tactics* covers unique search tactics participants engaged in during the search task that indicated exploration-types of behaviors (N = 10) (Table 46). There was one mention each of tactics such as looking for smaller banks or the town where the participant lived, looking for more information than what had already been found, reading all the way through articles to gain insight as to what to search for, searching for the risks of the task's financial product, searching for default rates of reverse mortgages or refinancing information for student loans, looking for personal finance websites, searching for lower interest rates, finding government programs for low-cost advising, looking for blogs, getting current interest rates from a bank, finding payment plans or procedures for payment plans, searching and not giving up until the participant found out the name of the loan, and searching for Annual Percentage Rate (APR).

Table 46. *Codes for Exploration Tactics*

Subnodes	No. Participants	No. References
E. Exploration Tactics	10	18
Find a lower interest rate	1	2
Look at smaller banks	1	1
Look at town where I live	1	1
Look for more information than I already found	1	1
Read article to figure out what to search for	1	1
What are the risks of the product	1	1
What are default rates for reverse mortgages	1	1
Search for refinancing information	1	1
Look for personal finance sites	1	1
Find government program for low-cost advising	1	1
Look for blogs	1	1
Go to a bank and get its interest rates	1	1
Find payment plans	1	1
Find procedure for making payments	1	1
Search until I find the name of this loan	1	1
Find out about APR	1	1

5.4.1.6. Querying tactics. Participants (N = 5) did not make a lot of statements about their querying process (code: Querying Tactics) (Table 47). Even though ILS researchers and practicing professionals place a great deal of emphasis on the querying process and the importance of being able to formulate effective queries for returning results, it seemed like developing queries was taken for granted by participants and not thoughtfully considered during the search process. Ways in which people queried including (all are N = 1 instances) matching words on a webpage to words in the task to get the correct name for the financial product, typing the product name into the search box plus the term *FAQ* to get frequently answered questions, searching the site of a large company about the topic, using the main point of the task for the query terms, using whatever knowledge the person had for the query, and using terms found on a webpage for the query. Though four of the five participants were conducting the payday loans task, there was too little data to make a meaningful interpretation of that fact.

Table 47. *Codes for Querying Tactics*

Subnodes	No. Participants	No. References
Querying Tactics	5	6
Match task words to website to get product name	1	1
Query – product name plus FAQ	1	1
Query a large well known company about this	1	1
Use main point of topic for the query	1	1
Use whatever knowledge I have for the query	1	1
Use words from webpage reading for new query	1	1

5.4.2. Strategies for evaluating information. The top-level node *Evaluating Information* was comprised of seven major subnodes, shown in Table 48. The seven subnodes consisted of an approach for starting the evaluation process by seeking to understand the financial product; evaluation strategies; three methods of website evaluation that related to attributes of the website, descriptions of the websites, or why the participant selected the website; comments about the relevance decisions participants made; and general viewpoints.

Table 48. *Code Categories of Strategies for "Evaluating Information" Node, by Number of Participants, Number of Uttered References*

Code Categories	No. Participants	No. References
1. Understand the financial product	27	54
2. Evaluation strategies	25	66
3a. Website selection	30	73
3b. Website descriptions	25	58
3c. Website attributes	13	22
4. Relevance decisions	26	50
5. General viewpoints	5	11

5.4.2.1. Starting the evaluation process. Some participants sought first to understand information about the financial product as they started their evaluation processes (N = 27). Table 49 shows the ways in which people sought to understand the financial product. Participants engaged in different kinds of tactics (N = 10), such as querying the product name to learn about it, like P02 did: *"I guess at first I just wasn't sure at all, so I was trying to think of something that I could search that would give a general overview of the topic."* In some cases, participants mentioned the tactic of seeking general background on the product as a common strategy they use, like P31, when she said, *"I just wanted more background information. What I usually do, I'll try to get the most basic information so I can at least have knowledge, like background knowledge."*

Table 49. *Codes for Starting the Evaluation Process: Understand the Financial Product*

Subnodes	No. Participants	No. References
1. Understand the financial product	27	54
Tactics	10	12
What people learned	9	14
Strategies	9	12

Other participants, rather than seeking to understand the financial product, jumped right into the evaluation process. There were numerous individual approaches. One participant used a specific search engine because the participant believed it gave the best quality results: (P07): "*Normally I will default to Google over the other search engines because I feel it does a better job. Just generally I think it's the way that it algorithms approaches searching. It gives you the highest quality hits to start."* Another participant started evaluating by searching for the least expensive ways to pay off student loans (P08) and used that as his evaluation criteria. Others indicated they evaluated based on the information they thought an individual might need, the different kinds of loans available, or based on what they thought they already knew about the product.

5.4.2.2. Evaluation strategies. After talking about how they started their evaluation processes, participants mentioned strategies they used for evaluating information. These are shown in Table 50. The largest subnode in this category was about how people decided on how much effort they were going to invest in the evaluation process (Pre-determine effort, N=12). Some participants (N = 4) said they decided to get more information after finding the first webpage or two. Other participants similar to these were the ones who described liking to read a lot of information when doing an evaluation task such as this (1), or reading through each category on pages to make sure they had evaluated all information (1). Other participants were the opposite and indicated they wanted to avoid long documents (1), not read through every page (1), use only one source (1), try to complete the task with one chart or table (1). Finally, one participant said that when there was only two minutes left, he wanted to read something entertaining. The next largest subnodes were those participants who indicated they would look for certain kinds of information (N = 11), such as reviews (2), article comments (2), multiple sources (3), different loan features (2) legal information (1), information in bold typeface (1), and related information (1). In terms of the next subnode, “look at specific webpage features,” these webpage features included the date or currency of the information, the first few sentence of the page, the bottom of the page, headings of articles, the URL that shows once the link is moused over, and other skimming behaviors. Some participants evaluated by confirming information, which was the fourth largest subnode (N = 5). In this case, participants confirmed information they already knew or information they had learned. Some participants talked about the importance of reading through webpages to be able to evaluate properly (N = 4). Others talked about comparing information such as similar products, the same products, or consistency across the product information they found (N = 4). Participants also sought out certain kinds of websites (N = 4) such as government websites (2) or well-known companies (2). Some participants said that in a real-life situation, there was information they would go back to, to evaluate (N = 1) or go back to and evaluate later because now it was too much effort (N = 1).

Table 50. *Codes for Evaluation Strategies*

Subnodes	No. Participants	No. References
Evaluation strategies	25	66
Pre-determine effort	12	17
Look for certain kinds of information	11	15
Look at specific webpage features	7	12
Confirm information	5	5
Read information	4	4
Compare information	4	4
Look for certain kinds of websites	4	7
Go back later to evaluate	1	1
Too much effort	1	1

5.4.2.3. Criteria participants used for evaluating websites. There were three categories of criteria people used when evaluating websites. Table 51 shows the first category, *website domain and content criteria,* which covers how people selected websites.

Table 51. *Codes for Website Domain and Content Criteria*

Subnodes	No. Participants	No. References
Website criteria – why participant clicked on it	30	73
It was .gov	23	43
It was .org	5	6
Wikipedia	4	4
Has pros and cons	4	4
Investopedia	1	1
It was about outer banks	1	1
I knew this site (AARP, Forbes)	2	2

In combination with the 23 participants who identified government websites (.govs) in their criteria for selecting a website, of the 44 participants in the interview sample, 38 (86%)

mentioned U.S. government websites and government-sponsored information as some part of their information searching and evaluation processes. When talking about their evaluation behaviors, a number of participants expressed the need to learn general information and background about the financial product in the task. When selecting the kinds of websites that could provide the best information for completing their tasks, participants identified government websites and information as reliable, trustworthy, and unbiased, such as one participant who said, "*I look down (the SERP) and typically I'm trying to find government sources and skip the commercial sources. Just because I want unbiased sources and not, you know, somebody that has an interest in a slant on the information"* (P35). Adjectives used to describe government websites and information are shown in Table 52, in order by the number of participants who mentioned each. Participants valued the government as a source of credible, reliable, trustworthy information that would provide relevant information on its websites, especially for personal finance topics such as when P32 said, " . . . *I know that with financial stuff, the government websites are really helpful*".

Table 52. *Word Describing Government Websites and Information*

Adjective	No. Participants
Reputable	9
Reliable	6
Trustworthy	5
Useful	4
Legitimate	2
Neutral	2
Official	2
Professional	2
Unbiased	2
Comprehensive	1
Credible	1

Table 53 shows the second category of website quality criteria. Participants described websites as “bad” for numerous reasons, including: the site asks questions but does not give information, does not have any numbers, has inaccuracies, has a calculator, is an online book that my friend might not be able to access, it was a scam-type site, it was too long, looks like an infomercial, had popups, and was bland. Specific sites that were mentioned included the FTC, National Coalition on Aging (NCOA), NerdWallet, Yellow Pages, Investopedia, and Huffington Post. Sites that were described as “good” included those that answered a lot of the participants’ questions, gave topic overviews, talked about advantages of the product, or were .coms. One participant (P21) commented that .coms were useful when trying to find out negatively-skewed information because .edu sites “prefer not to say anything negative against all these financial programs.” NerdWallet was described as reliable (P19) and .orgs and .govs were described as reputable and also trustworthy. Sites that did not have ads were also described as trustworthy (P13). One participant (P22) described Reddit.com as trustworthy. Sites described as not trustworthy were .coms and those that displayed ads. Sites that had a lot of information, charts, and downside risks were described as useful.

Table 53. *Codes for Website Quality Criteria*

Subnodes	No. Participants	No. References
Website quality descriptions	25	58
Bad	9	15
Specific sites mentioned	7	7
Good	6	7
Reliable	3	3
Trustworthy	3	3
Not trustworthy	3	3
Useful	3	3
Helpful	2	3
Layout	2	2
All others	5	9

Table 54 shows the content and layout features that participants paid attention to during their evaluation processes. In terms of content, participants described evaluating websites based on the product features in the content, personal stories, and using Tom Selleck as the reverse mortgage spokesperson. Several participants (N = 3) commented about page layouts that had too many pictures or too much information or pages that they liked because the page did not require a lot of scrolling. Headings were mentioned in reference to participants who skimmed headings to make evaluations. This was also the case with the code “first few sentences”. Other attributes that participants talked about using for their evaluations were sites with FAQs, charts and tables, and those with comments, white space, and other individual characteristics.

Table 54. *Codes for Website Attributes*

Subnodes	No. Participants	No. References
Content and layout features	13	22
Content	4	4
Page layout	3	3
Headings	2	2
First few sentences	2	2
FAQs	2	2
Charts ad Tables	2	2
Comments, white space, and other codes	5	7

Participants prioritized government websites and information sources in their relevance criteria over other sites and sources. Some participants' statements clearly indicated this prioritization, such as: "*Well, **the first reason** I thought it was relevant was it was a government resource*" (P22) and "*. . . **the first thing** I clicked on was a government website. I feel like this is a reliable resource"* (P26). In other cases, participants used their previous knowledge about specific government agencies to predict the quality of information on a website. For example, when the FTC website showed in the screen recording, P25 said he knew that it was "*definitely going to be pretty reliable, like better than Credit.com or PaydayLoanInfo.org."*

5.4.2.4. Reasons for relevance decisions. Participants gave a variety of reasons for their choices of the four relevance grades. The codes are shown in Table 55. The most common reason for grading a website *very relevant* was a tie between a website having a lot of information and the website being the FTC. Other reasons included: there was a list of lenders, it was a local resource, it had charts, it was an easy to understand format, it was a .gov, it answered the task questions, and it was the CFPB. Reasons given for making a page *relevant* included that it was from the government, it was personally relevant to the participant, a friend had a good experience with the company, it related to the task questions, it covered different issues on the topic, and it talked about the risks of the product. In terms of the *somewhat relevant* decisions, participants said that reasons included because the website was at the top of the SERP, it had a lot of links, or because it did not exactly match the task. For those webpages graded *not relevant,* participants included reasons such as they had changed their mind about what they wanted, there was too much information on the site, there were too many ads, there was a calculator, it had legal information, the website was too long, the content was on another site read earlier, or it was not related to advantages and disadvantages. Several participants commented that certain websites would have been more relevant if they had had more information, had better categories and sections, or if they were a .gov site.

Table 55. *Codes for Relevance Decisions*

Subnodes	No. Participants	No. References
Relevance decisions	26	50
Very relevant	13	14
Relevant	11	15
Not relevant	7	9
Somewhat relevant	4	4
Would have been more relevant if . . .	2	5
Attributes of relevance	2	2

5.4.2.5. General viewpoints. During the stimulated recall, participants also expressed different viewpoints that were evaluative in nature and so those are included here (Table 56). These were general comments such as "companies can be paid to say anything" or comments about the government information being unbiased and regulated.

Table 56. *Codes for General Viewpoints*

Subnodes	No. Participants	No. References
General viewpoints	5	11
.coms provide useful information	1	2
Companies can be paid to say anything	1	2
Companies selling stuff are useless	1	1
Federal sites don't revenue off of clicks	1	1
Government is not biased	1	1
Government is regulated	1	1
Payday loan are too much interest and fees	1	1
Sites that allow comments are biased	1	1
.nets are better than .coms	1	1

CHAPTER 6: DISCUSSION AND IMPLICATIONS

In this section the findings of the study are discussed with respect to insights about the outcomes, implications of the findings, and limitations of the study design, with ideas on improvements in future study designs throughout the text.

6.1. Discussion of Findings Related to the Models and Hypotheses (RQ1)

The data analysis indicated very little about the influence of the independent variables on search performance, relevance assessment, and mental workload. In this subsection, the models are discussed in the order they were proposed and studied.

6.1.1. Perceptual speed model. There were three hypotheses for the perceptual speed model. This section discusses the main findings and insights about them.

6.1.1.1. Hypothesis #1. In the first hypothesis, it was expected that participants with higher perceptual speed would interact more with the search system while searching, as evidenced by issuing longer queries, have more SERP clicks, viewing more unique URLs per query, and viewing more unique URLs overall per task. The outcome was that there was no difference. This hypothesis was based on findings from the literature, in particular those from a previous study's (Brennan et al., 2014)'s significant results that had large effect sizes.

There could be numerous reasons for the difference in findings between the two studies. For example, the structure and lengths of the search tasks in the two studies were different. The average session length in Brennan et al. (2014) varied widely across the three tasks, from 3.8 minutes on average (SD = 2.5) for the least complex task to 9.1 minutes (SD = 3.2) for the midlevel complexity task ("analyze") which was also the task that participants recorded the highest mental workload for, and 7.0 minutes (SD = 3.2) for the highest level of complexity ("create"). By adding the three averages, the total average time that participants searched in the study was about 20 minutes. In the current study, the task complexity was kept constant across the three tasks and the time participants spent on each task was longer (Table 57).

Table 57. *Means and Standard Deviations for Time Spent on Tasks in Minutes*

	Task Topic			
	Reverse Mortgage	Payday Loan	Student Loan	Total
Task Time	11.2 (2.0)	10.0 (2.9)	10.8 (2.3)	10.7 (2.4)

Participants in the current study spent more time on all tasks and also spent more time overall on the search session, that is, they spent 32 minutes overall versus the previous study of 20 minutes. This difference between the two studies may be important for several reasons. First, it may be the case that participants with higher perceptual speed may interact more quickly "out of the gate" so to speak, at the onset of the task, than participants with lower perceptual speed and then slow down in their interactions as the task time grows. While this is conjecture, at least one study found variances in how strongly cognitive abilities affect users' activities at different stages in visualization tasks (Steichen et al., 2013). Cognitive abilities may vary in their influence on users depending on different factors or they may vary in the strength of their affect on different users based on their ability levels or some other factor. For instance, it may be the case the users with lower perceptual speed need "warm up time" to get started on tasks and then speed up as the task progresses. Whether participants with higher perceptual speed slow down or those with lower perceptual speed get faster, it may simply be that with longer length search tasks such as those from this study, the interaction differences between high and low groups smooth out over time and this may be a natural phenomenon. One way to test this idea would be to design a study in which the volume of search interactions (i.e., clicks, queries, URL visits) are captured at different fixed intervals during the search sessions and then compared.

Another possible reason the results of the studies are so different may have to do with task complexity. In Brennan et al. (2014), significant main effects were found for task complexity on search interactions and also significant main effects for perceptual speed on search interactions. It may be that varying the level of task complexity is a stimulus for evoking difference in perceptual speed ability that do not occur when task complexity is kept constant.

6.1.1.2. Hypothesis #2a and b. The second hypothesis was about relevance assessments and had two parts. For Hypothesis 2A, it was expected that participants with higher perceptual speed would bookmark their first relevant pages faster. I used a proxy for *time to first relevant document* because there was no reasonable way to calculate this measure for the study. The proxy was *time to first click*. The findings were not significant and were opposite of the direction proposed in the hypothesis. Despite this, the differences in time to first click were fairly large between the two groups and seem to be worth noting. Participants with higher perceptual speed took more time across all three tasks than those with lower perceptual speed. The amount of extra time ranged from about 7.5 seconds longer on the reverse mortgage task, to 11 seconds longer (which was twice as long) on the payday loan task, to 16.5 seconds longer (more than three times longer) on the student loan task. This may have been an indicator that the higher ability participants were being more careful in selecting which SERP links to click which resulted in the longer first click times.

One of the studies this hypothesis was based on was Al-Maskari and Sanderson (2011). There are several differences to note between the two studies. In Al-Maskari and Sanderson

(2011), the researchers found that the correlation between the *Finding A's* test (the same one used in this study) for their N=56 users was not significant for the *Time to First Relevance Document* measure. Upon finding this, the researchers combined participants' scores for the Finding A's test with the two other Perceptual Speed tests in the Ekstrom Kit and a created a new measure which they called "Overall Perceptual Speed". This measure was used for their Mann-Whitney tests and the subsequent significant finding for the *Time to First Relevant Document* measure. In the current study, the *Finding A's* test was also used, but the tests were scored according to the instructions in the Ekstrom Kit Manual (Ekstrom et al., 1976b) and not re-calculated any further. Thus, there is no way to accurately compare the scores of the studies.

Hypothesis 2b was also about relevance and this was related to precision. This hypothesis was based on Allen (1994), who found that participants with higher perceptual speed achieved greater precision and recall in an experimental system. The finding was in the direction of the stated hypothesis; participants with higher perceptual speed ability scored higher IUP (M = .362, SD = .141) than those with lower ability (M = .349, SD = .111). The difference between the two groups was not statistically significant and it was also not large, practically speaking. It is likely that this result stemmed from the lack of high agreement between the two expert assessors, which made achieving high IUP very unlikely. It is also likely, given the trend in other findings, that the topic of debt-related information searching was equally challenging for all participants and so they all performed similarly. Figure 18 shows the comparison of the participants' and experts' relevance grades of the webpages selected by participants. Participants marked most of their documents relevant or very relevant, which is not surprising, given the instructions they were given, which were to find the best webpages about each topic (Appendix K). A form of the Hawthorne effect may have influenced participants' relevance scoring; because participants knew the instructions were to find the best webpages, they may have been more likely to believe that the webpages they found were the best ones because they wanted to do the best job while being observed in a research setting. It is also possible that having a high level of trust in Google may have biased participants to grade webpages as better than the webpages actually were. Evidence supporting this is that when responding to the post-task questionnaire on the system's ability to retrieve relevant documents, the average score was 4.3 on a 5-point scale and this was higher than participants rated their own abilities to find relevant webpages, which was 3.9 on the same scale. The qualitative findings also pointed to evidence of participants' trust in Google's ability to retrieve the most relevant results.

While participants graded most webpages relevant or very relevant, the experts graded most pages somewhat relevant or not relevant (Figure 18). Even though the assessors had lower agreement than expected, there was still enough consensus between them that indicated an overall lack of quality in most of the webpages selected and graded by participants during the search sessions. This bears strong implications for information searching in the personal finance domain. Participants believed they were successful in finding relevant and very relevant webpages (i.e., indicated by the post-task questionnaire score of 3.9/5.0) and that the system performed well (post-task score of 4.3/5.0), but according to topic experts, the webpages were less than relevant in many cases.

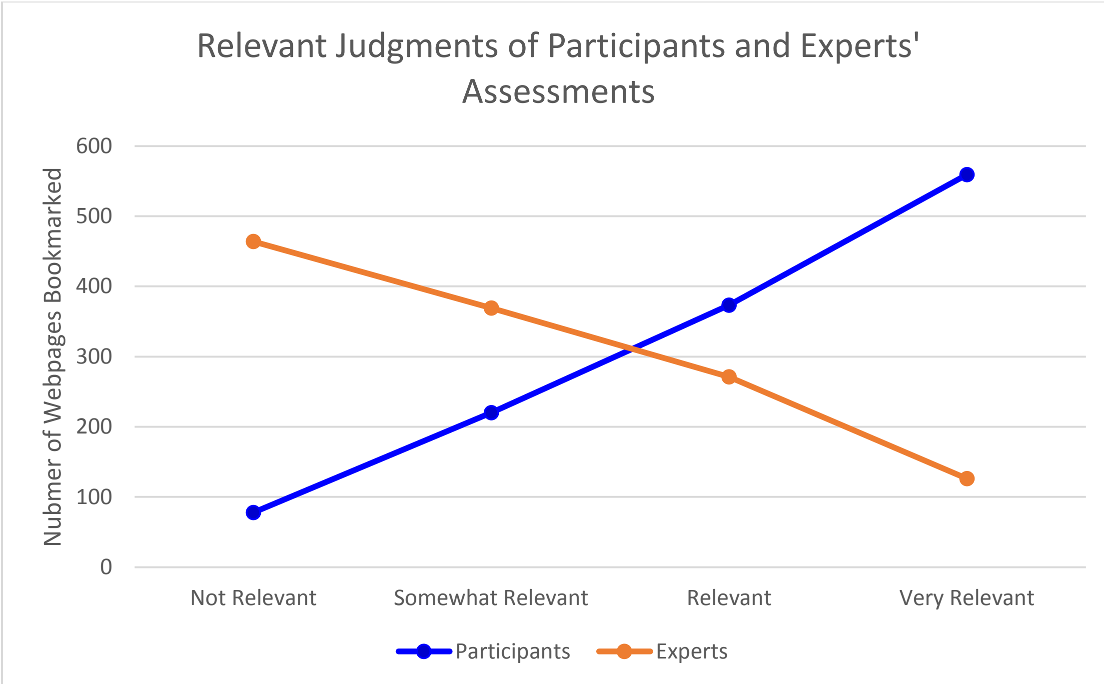


Figure 18. Participants' and experts' relevance judgments compared.

6.1.1.3. Hypothesis #3. The third hypothesis dealt with the impact of perceptual speed on mental workload. Mental workload was operationalized with responses to the NASA-TLX mental workload questionnaire and eye fixation durations. In terms of the NASA-TLX, participants with lower perceptual speed reported greater levels of mental workload (M = 4.2, SD = 1.0) than those with higher perceptual speed (M = 3.9, SD = .84) but the difference was not significant. This hypothesis was based in part on Brennan et al. (2014) in which participants with lower perceptual speed reported greater mental workload than those with higher perceptual speed. There could be a variety of factors for why there was not much difference in the two groups of participants in the current study. It could be that the tasks themselves did not incur much mental workload across all participants. It could also be that the tasks were complex and the topics created a lot of uncertainty in participants, which created a floor effect on their behaviors. One way to find out about differences in mental workload in personal finance might be to vary the degree of complexity of the tasks to see if there are task differences across the two groups. Another way to look at this might also be to look at everyone's last search task only to see the differences in mental workload or to look at the last half of each search task, when participants would be presumably more taxed than in the beginning. However, this may not hold for unfamiliar tasks in which people spend more time learning during the beginning of the task and have a greater understanding during the second half which may lead to greater engagement and a lessened experience of mental workload.

The mean fixation duration measure and standard deviation of fixations durations were not significant, however both measures were in the direction of the stated hypothesis and it could be that the effect of perceptual speed on participants' mental workload was not strong enough for an effect to be detected in the mean fixation duration or again, that a floor effect was created for all participants, which made differences undetectable. It could also be that by measuring fixation durations at the task level made it such that the strength of the effects were muted and so a possible future analysis could be to measure fixation durations at the webpage level.

6.1.2. Working memory model. There were two hypotheses and one exploratory measure for the working memory model.

6.1.2.1. Hypothesis #4a and b. Hypothesis 4(a) stated that participants with higher working memory ability would issue more unique queries and open more unique webpages than those with lower working memory ability**.** The outcome of this measure was the opposite of the hypothesis. Participants with higher working memory ability issued fewer unique queries on average per task (M = 5.3, SD = 2.2) than those with lower working memory ability (M = 6.4, SD = 3.0), and opened fewer unique webpages (M = 10.7, SD = 4) than those with lower abilities (M = 10.9, SD = 3). The independent samples t-test indicated that these differences were not significant. The differences between the groups were small, particularly in the case of the unique webpages measures, where the difference between the two groups is less than one webpage. This hypothesis was based on the work of Gwizdka (2013) who found that in an experimental system, users with lower working memory selected fewer word tags and opened fewer documents than those with higher working memory (the act of selecting and de-selecting tags was considered a proxy for querying behaviors). Even though this was the same psychometric test that was used in Gwizdka (2013), it may be that the measure differences do not translate when going from an experimental system to the open web for the search system.

It could reasonably be said that there were actually no differences at all across my participants, given how close the results numbers were. It is possible that, similar to the finding from Hypothesis #1, any possible differences between the two groups were erased by other effects of task complexity, task topic, or other factors. It may also be the case that this psychometric variable does not impact how many webpages a participant selects in a relevance-type task like the ones in this study.

Hypothesis 4(b) investigated the effect of working memory on search behaviors by measuring eye movements. Participants with lower working memory were expected to have larger fixation counts on average per SERP and per webpage as well as longer average lengths of fixation durations per task than those with higher working memory, but that was not the case for any of the measures. This hypothesis was based on readings from Holmqvist and Nystrom (2011), Meghanathan et al. (2015), and Peterson et al. (2008). It made sense to hypothesize that those participants with lower working memory would have longer average fixation durations because their memory load would be greater than those with higher working memory. Based on the data from the dissertation study, this was not the case. Participants with higher working memory ended up having larger fixation counts on average and longer average fixation durations, though the differences between the averages of the two groups was small. This may mean that all participants experienced equivalent amounts of memory load. This may be an indicator that the task topic or the complexity of the task or a combination of both of those factors was so challenging for all participants that there was no way to differentiate between the groups.

There are numerous elements about the current study that differ from the studies upon which this hypothesis was based. First, as previously mentioned, the measurement of eye gaze data at the task level rather than the webpage level may have been too high of an aggregate to capture participants' individual behavioral differences. Second, the tasks in the studies upon which this hypothesis was based were visual search tasks, not information search tasks and this may have mattered because information search tasks are more visually complex than those encountered in experimental psychology studies:

> From an eye monitoring perspective, information retrieval seems to encompass both a visual search scenario as well as reading, so it is expected that the average fixation duration will fall within the range of these two groups . . . It is necessary for the eye to move rapidly during reading, while in visual search and scene viewing, it is less imperative that the eye quickly scans the entire scene, but rather that user can absorb key information from certain regions (Granka, Feusner, & Lorigo, 2008, pp. 348-349).

6.1.2.2. Hypothesis #5. The next part of the working memory model dealt with relevance assessments. In this hypothesis it was expected that similar to MacFarlane et al. (2012), participants with lower working memory would grade fewer webpages to be *not relevant.* The findings of the study were in line with the hypothesis. Participants with lower working memory bookmarked fewer pages as *not relevant* than those with higher working memory, though the result was not statistically significant. In addition, the difference between the groups was about one bookmark; that is, the lower group bookmarked one less page as *not relevant* than the higher group. Whether or not this difference is meaningful is inconclusive.

However, further exploration of this finding could include a more controlled design for evaluating webpages, such as using a closed corpus of documents of varied levels of expert-determined relevance, along with a pre-screened pool of participants belonging to lower and higher working memory ability groups.

6.1.2.3. Exploratory measure #1. The final aspect of the working memory model was Exploratory Measure #1, that explored a potential influence of working memory on mental workload. Two measures were explored. The first was the relationship of working memory to the mental workload score from the NASA-TLX questionnaire. Participants with lower memory reported overall greater mental workload, though the result was not significant. The second measure was to see if there was a difference in the amount of time participants spent on SERPs based on their working memory levels. This result showed that across the tasks, participants with lower working memory spent, on average, an additional 1 minute, 22 seconds on SERPs than their peers with higher working memory. This result was also not statistically significant. The finding was similar to the findings of Brennan et al. (2014), in which participants with lower abilities for memory, perceptual speed, and visualization spent more time on SERPs during all tasks. This finding may be similar to Gwizdka (2017), who found that in the last phase of search tasks, searchers with lower working memory increased the amount of time they spent reading SERPs, as evidenced by longer *reading fixation durations*. While in the current study it was not possible to determine reading time through the measure *reading fixation durations* as in Gwizdka (2017), it may still be interesting to conduct further analysis of fixation durations of high and low working memory groups for some appropriately specified final time segment of the search tasks to see if there is evidence of increased fixation durations on SERPs for the participants in the low working memory group.

Overall, the combination of these findings, while not statistically significant, suggests that there may be a link between working memory and mental workload. Participants with lower working memory may spend more time on SERPs to either remember where they are going or to go back and review the results again. It may be the case that the indicators for either working memory or for mental workload are not sensitive enough to detect the subtle differences that may impact users at different levels. Beyond finding a significant effect it would also be important to determine if significant effects of working memory are actually meaningful in real-life terms, such as accuracy, satisfaction, and other factors.

6.1.3. Financial knowledge model. There were two hypotheses and one exploratory measure for the financial knowledge model.

6.1.3.1. Hypothesis #6. The sixth hypothesis dealt with the influence financial knowledge might have on searching behaviors. There has not been any research in IIR that specifically investigates the search behaviors of users based on their empirically tested financial knowledge and so this hypothesis was based on a large-scale log study by White et al. (2009), studies of other kinds of domain knowledge (Freund & Toms, 2006; Hembrooke et al., 2005; Zhang et al., 2005), and a secondary data analysis of consumers in Germany (Berger & Messerschmidt, 2009). Apart from the German study, which indicated that consumers with greater financial knowledge search more extensively, the remaining studies all had findings related to differences in language-related behaviors, such as length and number of users' queries and vocabulary differences in query terms. The hypothesis for my study stated that participants with higher scores on the financial literacy test would issue longer and more queries than those with lower scores on the test, but the results showed that there was no difference between the groups on these two measures.

The lack of meaningful differences in queries could be related to the fact that the participants in the current study were not considered experts in the subject domain, whereas in several of the studies mentioned, participants were identified as having ‘expertise’ in a particular domain based on being students who were majoring in that subject matter (Zhang et al., 2005) or professionals who were working in their area of expertise (Freund & Toms, 2006). In another study (Hembrooke et al., 2005), participants selected search topics in which they felt they were experts. It could be that using an objective test of knowledge such as the one in this study does not effectively identify the level of expertise in financial matters that would make a difference in people’s search behaviors.

6.1.3.2. Hypothesis #7. The seventh hypothesis dealt with the influence that financial knowledge had on relevance behaviors. The studies that this hypothesis was based on were mentioned previously in Hypothesis #6 and all had findings related to webpage behavior, such as the number of webpages visited or bookmarked. Again, the findings of those studies were that those with greater domain expertise viewed or selected more webpages than those with less domain expertise. In the dissertation study, participants with higher financial knowledge bookmarked about one additional webpage per task more than those with lower financial knowledge and this finding was not significant. Similar to Hypothesis 6, it could be that the delineation of participants as having high and low levels of a certain kind of financial knowledge test did not adequately capture the domain knowledge that results in meaningful behavioral differences in users.

6.1.3.3. Exploratory measure #2. Exploratory measure #2 investigated the possible influence of financial knowledge on mental workload using several different dependent variables. The variables were scores on the NASA-TLX mental workload questionnaire, scores on the post-task difficulty question, and IUP scores. Each independent sample *t*-test revealed similar findings to previous hypotheses – there were very small, if any, differences in the means for these measures between the high and low financial knowledge groups. It seems intuitive that having greater domain knowledge should reduce a person's experienced mental workload on a given task, however the operationalization of domain knowledge in this study may not have been valid.

While there has been a lot of study on domain expertise, there have been no studies of directly-tested financial knowledge in information science or other disciplines. It seemed that the arithmetically-derived grouping, "high knowledge," did not accurately operationalize real-world expertise in the personal finance subject domain. An alternative approach could be to compare information searching by people who are financial hobbyists with those who do not regularly search on financial topics.

6.1.4. Interaction model for perceptual speed and financial knowledge. There were two hypotheses and one exploratory measure for the financial knowledge model.

6.1.4.1. Hypothesis #8a, b, and c. The eighth hypotheses was about the order of effects in which participants with different levels of perceptual speed and financial knowledge would issue more queries than other combinations of different levels. None of the hypotheses were supported and the direction of the findings was inconclusive. This is not surprising given the lack of clear findings from earlier hypotheses for these independent and dependent variables.

6.1.4.2. Hypothesis #9a and b. Hypothesis 9 addressed the interaction model in relationship to relevance assessments. Similar to Hypothesis 8, there were findings were inconclusive, which is likely the result of the lack of clear findings about the individual measures in other analyses.

6.1.4.3. Exploratory measure #3. No exploratory measures were calculated because there was not enough evidence to inform the direction which any effects might take.

6.1.5. Interaction model for working memory and financial knowledge. This model is undefined and the exploratory measures were not undertaken, as a result of non significant findings on the two independent variables in earlier hypothesis. Without compelling evidence, whether statistically significant or practically meaningful, it did not make sense to perform statistical analysis. Until there is better direction in the future from evidence of either or both of the two IV's having an impact on mental workload, this model will remain un-investigated.

6.1.6. Additional data analysis techniques. Upon reviewing the initial data analysis results with so many non-significant findings, I conducted various other types of data analysis to determine if there might be other possible explanations or answers. Partial correlations were calculated, but since the correlations of the IVs and DVs were very low and not significant, it was not surprising that the partial correlations did not yield any new information either about the relationships among the variables. Regression analyses were run on the hypotheses but also did not produce significant nor meaningful results. Finally, several calculations of multi-level models were conducted, but these also did not produce results showing significant relationships, and given that the multi-level models used the measures at the task level, this approach had the same limitation as the original analysis that used task level data.

6.2. Discussion of Findings Related to Qualitative Analysis (RQ2)

There were many findings in the qualitative portion of the study that are similar to other studies that have investigated users' searching and assessing behaviors. For example, participants spoke about strategies for assessing webpages that incorporated the content, structure, and quality of webpages, similar to the findings from Tombros et al. (2005). Others commented on design elements of webpages, which was also a finding in Xie et al. (2010) and Xie and Benoit (2013). While it is confirming to see these similarities, the value of this dissertation study lies more in the unique findings related to searching and evaluating strategies as they relate to debt-related task topics in the personal finance domain. Therefore, this section discusses the highlights of those findings.

6.2.1. Finding information. Participants used different strategies for starting their search tasks and also for finding information.

Similar to findings from Thatcher (2006), participants had clear strategies for starting their searches. What was noticeably absent from this data, however, was any kind of serious concern on the part of participants when it came to entering queries into the search system. Developing queries seemed to be a low-cost activity for participants. They were quick to use whatever language on the screen seemed best to them for their queries, whether that was from the search task description, the SERP snippets, or the section titles of webpage documents. Participants typed in many queries and they did so quickly, often in lieu of scanning further down on the SERP. Many used the query auto-completion feature of Google. It may be that the combination of being able to type quickly, use the query auto-completion feature, or click on the "did you mean --?" spell-correct feature of Google has made querying such a low-cost activity for participants that it is faster and easier for them to type a new query to get new results to appear at the top of the SERP than it is to review all the links down the page on the SERP. For exploratory search tasks, querying behaviors such as these which emphasize precision over recall will be less effective for users (Marchionini, 1995). Optimal recall-oriented behaviors involve spending more time developing queries and viewing deeper-level links on SERPs. For example, Azzopardi et al. (2013) tested query interface designs and SERP-viewing behaviors and found that when users spent more time on querying, they investigated results farther down on the SERP. Qvarfordt, Golovchinsky, Dunnigan, and Agapie (2013) found that a novel query widget enabled users to find more useful documents by searching further down on the SERP. It may be that the design of the ubiquitous small, rectangle search box forces users to engage in precision-optimized search behaviors that sabotage their goals in exploratory search tasks. It has also been hypothesized that fast "rapid-fire" querying and shallow examination of SERPs might be associated with stress (Edwards et al., 2015) and while that premise has not yet been substantiated, the qualitative findings in this dissertation suggest that when users search and evaluate unfamiliar, complex topics in personal finance, the uncertainty (or perhaps stress) of the whole environment may lead to this kind of excessive querying behavior that is less effective for exploratory tasks that are more cognitively demanding and require more patient searching behavior.

Many participants started out their searches by looking for general information on the tasks topics. If they were uncertain about the product in the task scenario, they searched for product definitions, overviews, and basic information before addressing the issues in the task scenario. They often articulated the desire to avoid advertisements and commercial websites. They looked for websites they perceived to be "neutral" and "unbiased," which most often meant websites run by the federal government. The implication about this manner in which many participants attributed neutrality and lack of bias to federal government websites is that citizens put high trust in the government to provide them with reliable information about financial topics with which they are unfamiliar.

Another finding that is quite clear is how much participants trust Google. Even though participants were instructed to use any search engine they wished to, no one switched from using

Google. Some participants explicitly commented about trusting Google while others implied the trust that the search engine would return best results at the top of the SERP (but under the advertisements). Even when participants found very relevant information on websites run by the FTC, CFPB, or Department of Education, instead of using the search features on those websites to search for more information, they opted to leave those websites and go back to Google to search. This seems to be partly driven by the desire to find different sources of information from different entities. Participants valued diversifying their information sources, with many talking about the need to view multiple sources before deciding which information was the best or whether they could trust information they found on some of the websites.

Later in the search tasks, participants who previously had avoided lenders and banks on the SERP began to look for information from these kinds of websites. When this happened, the lenders, banks, and credit unions they sought out were ones where they held their own accounts or where family members held accounts.

Another finding was about participants who talked about exploring behaviors. Exploring took place once participants had developed a basic set of knowledge about the topic and included strategies to find out more about the risks of the products, the current offerings available for the products, and searching for ramifications of the products such as not paying off the loans on time. The pattern of developing a knowledge base and then diving into deeper areas of detail about products seems to fit into the framework of *searching as learning* (Eickhoff, Gwizdka, Hauff, & He, 2017).

In summary, when participants searched on financial topics with which they were less familiar, they stuck to basic strategies and tactics which meant a search that looked something like this: start with Google, avoid ads, define the topic, see what the government says about it, keep using Google to get more information, and once enough learning has taken place, dig deeper in more specific (i.e., commercial) places. Further investigation into this phenomenon could also take into account variables such as topic uncertainty or search uncertainty.

6.2.2. Evaluating information. Participants described basic strategies for evaluating information.

Participants decided ahead of time how much effort they were going to put into the evaluation process. In some cases, the effort had to do with how much reading they were planning to do or the lengths of documents they were willing to look at. One of the prominent themes in participants' comments was that of seeking information to ground themselves in the topic, either by looking for definitions or getting some kind of background information. This may be a common strategy for users in exploratory search tasks where users' uncertainty is high. If so, this has implications for systems design to create space for fundamental information gathering. The advent of SERP entity cards partially addresses this, but they are limited by the extent to which they reflect the searcher's topic (Bota et al., 2016).

This theme of fundamental information gathering also may have had implications in terms of how participants evaluated information when deciding on which level of relevance assessment to assign a bookmark. For those participants who sought background information on the financial products, webpages that educated them may have manifested in a kind of cognitive relevance that influenced those participants to evaluate those pages by giving them bookmarks at higher relevance grades than participants who did not need or seek background information. In addition, experts were instructed to evaluate the information on the webpages by focusing on the quality of the information in general, which meant making their relevance assessments from a domain relevance perspective, which could explain why the experts' relevance assessments were so different from participants'. This is a common challenge in trying to collect gold standard relevance judgments that are used to evaluate the information that laypeople find.

In order to evaluate information to make relevance assessments, participants described user-defined relevance criteria such as page layout, sections titles, and so forth that are the same as those from the literature (summarized earlier in Figure 2 – "Compilation of user-defined relevance criteria,"). For example, participants pointed out features of webpages related to layout (Balatsoukas & Ruthven, 2012), such as the amount of whitespace and use of stock photography.

Participants at times also looked to more subjective information about the financial topics, in the form of reviews, article comments, and even posts on https://www.reddit.com/. This seemed partly to be a means for people to find out more than just basic information about financial products that may be prone to unscrupulous practitioners. This finding also has implications for financial literacy educators to be able to teach consumers how to distinguish opinion from fact.

In addition, the website domain was an indicator of the relevance of the information and this mostly applied to the .gov and .org domains as being more relevant than others. Both the FTC and CFPB were mentioned numerous times as websites that had credible, relevant information.

Interestingly, even though 43% of participants held mortgages on homes they owned at the time of the study, none of the participants in the stimulated recall for the reverse mortgage lending task indicated that they sought information from a lender they used personally. A future study could investigate the extent to which people seek general information from companies they use versus general resources on the Internet.

Participants prioritized government websites and information sources in their relevance criteria over other sites and sources. Some participants' statements clearly indicated this prioritization, such as: "*Well,* ***<u>the first reason</u>*** *I thought it was relevant was it was a government resource*" (P22) and "*. . .* ***<u>the first thing</u>*** *I clicked on was a government website. I feel like this is a reliable resource"* (P26). In other cases, participants used their previous knowledge about specific government agencies to predict the quality of information on a website. For example, when the FTC website showed in the screen recording, P25 said he knew that it was "*definitely going to be pretty reliable, like better than Credit.com or PaydayLoanInfo.org."*

Participants made statements and comments about government websites that reflected possible beliefs about how the federal government conducts website marketing differently than commercial websites. For instance, one participant said *"they don't gain revenue off of clicks,"* (P21) and another surmised that government organizations are *"presumably not paid by the lender"* to create website content about reverse mortgages (P16). One participant said he went to government websites because he believed he would find low-cost options for financial advising.

In summary, participants had varied approaches to evaluating information that involved deciding ahead of time how much effort to put into the evaluation, spending time learning about the topic, getting objective and subjective information, and seeking to avoid websites that might be scams.

6.2.3. Qualitative analysis of cognitive abilities and financial knowledge. Research Question #2 also asks about the impact that cognitive abilities and financial knowledge may have had on participants' search and evaluation strategies. Numerous approaches were attempted to code the transcription data to explore this question but as of this writing, a satisfactory approach has not been found. The first approach that was to code instances of participants' comments that seemed to indicate the constructs of speed, memory, and financial knowledge. This included coding utterances related to finding something right away on the SERP, which would be an example of perceptual speed. Examples of memory ability were codes for participants who mentioned they could not remember something, such as what they were searching for. For financial knowledge, codes were related to comments about whether or not a participant had heard of payday loans or reverse mortgages. This produced very little in the way of patterns or insights. In the future, this effort may be attempted again, along with the help of one or more experts in qualitative transcription coding. Another option would be to tailor the protocol of the stimulated recall technique to include prompts designed to elicit information about abilities- and knowledge-driven strategies.

The second approach to code the independent variables in the qualitative part of the study was to divide up the participants by the different high and low groups that were created during the hypothesis-testing phase of the study. That meant that there were high and low perceptual speed groups, high and low working memory groups, and high and low financial knowledge groups. Once divided, each group's responses were examined separately for themes but this approach also did not spawn any clear or useful insights within the data.

6.3. Limitations of the Methods

Rigorous and thorough efforts were made in designing and executing this study. Despite this, some of the main findings of the study were inconclusive. This has prompted deep reflection and retrospection about all aspects of the study, ranging from the original constructs chosen for investigation to the measures selected for the dependent variables to the procedural details, from start to finish. If the study of individual differences in information searching and assessment is to move forward, it is essential to understand which methods work best and with which user populations and contexts. It is hoped that the examination and explanation of the limits of this study can be used to design studies whose findings can more effectively address real-world problems that users face when searching for and evaluating information online, especially in personal finance topics. In this section limitations about the study design, search tasks, procedure, participants, and data analysis techniques are discussed.

6.3.1. Study design. The study was a quasi-experimental lab study that measured the effect of three independent variables on three dependent variables of participants conducting assigned search tasks on the Internet. What made the study *quasi*-experimental was the fact that the IVs were not randomly assigned, as they are in experimental studies. This is because the IVs of this study occur naturally in the world as innate abilities or knowledge acquired over a long period of time, so it is not possible to assign them to participants. Instead, the IVs are measured as part of the study procedure and then IV groups are created post hoc based on the test scores of the participants. The limit to doing this is that there is no control group against which the IV groups can be measured. Without having that baseline of performance, the only way to understand performance is to compare groups of participants in the study to groups of other participants in the study. The limitation is that this makes the outcome measure relative rather than absolute. It is relative to the particular sample of participants who may or may not represent a typical sample of users. It is relative to aspects of the task design such as complexity or structure. This also means that the study is not replicable. The best that can be done is general comparison of findings with cautions about differences between the studies. One implication is that there is no way to establish standards for human interaction in Internet-based systems.

The cognitive ability instruments were psychometric tests designed to measure perceptual speed and working memory. After examining the cognitive abilities literature and its theories, I decided to ground this work with Carroll's Three Stratum Theory because the factored framework of the model seemed like it would be ideal for being able to pinpoint specific cognitive abilities and tie them to core activities in searching and assessing.

The perceptual speed test, called *Finding A's,* is a paper and pencil test from the Ekstrom Kit and the Memory Span test is a CD-based test from CogLab. Both have been used in previous ILS and IIR studies with mixed results. *Finding A's* (PS-1) is one of three tests that can be used to measure perceptual speed in the Kit, along with the *Number Comparison* (PS-2)and *Identical Pictures* (PS-3) tests. Guidance for these tests indicated that it was appropriate to use one of the three tests and it was recommended that when the test is administered, that both parts of the test are given. Past research has uncovered a link between these measures of perceptual speed and the volume of interactions that users make such as mouse clicks, queries, and page views. It makes sense that perceptual speed, which is equated with cognitive speediness (Carroll, 1993; Salthouse, 2017), would enable people with higher perceptual speed to scan a SERP or a webpage faster than those whose speed was slower. What has yet to be understood, however, is the extent to which this speed advantage amounts to a real-world advantage for those with higher speed. As described in the literature review, Al-Maskari and Sanderson (2011) found that users with higher perceptual speed were faster at finding their first relevant documents. However, these users did not find significantly more webpages and both the high and low groups of users rated their satisfaction levels the same. Other studies such as Brennan et al. (2014) and Turpin et al. (2016) found that higher perceptual speed users interacted more with the search systems, but neither measured other outcomes that might link the speed advantage to a more tangible outcome such as finding more or better information. In the case of the dissertation study, the difference between the two perceptual speed groups was not significant and additionally, the group with higher perceptual speed interacted slightly less than those with lower perceptual speed. It may be the case that this ability fluctuates within people. There may be a dynamic nature to this ability that cannot be measured using a static, paper-and-pencil, timed test. Another notion to consider is whether a paper-and-pencil test is the most valid instrument for testing cognitive speed in digital environments. It also may be that other factors influenced the outcomes, such as users' levels of engagement.

Another issue with the use of this and other cognitive abilities tests in information searching is that there is no ground truth about what an "average" level of perceptual speed in the context of searching. Unlike the field of psychometrics, where intelligence quotient (IQ) tests center around a score of 100 as the point of average intelligence, there is no such thing in IIR as an "average" searcher. There is no established amount of time, context, or amount of interactions that is considered the norm. This makes understanding abilities in searching challenging because it will always be the case in a study that participants can only be compared to other participants in the same study and not to a commonly held standard. How long should it take someone to find the "right" webpage for a fact-finding task? An exploratory task?

The second cognitive ability measured was working memory, which was operationalized using a memory span test. The reason working memory is an important variable in information searching and evaluation has to do with the well-known fact that human memory has capacity limitations. The implication of overloaded memory in online searching and assessment is that the user will be unable to inhibit off-path information or attentional intrusion. This kind of poor inhibition affects the brain's ability to encode information for later retrieval (Hasher & Zacks, 1988). When measuring memory, typically a measure is taken and then correlated with a person's performance on a task. In psychology research, the tasks are single trial object representations, which are vastly more simple than someone searching for information on a SERP. The idea behind psychometric measurements in psychology is that the right instrument will measure the immutable trait called “memory” in a user and this trait will transcend the specific task. In other words, the measure is measuring not the interaction of that person’s specific set of skills relating to immediate task being performed, but instead, is measuring the immutable capacity of that person’s ability to perform all such tasks. In considering this regarding information search tasks, it seems unlikely that a single measure of memory such as a memory span test, would be able to capture that immutable capacity given the complexity of even the most simple search tasks on a search engine. It is no wonder that much of the research on memory has either been inconclusive or has found results for single, study-specific contexts. It could be that it is not possible to measure working memory using a single measure and that instead the more appropriate way to think about working memory in Internet searching and evaluation is as a component ability with component processes, such as that proposed by Kyllonen and Christal (1990) or by Baddeley and Logie (1999).

Of course, there are numerous approaches to studying working memory depending upon which theory one is following (Baddeley, 2012), but a well-documented view of working memory in the psychology literature proposes working memory is comprised of a memory storage unit (or “store”) and also a cognitive processor (Baddeley, 2007). In numerous studies that subscribe to this view, it was found that study participants were able to successfully complete different kinds of tasks when only memory span was taxed, but that when additional load was put on processing memory, participants’ performance diminished (Han & Kim, 2004; Meghanathan et al., 2015; Peterson et al., 2008). While it is acknowledged that these studies conducted multiple trial tasks within the psychology visual search experimental paradigm (and thus, they were not information search tasks that require reading and other mental processing), it still seems worth considering that the memory span measure alone, such as was used in the dissertation study, is not sufficient for operationalizing the whole of working memory in information search study participants.

For these two cognitive abilities and perhaps others, it could be that we have not yet identified the true cognitive constructs for searching in the Internet's hypertext environment nor the cognitive constructs for evaluating webpages in this environment. It may be that new constructs need to be developed such as digital memory span, digital working memory, and digital processing speed. The present study raises the possibility that the use of psychometric tests in IIR studies calls for greater clarity or caution in their use, taking a step back to first really define the constructs we are trying to study and if these are in fact, the right constructs to be studying.

The financial knowledge measure was a set of test questions created by combining nationally administered surveys in the U.S. and worldwide. This is one of only a few efforts in IIR or the broader area of information science to understand personal finance-related information searching and assessment on the Internet. As a result, there was no guidance to use for operationalizing financial knowledge and applying it to finance-related search tasks. This study was the first attempt at doing so and the findings of the study were inconclusive so there is no way to determine the validity of the instrument. In addition, other user studies of different kinds of domain knowledge (Freund & Toms, 2006; Hembrooke et al., 2005; Zhang et al., 2005) operationalized domain expertise using different approachES such as allowing users to self-select their areas of expertise or identifying the users' membership in a university major or an occupation. This seems to be an important point. The participants in the current study, when asked to assess their overall level of knowledge about finance (Appendix B), rated themselves much differently than the U.S. sample or the North Carolina sample. The participants of this study rated themselves lower on overall knowledge, lower on dealing with day-to-day financial matters, and lower on being “pretty good at math”. In addition, when asked the question, “Who in your household is most knowledgeable about saving, investing, and debt?” (Appendix C) close to half the participants in this study (47%) selected the response, “Someone else” while only 9% of the U.S. sample selected this response and only 6% of the North Carolina sample selected it. Based on this, it seems reasonable to conclude that the participants in this study did not see themselves as having high financial knowledge. It makes sense then, that participants did not have high domain knowledge when thinking about the search strategy results in the qualitative part of the study. This may have been similar to the findings of Hölscher and Strube (2000), in which they found that experts had more sophisticated strategies and were also more flexible with their strategies, whereas novice Web searchers tended to search in the same manner every time, whether or not the search was proving effective. The difference is that the participants in the current study were not novices to searching, but they were all novices to searching for debt-related financial information online, and so there were minimal differences between the high and low financial knowledge groups.

A measure and instrument that the results of this study call into question are the mental workload measure and subsequent use of the NASA-TLX mental workload questionnaire. The mental workload construct was developed in the field of human factors research and has primarily been used in human-machine research in aeronautics and industrial machine operations. It may be that this human factors construct does not translate to the world of online searching and assessment and that a better construct to consider would be that of cognitive load.

Cognitive load and theories related to it (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Sweller, 1988; van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009; Wiebe, Roberts, & Behrend, 2010) were developed in the field of education where the concept of "load" is based in mental activities without the physical activity component of the human factors concept of load. Thus, cognitive load may be more in line with the mental activities of Internet searching and information assessing, as well as searching as learning. It might also be valuable to consider the concept of workload (or cognitive load) as a multidimensional vector rather than a scalar quantity, given the multi-layered, spatial, networked environment of the Internet. In terms of the instrument itself, it is not clear whether this index is a valid measure for the mental workload of searching. Study findings in IIR have varied greatly for the TLX. Fortunately, there are new efforts underway to develop information searching-appropriate measures for mental workload, such as by Wilson, Sharon, Maior, Midha, Craven, and Sharples (2018).

Another limitation of the study was the use of the two expert assessors. The lack of assessor agreement was disappointing. While both experts have extensive backgrounds in personal finance, expert #2 was more negative overall in her assessments than expert #1 and there was no clear way to determine why. In addition, expert #2 was much more negative on the reverse mortgage and payday loan tasks than on the student loan tasks. A better approach for this process would have been to allow time for the two experts to calibrate their judgments in order to achieve better agreement. Unfortunately, the low agreement had major ramifications in the study, as it was not possible to determine the extent to which the different groups of participants differed in their assessments of the webpages. An interesting anecdotal finding, however, was that both assessors commented to me that the overall quality of the set of webpages was poor and this was evident in the difference in the trends of the scores of the participants versus the two assessors (see Figure 18). This finding should be pursued in future research, as it has important implications for consumers who need information about personal finance. If consumers trust the system to retrieve good information for them and they evaluate the information as relevant, but from the standpoint of a subject matter expert the information is actually poor, then there are problems that could arise from this situation.

Another limitation to the study design was using the open Internet and live search engines (i.e., Google). There was no defined document corpus and no ability to control what participants viewed on the screen. For example, it was not possible to know things such as whether Google was using any experimental algorithms during the course of the study, which lasted about six weeks. All participants were not exposed to the same set of documents and it is not known what impact this had on the outcomes of the study.

6.3.2. Search tasks. There were several limitations related to the search tasks. Designing the tasks to be complex meant that people did not get any kind of mental break in which they could search for something easy and have immediate success. People expected the tasks to be somewhat hard for finding relevant webpages, but then in the post-task difficulty question they responded that they actually found the tasks to easier than they had anticipated. Since participants indicated that they had low prior knowledge of the task topics, this finding about the post-task difficulty may have been an indicator that participants were relieved after the tasks that the searching was easier than they expected. It may also be an indicator of other user experiences as well, such as searching as learning (Eickhoff et al., 2017) or user engagement (O'Brien & Toms, 2008).

In addition, designing all the tasks to be equally complex meant that when it came time for participants to answer the mental workload questions, they may not have had enough of a frame of reference of what less demanding questions felt like.

Other limitations related to the tasks had to do with participants' perceptions of the tasks, based on the pre-task questionnaires. In particular, participants rated themselves as having low knowledge of the three topics on average (2.3 on a 5 point scale) and as having low relevance to their lives (2.5 on a 5 point scale), which may have been an indicator of how motivated they might have been to search on the tasks. In addition, two of the three task scenarios situated the participant as helping someone else (e.g., a grandfather and a friend), which may have reduced the motivation they might have had to search for their own interests. This may have created distance from the task and influenced participants to be less interested in it.

The type of debt in the topics was different kinds of loans. The notion of borrowing money is more abstract than the idea of purchasing a desirable consumer good using a credit card and this may have also distanced participants from the tasks. Participants may not have had enough of an emotional investment in the tasks to be fully engaged.

Overall, the complexity level of tasks combined with the abstractness of the task topics and the distanced scenario may have created a ceiling effect, in which everyone's interactions were similar because everyone had surpassed their optimal searching situation.

6.3.3. Procedure. For the procedure, participants came to the lab at UNC for two separate sessions, first for the search session and then for the testing session. Rather than using this approach, another way to conduct the study would have been to recruit a much larger number of participants to a group testing session, where the psychometric and financial test could have been administered. From that session, the lowest and highest groups of abilities and financial knowledge could have been chosen, to ensure greater diversity in participants' abilities.

6.3.4. Participant characteristics. Some of the characteristics of the participants may have limited the study as well. For example, more than half of the participants were college-educated and 15 had post-graduate degrees. This may have reduced the intellectual heterogeneity of the sample needed to have diverse cognitive ability scores. The perceptual speed test, *Finding A's*, was also used in a study by Turpin et al. (2016). The sample in that study (N = 16) had an average score on this test of 51.94. The average score of the participants in the current study was 63.10, a higher scoring sample.

Another limitation was mentioned earlier and that was the participants' responses to the financial self-rating questionnaire (Appendix B) and part of the financial experience questionnaire (Appendix C). When asked to assess their overall level of knowledge about finance (Appendix B), participants rated themselves much differently than the U.S. sample or the North Carolina sample. The participants of this study rated themselves lower on overall knowledge, lower on dealing with day-to-day financial matters, and lower on being "pretty good at math". In addition, when asked the question, "Who in your household is most knowledgeable about saving, investing, and debt?" (Appendix C) close to half the participants in this study (47%) selected the response, "Someone else" while only 9% of the U.S. sample selected this response and only 6% of the North Carolina sample selected it. Based on this, it seems reasonable to conclude that the participants in this study did not see themselves as having high financial knowledge. So it may be the case that the high and low grouping for financial knowledge was not useful because everyone may have had lower than average knowledge in this area.

6.3.5. Data analysis techniques. The results reported in this dissertation are of analyses across task-level results. Analyzing data at this higher aggregate level may have made more subtle differences undetectable, particularly regarding eye movements. It may be the case that by analyzing the tasks in segments such as in thirds or some other level of granularity, that differences between users would become more obvious. In this way, variations that occur early-on in tasks might be detectable that would otherwise be diluted by later-stage task behaviors where the effect is no longer occurring. I did try to segment the data this way, however, the cut-points did not fall in any natural place. For example, cutting the task in half meant that some data had to be eliminated, such as whole webpages that the participants were currently reading, which then impacted the rest of the results for those participants. It was not possible to determining a clear place within the search tasks for dividing up the tasks.

CHAPTER 7: CONCLUSION

This dissertation study set out to understand the influences that cognitive abilities and financial knowledge have on outcomes related to search, assessment, and mental workload of adults searching online for debt-related personal finance information. The results of the study were mixed. The most important finding of the study was that the topic of personal finance, specifically in the realm of financial loans such as mortgages, student loans, and payday loans, is more challenging for people than they realize. Participants reported low prior knowledge of all the task topics and used simple search strategies such as avoiding advertisements on the SERP, relying heavily on the first result of the SERP, and reformulating queries rather than investigating the SERP at deeper levels. Even with self-rated low knowledge of the topics, many participants did not seek out general kinds of knowledge-building resources online but instead began evaluating webpages immediately based on task criteria. As well, when participants did find highly relevant information on websites such as www.ftc.gov and www.cfpb.gov, they did not continue to search deeper on those sites, but instead went back to Google and started their searches over. Participants rated most of the webpages they found as relevant or very relevant but expert assessors rated most of those same pages as only somewhat relevant or not relevant.

In the study, a theoretical model was proposed and in this model, financial knowledge acted as a moderating variable on the effect that cognitive abilities would have on search and evaluation behaviors as well as mental workload. The aspect of financial knowledge studied was debt literacy and the factors of cognitive abilities modeled were perceptual speed and working memory, while the information tasks were about debt-related loan information and products. None of the hypotheses were supported.

In addition to the quantitative measures from log and eye tracking data and self-report questionnaires, a qualitative component was included in the study to explore participant's conscious strategies while searching. Through the qualitative work, insights about how people think about debt-related topics and how to search for them were gathered.

7.1. Contributions

This study makes several contributions to research. This was a first effort at bringing together a set of known cognitive ability variables – perceptual speed and working memory – known to influence different search behaviors, to explore how they influence searching and assessment of personal finance-related topics. As part of this effort an original definition for the term *cognitive abilities* was proposed which captures both the inherent and learned aspects of ability and situates abilities within the context of cognitive tasks, making the definition more suitable and useful in IIR and ILS studies than non-task based definitions.

Exploring searching and assessment using the subject domain of debt-related personal finance personal is also a main contribution of this dissertation. Personal finance as a domain or as a task topic has received very little attention in IIR and ILS, despite its importance as a foundation to being able to live a life of health, well-being, dignity, and autonomy in modern societies throughout the world.

Next, this study contributed a main model for studying topics in personal finance domain that is specific to the influence of cognitive abilities and financial knowledge on three outcomes: search behavior, relevance assessment, and mental workload. This was the first attempt in ILS research at creating such a set of models in the context of any aspect of the personal finance topic domain. The testing of hypotheses on the model were unsuccessful which now provides the first ever information that can inform future model design and hypothesis development in this area.

Another contribution of this study is that it was conducted with working adults, an underrepresented population in the research on information searching. Much of the research on information searching has used convenience samples of university undergraduate or ILS students and faculty, which has prompted a call for more research of the general population so that findings can be generalized more broadly (O'Brien et al., 2017).

Finally, the study provides qualitative insights about participants searching in complicated topic domains. Participants had low prior knowledge of the task topics and relied on basic search strategies for finding and evaluating information. The findings may also point to participants' habituated response to searching on Google, in which participants trust and rely on information from the first non-sponsored search results, while evidence from expert assessors indicated that much of this information was less than adequate for addressing the task scenarios. This has implications for the algorithmic integrity of top-level search results in life-critical areas such as personal finance and calls into question the current paradigm that favors monetarily optimized or page-link prioritized search results over objective, expert-verified, neutral results. It also has implications about the use of general purpose search engines designed to retrieve basic, easy-to-read information for average users. An argument can be made for the creation of a new type of vertical that uses domain-specific algorithms and indexing for retrieving personal finance-related content. Overall, findings from this dissertation provide a basis for further substantive research in information science and consumer finance.

An unintentional contribution of this study is that it exposes the complexity of studying information searching, relevance, and open web study designs. It exposed the challenges of studying individusl differences and basing research hypotheses on only the known, published significant results. With bias in publishing toward studies that show significant findings, it is not possible to know when a psychometric measure has had more failed uses than successful ones. Thus, it may be possible to misread signals from the literature that make it seems like there is more evidence to suggest a certain variable will impact a behavior but in reality there is no way to know. In addition, there may be so many differences between studies that is not useful to think about them as an accumulation of evidence pointing in the same direction. Psychometric tests need to be aligned with search task situations, which may be the most challenging aspect of individual differences research. It requires a level of precision that first-pass studies will not have, but their very nature of being exploratory in nature. Whittling down to the precise levels of measurement that work requires a series of studies over time. This may be why studies in psychology are so refined and focus on minute changes in behavior.

7.2. Future Directions of the Work

In addition to the contributions of the research, there are several avenues for promising future work. More research needs to be done at the most basic level to understand how users search for personal finance topics. There are many ways to do this and one good way would be to allow study participants to bring their own financial topics to search or to assign more basic-level topics for participants and allowing them to choose which topic they wish to search. Additionally, more in-depth qualitative inquiry will provide stronger insights into how people think about personal finance and evaluate information related to it.

More work needs to be done at defining the construct called *financial knowledge.* This may incorporate tests of financial literacy but should also allow for self-selection into “high” knowledge categorizations. It would also be useful for research to be conducted on specific populations that relate to the financial knowledge construct, such as homeowners or recent retirees.

Beyond the areas just mentioned, this research lays the groundwork in IIR and ILS for exploring a wide variety of phenomenon. By using concepts and theories from behavioral economics and neuroscience, it may be possible to explore ways in which people's perception of risk (Kahneman & Tversky, 1979) creates uncertainty and draws out their feelings of aversion to loss (Tversky & Kahneman, 1991) and risk (Peterson, 2007), which in turn can lead to the kinds of information avoidance behaviors previously explored in information and library science (ILS) (Case, Andrews, Johnson, & Allard, 2005). In this way, IIR and ILS researchers can take what we already know about human information behavior and apply it to a variety of phenomenon well-studied in consumer finance and economics.

APPENDIX A: DEMOGRAPHIC QUESTIONNAIRE

What is your gender?
❍ Male
❍ Female
❍ Non-conforming (please describe): _____________________

What is your age? _____

Which of the following best describes your or ethnicity? Select all that apply.
❑ American Indian or Alaskan Native
❑ Asian
❑ Black or African-American
❑ Hispanic or Latino/a
❑ Native Hawaiian or other Pacific Islander
❑ White or Caucasian
❑ Other (please describe): _____________________
❑ Prefer not to say

What is the highest level of education that you have completed?
❍ Did not complete high school
❍ High school graduate - regular high school diploma
❍ High school graduate - GED or alternative credential
❍ Some college, no degree
❍ Associate's degree
❍ Bachelor's degree
❍ Post graduate degree

Which of the following best describes your current employment or work status?
❍ Self-employed
❍ Work full-time for an employer
❍ Work part-time for an employer
❍ Homemaker
❍ Full-time student
❍ Permanently sick, disabled or unable to work
❍ Unemployed or temporarily laid off
❍ Retired

APPENDIX B: FINANCIAL KNOWLEDGE SELF-RATING QUESTIONNAIRE

Table 58. *Responses to Financial Self-rating Questions.*

Responses to financial self-rating questions: national, state, and dissertation data comparison

			NFCS - U.S.	NFCS - N.C.	Dissertation
Q#	**Question**	**Answer**	**(N=27,564)**	**(N=505)**	**(N=47)**
II.Q1	On a scale from 1 to 7, how would you assess your overall financial knowledge?	Low 1-3	7%	7%	20%
		Neutral 4	14%	14%	36%
		High 5-7	76%	77%	44%
		Don't know	2%	2%	0%
II.Q2a	I am good at dealing with day-to-day financial matters . . .	Disagree (Low 1-3)	7%	6%	13%
		Neutral (4)	12%	12%	11%
		Agree (High 5-7)	81%	82%	76%
		Don't know	1%	0%	0%
II.Q2b	I am pretty good at math	Disagree (Low 1-3)	10%	7%	16%
		Neutral (4)	11%	13%	18%
		Agree (High 5-7)	79%	79%	66%
		Don't know	0%	0%	0%

APPENDIX C: FINANCIAL EXPERIENCE QUESTIONNAIRE

Table 59: *Responses to Financial Experience Questions.*

Responses to financial experience questions: national, state, and dissertation data comparison

			NFCS - U.S.	**NFCS - N.C.**	**Dissertation**
Q#	**Question**	**Answer**	**(N=27,564)**	**(N=505)**	**(N=47)**
III.Q1	Who in your household is most knowledgeable about saving, investing, and debt?	You	54%	53%	40%
		Someone else	9%	6%	47%
		You and someone else, equally knowledgeable	34%	38%	13%
		Don't know	2%	3%	0%
III.Q4	Do you [your household] have a checking account?	Yes	91%	91%	100%
		No	7%	8%	0%
		Don't know	1%	0%	0%
III.Q5	Do you [Does your household] have savings account, money market account, or CDs?	Yes	75%	74%	98%
		No	23%	25%	2%
		Don't know	1%	0%	0%
III.Q9	Do you [or your spouse/partner] own your home?	Yes	60%	62%	38%
		No	39%	38%	60%
		Don't know	1%	0%	2%
I.Q10	Do you currently have any mortgages on your home?	Yes	57%	60%	44%
		No	41%	39%	56%
		Don't know	1%	1%	0%
		Prefer not to answer	--	--	2%
III.Q12	How many credit cards do you have?	1	18%	19%	23%
		2-3	33%	30%	32%
		4-8	22%	20%	19%
		9-12	3%	3%	0%
		13 or more	1%	1%	2%
		No credit cards	21%	26%	23%
		Don't know	1%	1%	0%
III.Q13	Do you [Does your household] currently have an auto loan?	Yes	30%	30%	34%
		No	68%	68%	66%
		Don't know	1%	1%	0%
III.Q14	Do you currently have any student loans? If so, for whose education was this/were these loan(s) taken out? (Select all that apply)	Yourself	19%	16%	43%
		Spouse/ partner	7%	6%	9%
		Your child(ren)	4%	3%	0%
		No, do not have	73%	77%	49%
		Don't know	1%	0%	0%
		Prefer not to answer	--	--	2%

APPENDIX D: PRE-TASK QUESTIONNAIRE

[Each pre-task questionnaire showed the current search task on the screen, above the 5 questions shown on this page.]

1. How much do you know about this topic?

	1	2	3	4	5
Nothing:I know details	❍	❍	❍	❍	❍

2. How relevant is this topic to your life?

	1	2	3	4	5
Not at all:Very much	❍	❍	❍	❍	❍

3. How are interested are you to learn more about this topic?

	1	2	3	4	5
Not at all:Very much	❍	❍	❍	❍	❍

4. Have you ever searched for information related to this topic?

	1	2	3	4	5
Never:Very often	❍	❍	❍	❍	❍

5. How difficult do you think it will be to search for information about this topic?

	1	2	3	4	5
Very easy:Very difficult	❍	❍	❍	❍	❍

APPENDIX E: SEARCH TASKS

Reverse mortgage task:Your grandfather told you he is thinking about taking out a reverse mortgage on his home. A mortgage broker at his church told him it was the perfect way to get extra cash to buy that sailboat he has dreamed about having for a long time. But you are not certain this is the right choice for your grandfather, so you need to find out more about reverse mortgages. Your goal is to find reputable, trustworthy webpages that can help you understand reverse mortgage loans. To guide you in your search, here are a few questions that such webpages would be able to answer for you:

- What are reverse mortgages?
- What are the advantages of reverse mortgages? Disadvantages?
- Is a reverse mortgage a good way to get extra cash for something like a sailboat?
- What are the names of some high quality companies that offer reverse mortgage loans where your grandfather lives, in Nags Head, North Carolina?
- What resources can your grandfather use if he has more questions?

Payday loan task: Your close friend seems to have gotten into a financial bind. He didn't have all the money to pay the rent on his apartment in Virginia and was worried he would get evicted, so he went online and took out a two-week loan for $500. The loan was due when he received his next paycheck, but by that time, the amount he owed was $600 because there were additional fees and interest added on. You realize that this seems like a great deal of additional interest and fees to pay on such a short-term loan. You want to help your friend understand what this kind of loan. Your goal is to find reputable, trustworthy webpages you can share with him so he can better understand the loan he has taken out. To guide you in your search, here are a few questions that such webpages would be able to answer for your friend:

- What kind of loan is this?
- What are the most important things he should know about this kind of loan?
- What resources can he use if he has more questions?

Student loan task: You have just finished your university degree and need to find out what your options are for paying off your student loans. You know practically nothing about this topic, because when you were in school you were so focused on your studies you didn’t have time to learn more about paying off your loans. Your goal is to find reputable, trustworthy webpages that will help you learn more about paying off your student loans. To guide you in your search, here are a few questions that such webpages would be able to answer for you:

- What are the most important things to know about paying off student loans?
- What kinds of options are there for paying off student loans?
- Are there any special programs to consider?
- Is there anything to watch out for, such as loan payoff scams or programs that will cost a lot more money than others?
- What resources can you use if you have more questions in the future?

APPENDIX F: POST-TASK QUESTIONNAIRE

Participant ID: _____

Please check the task you just completed:

❍ Task R
❍ Task M
❍ Task S

Thank you for completing this search tasks. Now we would like to ask you a number of questions regarding your experience while completing the search tasks.

1. How difficult was it to find relevant documents?

	1	2	3	4	5	6	7	8	9	10
Very Difficult:Very Easy	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

2. How would you rate your skill and ability at finding relevant documents?

	1	2	3	4	5	6	7	8	9	10
Not Good:Very Good	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

3. How would you rate the system's abilities at retrieving documents?

	1	2	3	4	5	6	7	8	9	10
Not Good:Very Good	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

APPENDIX G: NASA-TLX MENTAL WORKLOAD QUESTIONNAIRE

For this section, please rate your answers to the following questions on the scale provided:

	Very Low				Neither Low nor High					Very High
How mentally demanding was it to complete the search tasks?	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
How physically demanding was it to complete the search tasks?	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
How hurried or rushed did you feel while completing the search tasks?	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
How successful were you in completing the search tasks?	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
How hard did you have to work to accomplish your level of performance while completing this task?	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
How insecure, discouraged, irritated, stressed, or annoyed were you while completing this task?	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

APPENDIX H: RECRUITMENT EMAIL

To : UNC Employee Email List
Cc : dianek@email.unc.edu; kbrennan@unc.edu
Subject : Participants needed for Internet Search Study

----- Message Text -----

Are you familiar with Google and Bing and other search engines? Are you interested in earning some extra money and helping me with my research by conducting some searches? If so, I need your help understanding how adults search the Internet. This is for my dissertation research. In this study, you will be asked to search the Internet to address three information scenarios and then complete several skills and knowledge exercises.

This study takes place in 2 sessions. The **first session** lasts from **50-60 minutes** on one day and the **second session** takes about **20 minutes** to complete (the second session takes place on a different day, but after the first session). You will receive **$25 in cash** for participating!

This study will take place in Room 09 in the Interactive Information Retrieval Lab in Manning Hall at the School of Information and Library Science.

You must fulfill the following in requirements in order to qualify for participation:

- you must be 18 years or older
- you must have at least 2 years of experience searching the internet
- you must be a native-English speaking (in other words, English is your first language)
- you must be able to complete both sessions

Please email me at **kbrennan@unc.edu** to schedule your participation.

** You will not be offered or receive any special consideration if you take part in this research; it is purely voluntary. This study has been approved by the UNC Behavioral IRB (IRB Study 17-0077)

Many Thanks!
Kathy Brennan

APPENDIX I: FINANCIAL EDUCATION RESOURCES

American Institute of CPAs

https://www.360financialliteracy.org/

Consumer Financial Protection Board (CFPB)

https://www.consumerfinance.gov/

Federal Reserve Bank of St. Louis

https://www.stlouisfed.org/education

Financial Industry Regulatory Authority (FINRA) - Investor Education Foundation

http://www.finrafoundation.org/resources/education/learning/

http://www.finrafoundation.org/resources/education/modules/index.html

http://www.finra.org/investors/alerts

National Endowment for Financial Education (NEFE)

https://www.smartaboutmoney.org/

http://www.myretirementpaycheck.org/

National Foundation for Credit Counseling

https://www.nfcc.org/financial-education/

4-H Cooperative Extension

http://articles.extension.org/pages/61531/4-h-build-a-million

APPENDIX J: PARTICIPANTS' ON-SCREEN INSTRUCTIONS

Instructions for Beginning the Search Sessions

General instructions:

During your search sessions, I will be recording your search behaviors (like your mouse clicking and scrolling and the keyboard keys you press) and eye movements with special software. The software and equipment will not disturb you while you are searching online. But sometimes the software and equipment are more sensitive than I would like and they don't work properly unless I make sure that certain criteria are maintained. Please help me with this by making sure that you NOT do any of the following while you are searching:

- Do not re-size or move the browsing window.
- Do not press escape at any time during the session.
- Do not open multiple browsing windows.
- Do not use tabs.
- Do not touch any equipment attached to the computer monitor.

And please also help me by making suring that you DO the following while searching:

- Use the back arrow on the browser when you want to return to the search engine results page.
- Stay seated with your head in a still position.

A small paper with this list of do's and don'ts will be available to you, taped to the left side of your computer monitor.

You will be asked to conduct three search tasks based on the specific scenarios you will be provided. Before each search task, you will read through the scenario and then answer several pre-task questions about your level of interest in the task.

Searching and bookmarking instructions:

Once you've completed the pre-task questionnaire, you will be taken to Google's search engine main page. I am starting your search at Google as a matter of convenience; if you normally prefer using Bing for searching or Yahoo! or some other English-language search engine, you are free to use that search engine to complete your task.

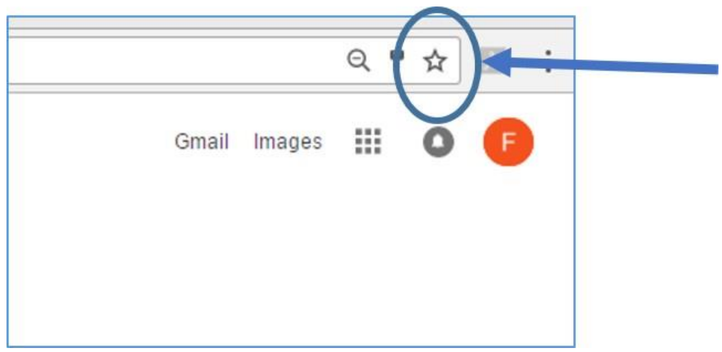
Press the

Page Down Key
to proceed to
the Next Page

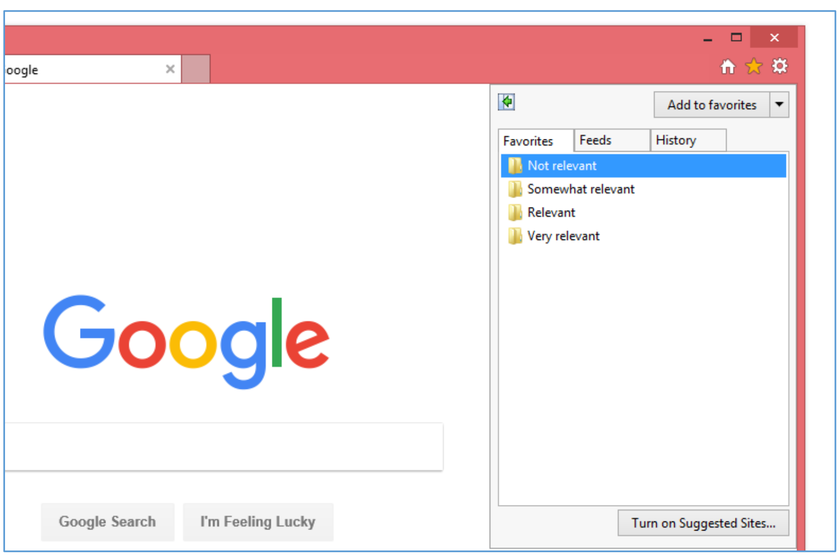
While you are searching, the task scenario will be available to you on the right-hand side of the page. Please do not move the task from its location. You may view it as often as you need to (or not).

To accomplish your task, you will need to bookmark webpages that you view according to four criteria: *not relevant, somewhat relevant, relevant,* and *very relevant.* Each criteria is the label for a bookmarking folder, so when you bookmark a page, you will bookmark it to the folder labeled in the way that you believe best describes the page you are viewing. Below are some images of the bookmarking tool and how you would go about bookmarking.

The bookmarking tool is a star in the upper right corner of the browser window. You can see it in this image in the middle of the blue circle that has an arrow pointing to it:



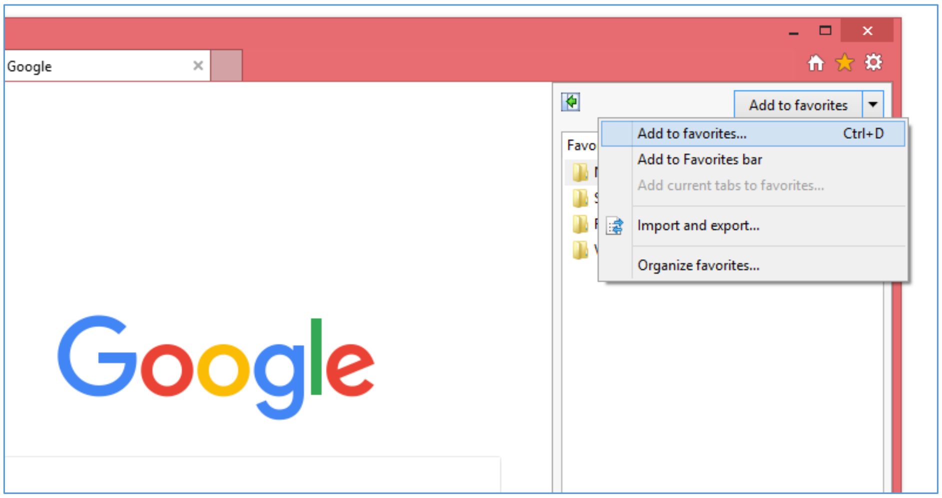
Here is an example using Google as the webpage you wish to bookmark. When you decide you want to bookmark this webpage, you click on the star once and a dropdown menu will appear that will look similar to this --



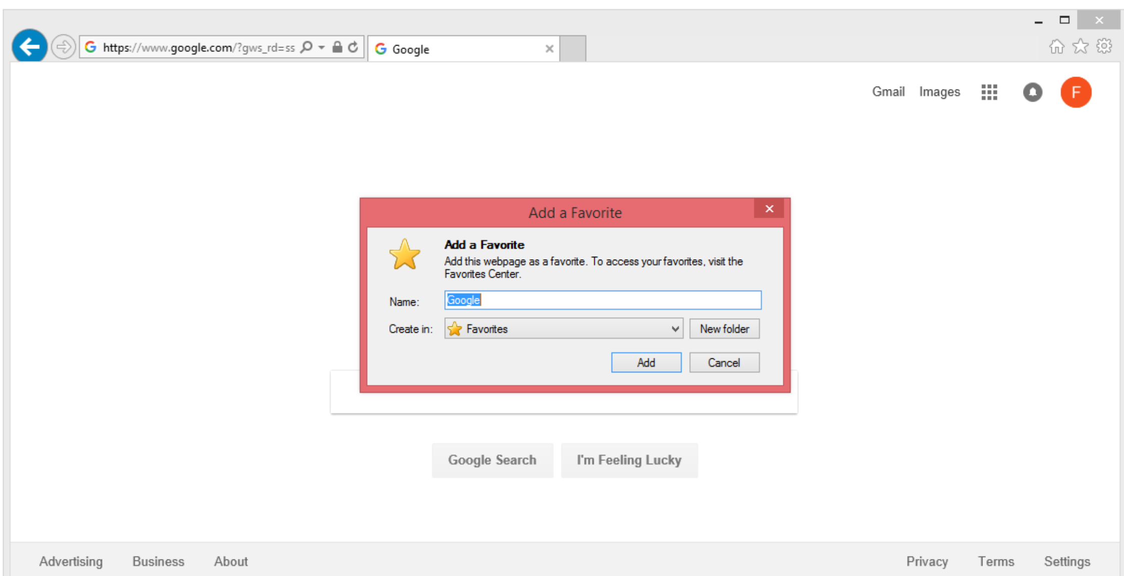
Press the

Page Down Key
to proceed to
the Next Page

Click on the button that says "Add to favorites,"



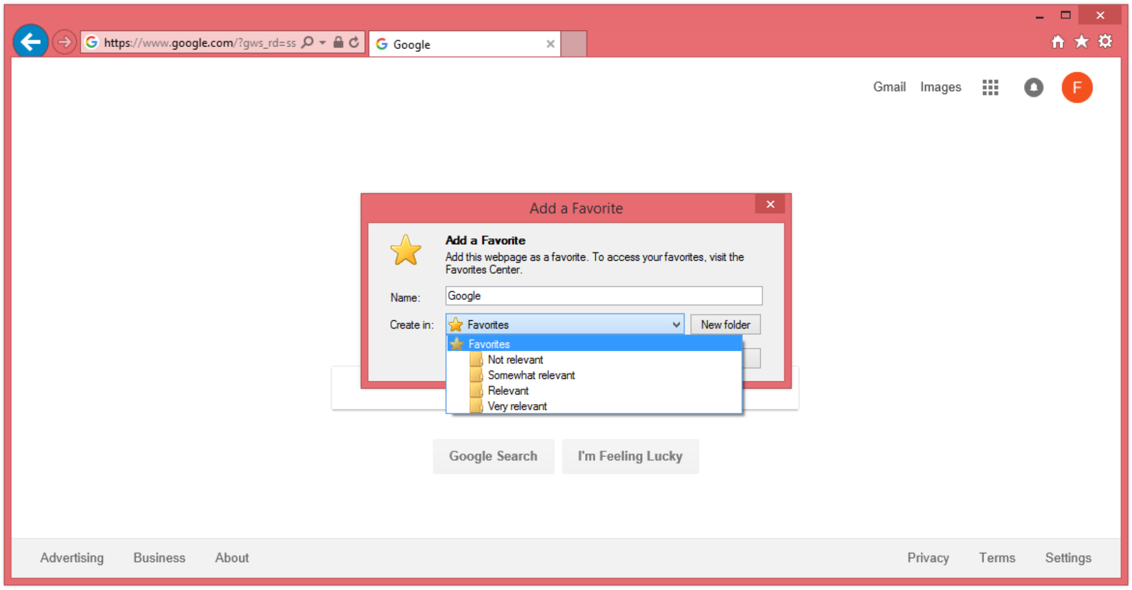
And a window will pop up onto the center of the window:



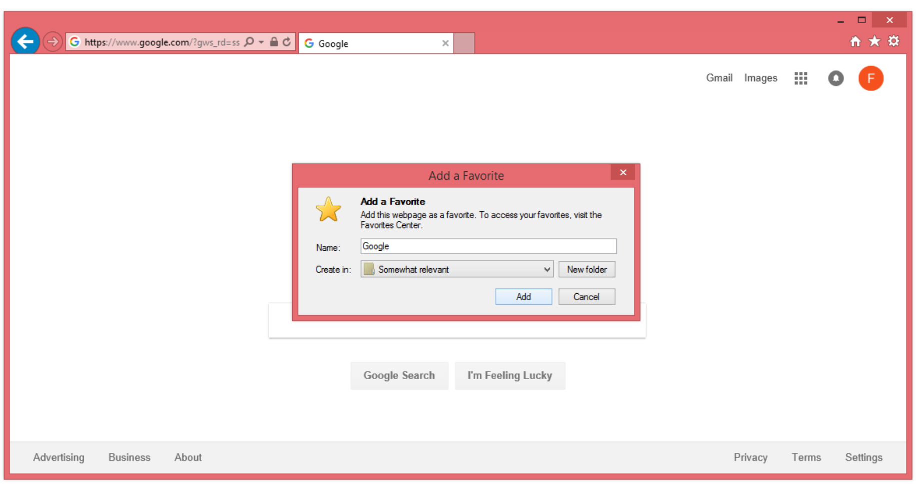
Click the box to the right of "Create in" and select which of the four categories of relevance (shown as *very relevant, relevant, somewhat relevant*, and *not relevant*) you wish to bookmark

Press the

Page Down Key to proceed to the Next Page the page as by clicking it with the arrow cursor. (If you see any other folders with other labels besides the four mentioned, you can ignore them).



In this example, the user wanted to save the webpage for Google as "Somewhat relevant," and did so by clicking on the "Add" button.



When you bookmark a page in the category of your choosing, you then go back to searching.

Press the

Page Down Key to proceed to the Next Page

Continue with your searching, according to what the task has indicated, and continue to bookmark all of the pages you view, EXCEPT you do not need to bookmark the Google search engine page (or Bing or Yahoo!). Just book the webpages you find that relate to the task.

You will have 12 minutes for each search task. For each task you will need to read and understand the task and then search the internet for webpages that will address the task. You will want to find the *best pages possible* for the task. Since I am interested in all of your search behaviors and decisions and so that I can make the most of your participation in this study, ***I am asking you to categorize every page you view with one of the four categories,*** so that I will understand which pages you think are the best (those are the ones you bookmark as ***very relevant***) of all the ones you viewed and which ones fit into the other three categories (*relevant, somewhat relevant*, and *not relevant*).

Remember to stay seated upright and keep your head still while you are searching the Internet, so that I can ensure that my equipment is behaving properly!

You will have 12 minutes for each search task. When you have been searching for 10 minutes, I will give you a verbal 2-minute warning. And when 12 minutes is up, I will ask you to stop. If you finish anytime before 12 minutes has passed, please turn to me and let me know you have finished with your searching.

Once you finish the task, you will be taken to a post-task questionnaire that will ask you questions about the difficulty and mental workload of the task. Please read these questions carefully and *answer them as they relate to the task you just completed* (not any of the other tasks).

After you've finished that questionnaire, we will move on to the second task and repeat this process. After that you will complete the third task in the same manner but at the end of the third task, we'll take a few extra minutes to process your thoughts (I will explain this to you when we get to the end of the third task).

***NOW PLEASE TELL THE RESEARCHER THAT
YOU HAVE FINISHED READING THE INSTRUCTIONS.***

APPENDIX K: STIMULATED RECALL MODERATOR'S GUIDE

Stimulated Recall Interview:

I am now going to read aloud some instructions to you.

During this next section of the study, I'm going to play back a screen recording of the actions you took during the last task you completed. While you watch this recording, I would like for you to state aloud why you took the actions shown on the screen and what you were thinking when you took those actions. I would like you to walk me through your decision-making processes you underwent as you searched.

At times I may shift the recording to another segment. At other times I might ask you about specific actions that you took during the experiment. I'll be reading these from my script and in some cases I may ask a question that you have already answered

There are no right or wrong answers here. I am simply looking for your thoughts as you review the steps you took during the experiment. Even minor thoughts will be helpful to this study.

[Scrub through the video pausing at the start of the last task (Google home page)]

"So this was your last task. The one you just completed."

[Allow recording to play until first query is entered and SERP is showing.]
[If the participants grows quiet, pause the playback]:

"Was it easy to find results that answered your query?

"What specifically made it [easy, not easy, etc.] to find results that answered your query – please give me specific example of what helped you and point with the mouse cursor at these examples on the results list."

[At Results selection on SERP if participant grows quiet]:

"Why did you choose this result?"

"How did you know it was a result you wanted? Please give me some specific examples of what helped you know this."

"Is there anything on this results page that provided a hint that you needed to select this result?"

"What did you expect to see on the webpage once you clicked this link?"

[At BOOKMARKING of a webpage if participants grow quiet]:

"Why did you choose this page as [highly relevant, relevant, somewhat relevant]? How did you know this page was [highly relevant, relevant, somewhat relevant]?"

"What did you expect to see on the page for a [highly relevant, relevant, somewhat relevant] webpage? Please give me specific examples in the recording of what helped you make this decision. Please point with the mouse cursor at these examples on the results page or full webpage."

[Where participant VIEWED a webpage BUT DID NOT BOOKMARK]:

"Why did you choose to not to bookmark this page in one of the relevance folders? How did you know that you needed to make this decision at this point?"

"How did you know you were not getting relevant information? What did you expect to see? Please give me specific examples of what helped you make this decision. Please point with the mouse cursor at these examples on the results page or full webpage."

"Is there anything on this webpage that provided a hint that you needed to leave the page and not bookmark it as relevant? Additional probing: please point at the feature or element that helped you realize that."

[ADDITIONAL PROBES to use to encourage participant to verbalize, if necessary]:

- *I'd like to hear what you're thinking*
- *If you could just say whatever words come to your mind*
- *Why was that choice attractive to you?*
- *Do you feel that choice turned out the way you expected it to turn out?*
- *What were you thinking when you made that choice?*

At end of Stimulated Recall interview:

"Thank you so much for giving your feedback. I have one final question to ask you before we end. This is a general question that applies to all three search tasks. I would like to know if you feel like you learned anything during your search session?[And if you did, could you describe to me what things you believe you learned?]

Great. Thank you. Now I just need to save these recordings to make sure I have all the data I need for the study."

1. **Stop the retrospective recording (Ctrl-Shift-Alt-F9 all at the same time)**
2. **SAVE THE RETROSPECTIVE RECORDING in Morae Recorder. Name it "PXX-retro"**
3. **Stop any other recordings**

Wrap-up with Participant

"Do you have any questions or comments about this part of the study?"

"Okay, now we need to schedule you for the session #2 of the study, your final session. Please mark on this sheet what time will work for you and I will send you a confirmation email to remind you of the date and time. Session #2 will take about 20 minutes, but I would allow for a half hour just in case it takes longer.

Also, at the end of session #2, I will give you $25 cash compensation for your participation and ask you to sign a receipt."
"Thanks again for participating in our study today.

[Get up and walk participant to the door.]

APPENDIX L: EXPERT ASSESSORS' INSTRUCTIONS

Expert Assessor Instructions

Thank you again for agreeing to be one of my expert assessors. In this document I've provided information and instructions for the expert assessment part of my dissertation. Hopefully I've explained everything you need to know to get started, but don't hesitate to ask questions about anything that isn't clear.

As a quick refresher, here is a short summary of the focus of my dissertation:

> *My dissertation explores the relationships of cognitive abilities and personal finance-related domain knowledge with users' search behaviors. The lab-based study measures three independent variables (financial knowledge, perceptual speed ability, memory span ability) and three dependent variables (search behaviors, evaluation behaviors, user's experience of effort), using measurements and data collection techniques chosen from an extensive literature review and first-hand use. These include questionnaires, psychometric tests, search interaction logging, eye tracking, and stimulated recall interviews.*
>
> *The goal of the research is to differentiate cognitive abilities-driven search behaviors from knowledge-driven search behaviors. This is important because the difference between the two also drives interface and systems design decisions. Individual differences in cognitive abilities will require interface and system modifications, whereas differences related to domain knowledge should first be addressed through improvements to users' levels of literacy – in this case, financial literacy – before modifications are made to systems or interfaces.*

Below I give a brief explanation of the setup of the search study and tasks, the criteria I gave participants for judging the information they found, and a general overview of the resulting data I gathered from participants. Then I explain the list of URLs I've sent you and provide guidance on how to go about providing your expert assessments. I've also footnoted some things along the way that have to do with the study design because I suspect you might be interested in such things (but if not, feel free to skip the footnotes).

Study setup and tasks:

For the study, I asked each participant to complete three search tasks[9]. Each search task was based on a scenario designed to mentally place the participant in a situation that would require her or him to search for information online at length, that is, beyond just a quick search. This design was successful, as I found that most participants searched for the full 12 minutes allotted for each task. By the way, there were 48 participants in the study and my final sample size was N=47, after I threw out the data from one participant who did not follow the study instructions.

[9] The order in which the tasks were given to participants was rotated to avoid a task order effect in the results.

Criteria for judging webpages:

To accomplish the tasks, I had participants use the bookmarking function in Internet Explorer to save the webpage URLs they viewed. They were instructed to assess the information in the webpages and bookmark each according to one of four criteria: *not relevant, somewhat relevant, relevant,* and *very relevant.* The only pages they didn't have to bookmark were the search engine results pages (SERPs) for Google or whatever search engine they were using.

Here is part of the instructions I gave to participants on how to go about bookmarking the different webpage URLs they found:

> "You will have 12 minutes for each search task. For each task you will need to read and understand the task and then search the Internet for webpages that will address the task. You will want to find the *best pages possible* for the task. Since I am interested in all of your search behaviors and decisions and so that I can make the most of your participation in this study, ***I am asking you to categorize every page you view with one of the four categories.*** That way I will know which pages you think are the best (those are the ones you bookmark as ***very relevant***) of all the ones you viewed and which ones fit into the other three categories (*relevant, somewhat relevant*, and *not relevant*)."

Resulting dataset:

After each participant's session, I copied and pasted all of the categorized URLs the person saved into a spreadsheet. After all participant sessions were complete, I organized the spreadsheet by the three search tasks. Not surprisingly, some URLs were common to numerous participants – I expected this to happen, since all participants completed the same three tasks[10]. To make your work easier, I eliminated all of these obvious duplicates in each task set[11]. After doing that, I ended up with the following sets of URLs:

Task R: This is the payday loan task. It has 231 URLs.

Task S: This is the student loan task. It has 206 URLs.

Task M: This is the reverse mortgage task. It has 204 URLs.

This is more bookmarks than I anticipated and may seem like a lot to you at first glance. For example, if you were to spend two minutes viewing each webpage, with 641 webpages that amounts to 21 hours of viewing time. Spread out over three months that would mean viewing

[10] The cache and browsing history of the web browser was cleared after each participant's session, so the duplicated bookmarks occurred naturally (i.e., not as the result of cached history influencing participants).
[11] You may still find some duplicate content, however, because I was only able to eliminate duplicate URLs, not duplicate content (such as might be the case for websites that buy and publish content from syndicate content providers). Hopefully this won't happen very much.

about 7 pages per day, totaling 14 minutes per day. However, I'm fairly confident that once you get started you'll come up with your own system for assessing the webpages and it won't take as long as you might think. It's possible that some of the pages may only need a few seconds of your attention. Also, you can expect to get faster at assessing as you proceed through each task. Overall, I am anticipating that this project will take you about three months (which would mean having it back to me sometime around the first or second week of September). If you need more time, that is completely fine with me. Just let me know whenever you can. ***The expert assessments are a critical part of my dissertation and I greatly appreciate your time and commitment to this. I will do whatever I can to help make this process go smoothly for you.*** I will check in with you from time to time over the course of the next three months to see what questions you might have and to troubleshoot any issues that may come up.

<u>Your expert assessments:</u>

The workbook is organized so that there is a tab for each task and then a tab with the list of URLs associated with it. On each worksheet of URLs, you will see there are seven columns:

- **1st column is "orig_ordr"** – please DO NOT change this item number under any circumstance or dissociate it from the link in that same row, as this is the only way I can match up your assessments with my raw data. For your purposes, the numbers in this column do not reflect any kind of prioritization or other useful meaning, because I randomized the participants and rotated the tasks. If you want to number the items or otherwise prioritize the list for your own purposes, feel free to add a new column or columns to do that. The most important thing is that the "orig_ordr" numbers stay associated with their URLs.

- **2nd column is the URL –** you should be able to just click or control-click on each URL and Excel should take you to your default browser and open up to the site. It might help to open URLs in Internet Explorer, since that was the browser I had to use for the study. If a particular site doesn't open, or something doesn't seem right about it, make a note in the column marked "notes". If you find this happens more than 3 or 4 times, send me an email so I can rectify it. Since these URLs were all accessed within the past eight weeks, the content shouldn't have changed since then. For Wikipedia pages, I've checked their history to make sure no major revisions occurred. You might find one or two of the pdfs don't work – just let me know and I can find them and send them to you. I had this problem myself and did my best to re-create the links for you.

- **3rd through 6th columns are relevance judgment criteria** – the headings are *notrel* (not relevant), *somrel* (somewhat relevant), *rel* (relevant), and *veryrel* (very relevant). You can mark an X in the cell under your judgment for each URL. Each URL should only have one judgment.

- **7th column is for comments and notes** – this is space for you to make whatever notes or comments you'd like me to see, about anything regarding that particular URL.

Overall, I am relying on you to apply to this project your many decades of expertise around general and specific topics related to personal finance and money. In other words, it is fine to be stringent in your assessments of the information. While I've asked participants to imagine *themselves* in the three scenarios, what I'm asking you to do is to evaluate what information would be best for people *other than yourselves* in those circumstances, which is a slight nuance. The easiest example of this is the student loan scenario. In that scenario, the participant is told that she or he is to act like they know very little about student loans. As my expert, I'm asking you to evaluate the webpages you read in terms of the information that would best meet the needs of a person who knows very little about student loans.

At the end of the project please feel free to share whatever you'd like to share with me about how you went about the assessment process.

REFERENCES

Agarwal, S., & Mazumder, B. (2013). Cognitive abilities and household financial decision-making. *American Economic Journal: Applied Economics, 5*(1), 193-207.

Al-Maskari, A., & Sanderson, M. (2011). The effect of user characteristics on search effectiveness in information retrieval. *Information Processing & Management, 47*(5), 719-729.

Allen, B. L. (1992). Cognitive differences in end user searching of a CD-ROM index. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, Copenhagen, Denmark.

Allen, B. L. (1993). Logical reasoning and retrieval performance. *Library & Information Science Research, 15*(1), 93-105.

Allen, B. L. (1994). Perceptual speed, learning and information retrieval performance. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Allen, B. L. (1998a). Designing information systems for user abilities and tasks: An experimental study. *Online & CD-ROM Review, 22*(3), 139-153.

Allen, B. L. (1998b). Information space representation in interactive systems: Relationship to spatial abilities. *Proceedings of the 3rd ACM Conference on Digital Libraries*, Pittsburgh, PA.

Allen, B. L. (1998c). Visualization and cognitive abilities. *Proceedings of the 34th Clinic on Library Application for Data Processing*, Urbana-Champaign, IL.

Altman, D. G., & Bland, J. M. (1995). Statistics notes: The normal distribution. *BMJ: British Medical Journal, 310*(6975), 298-298.

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

Arguello, J. (2015). Improving aggregated search coherence. *Proceedings of the European Conference on Information Retrieval*, Vienna, Austria.

Aula, A., & Käki, M. (2003). Understanding expert search strategies for designing user-friendly search interfaces. *Proceedings of the IADIS International Conference WWW/Internet 2003,* Algarve, Portugal.

Aula, A., Majaranta, P., & Raiha, K.-J. (2005). Eye-tracking reveals the personal styles for search result evaluation. *Proceedings of the 10th IFIP TC13 International Conference on Human-Computer Interaction*, Rome, Italy.

Azzopardi, L., Kelly, D., & Brennan, K. (2013). How query cost affects search behavior. *Proceedings of the 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Baddeley, A. D. (2007). *Working Memory, Thought, and Action*. New York, NY: Oxford University Press.

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The Multiple-Component Model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). Cambridge: Cambridge University Press.

Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology, 63*(9), 1728-1746.

Baldwin, N. S., & Rice, R. E. (1997). Information-seeking behavior of securities analysts: Individual and institutional influences, information sources and channels and outcomes. *Journal of the American Society for Information Science and Technology, 48*(8), 674-693.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science, 30*(4), 205-214.

Belkin, N. J., Kelly, D., Kim, G., Kim, J.-Y., Lee, H.-J., Muresan, G., Tang, M.-C., Yuan, X.-J., & Cool, C. (2003). Query length in interactive information retrieval. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*.

Berger, S. C., & Messerschmidt, C. M. (2009). Babbling before banking? Online communities and pre-purchase information seeking. *International Journal of Bank Marketing, 27*(6), 446-466.

Bhavnani, S. K. (2001). Important cognitive components of domain-specific search knowledge. *Proceedings of the Text REtrieval Conference (TREC)*, Gaithersburg, MD.

Bhavnani, S. K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems,* Minneapolis, MN.

Biza-Khupe, S. (2014). Credit search. *The Journal of Applied Business and Economics, 16*(3), 116-134.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation, 56*(1), 71-90.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology, 54*(10), 913-925.

Borlund, P. (2016). A study of the use of simulated work task situations in interactive information retrieval evaluations. *Journal of Documentation, 72*(3), 394-413.

Bota, H., Zhou, K., & Jose, J. M. (2016). Playing your cards right: The effect of entity cards on search behaviour and workload. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, Carrboro, NC.

Brennan, K., Kelly, D., & Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. *Proceedings of the Information Interaction in Context Conference*, Regensburg, Germany.

Bricker, J., Dettling, L., Henriques, A., Hsu, J., Moore, K., Sabelhaus, J., Thompson, J., & Windle, R. (2014). *Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances*. Washington, DC: Board of Governors of the Federal Reserve System.

Brown, S., Taylor, K., & Wheatley-Price, S. (2005). Debt and distress: Evaluating the psychological cost of credit. *Journal of Economic Psychology*, *26*(5), 642-663.

Buscher, G., Cutrell, E., & Morris, M. R. (2009). What do you see when you're surfing? Using eye tracking to predict salient regions of Web pages. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, Massachusetts.

Buscher, G., Dumais, S., & Cutrell, E. (2010). The good, the bad, and the random: An eye-tracking study of ad quality in Web search. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, Geneva, Switzerland.

Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology, 56*(10), 1050-1061.

Carroll, J. B. (1993). *Human Cognitive Abilties: A Survey of Factor-analytic Studies*. NY: Cambridge University Press.

Case, D. O., Andrews, J. E., Johnson, J. D., & Allard, S. L. (2005). Avoiding versus seeking: The relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. *Journal of Medical Library Association, 93*(3), 353-362.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems, 27*(6), 1065-1073.

Cattell, R. B. (1966). Stephanos, studies in psychology: Presented to Cyril Burt. In C. Banks (Ed.), *Studies in Psychology*. New York: Barnes & Noble.

Costello, K. L. (2015). *Social information behaviors in the context of chronic kidney disease: Information seeking and disclosure in online support groups.* (Unpublished Dissertation), University of North Carolina at Chapel Hill, NC.

Crystal, A., & Greenberg, J. (2006). Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology, 57*(10), 1368-1382.

DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*(3), 1616-1641.

Dik, B. J. (2007). Ability tests. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (Vol. 3, pp. 2-5). Thousand Oaks, CA: SAGE Publications, Ltd.

Downey, D., Dumais, S., Liebling, D., & Horvitz, E. (2008). Understanding the relationship between searchers' queries and information goals. *Proceedings of the 17th ACM Conference on Information and Knowledge Management*.

Drentea, P., & Lavrakas, P. J. (2000). Over the limit: The association among health, race and debt. *Social Science and Medicine, 50*(4), 517-529.

Dumais, S. T., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for web search*. Proceedings of the 3rd Symposium on Information Interaction in Context*, New Brunswick, NJ.

Edwards, A. (2015). *Engaged or frustrated? Disambiguating engagement and frustration in search.* (Unpublished Dissertation), University of North Carolina, Chapel Hill, NC.

Edwards, A., Kelly, D., & Azzopardi, L. (2015). The impact of query interface design on stress, workload and performance. *Proceedings of the 37th European Conference on IR Research*, Vienna, Austria.

Eickhoff, C., Gwizdka, J., Hauff, C., & He, J. (2017). Introduction to the special issue on search as learning. *Information Retrieval Journal, 20*(5), 399-402.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976a). *Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976b). Manual for Kit of Factor-Referenced Cognitive Tests. Princeton, NJ: Educational Testing Service.

Ericsson, K. A., & Simon, H. A. (1996). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, *324*(7337), 573-577.

Fidel, R. (1985). Moves in online searching. *Online Review, 9*(1), 61-74.

Field, A. (2005). *Discovering Statistics Using SPSS: And Sex, Drugs and Rock 'n' Roll* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.

Field, A. (2009). *Discovering Statistics Using IBM SPSS Statistics: And Sex and Drugs and Rock 'n' Roll* (3rd ed.). London: Sage Publications Ltd.

Francis, G., Neath, I., & VanHorn, D. (2008). CogLab on a CD: Version 2.0. Belmont, CA: Wadsworth.

Freund, L., & Toms, E. G. (2006). Enterprise search behaviour of software engineers. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, WA.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences, 110*(28), 11267-11271.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489.

Granka, L., Feusner, M., & Lorigo, L. (2008). Eye monitoring in online search. In R. Hammoud (Ed.), *Passive Eye Monitoring: Algorithms, Applications and Experiments* (pp. 347-372): Springer.

Guo, Q., & Agichtein, E. (2012). Beyond dwell time: Estimating document relevance from cursor movements and other post-click searcher behavior. *Proceedings of the 21st International Conference on World Wide Web*, Lyon, France.

Gwizdka, J. (2009). What a difference a tag cloud makes: Effects of tasks and cognitive abilities on search results interface use. *Information Research, 14*(4).

Gwizdka, J. (2010). Distribution of cognitive load in Web search. *Journal of the American Society for Information Science and Technology, 61*(11), 2167-2187.

Gwizdka, J. (2013). Effects of working memory capacity on users' search effort. *Proceedings of the International Conference on Multimedia, Interaction, Design and Innovation*, Warsaw, Poland.

Gwizdka, J. (2014). Characterizing relevance with eye-tracking measures. *Proceedings of the 5th Symposium on Information Interaction in Context*, Regensburg, Germany.

Gwizdka, J. (2017). I can and so I search more: Effects of memory span on search behavior. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, Oslo, Norway.

Han, S.-H., & Kim, M.-S. (2004). Visual search does not remain efficient when executive working memory Is working. *Psychological Science, 15*(9), 623-628.

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

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of learning and motivation* (Vol. 22, pp. 193-225): Elsevier.

Hassan, A., White, R. W., Dumais, S. T., & Wang, Y.-M. (2014). Struggling or exploring? Disambiguating long search sessions. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, New York, NY.

Hastings, J., Madrian, B., & Skimmyhorn, W. (2013). Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, 347-373.

Hawk, W. B., & Wang, P. (1999). Users' interaction with the World Wide Web: Problems and problem-solving. *Proceedings of the 62nd ASIS Annual Meeting*, Washington, DC.

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

Hjørland, B. (2010). The foundation of the concept of relevance. *Journal of the American Society for Information Science and Technology, 61*(2), 217-237.

Holmqvist, K., & Nystrom, M. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, United Kingdom: Oxford University Press.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, *33*, 337-346.

Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht, The Netherlands: Springer.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management, 36*(2), 207-227.

Jex, H. (1988). Measuring mental workload: Problems, progress, promises. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 52, pp. 5-39): Elsevier.

Jiang, J., He, D., & Allan, J. (2014). Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, Queensland, Australia.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. *Proceedings of the 28th Annual International*

ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil.

Jump$tart. (1997). Survey of Financial Literacy Among High School Students. Washinton, DC: Jump$tart Coalition for Personal Financial Literacy.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

Kelly, D. (2009). *Methods for Evaluating Interactive Information Retrieval Systems with Users*. Hanover, MA: now Publishers, Inc.

Kelly, D., Arguello, J., Edwards, A., & Wu, W.-c. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. *Proceedings of the 2015 International Conference on the Theory of Information Retrieval*, Northampton, MA.

Kelly, D., & Azzopardi, L. (2015). How many results per page? A study of SERP size, search behavior and user experience. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, Chile.

Kim, J., Sorhaindo, B., & Garman, E. T. (2006). Relationship between financial stress and workplace absenteeism of credit counseling clients. *Journal of Family and Economic Issues, 27*(3), 458-478.

Klöckner, K., Wirschum, N., & Jameson, A. (2004). Depth-and breadth-first processing of search result lists. *Proceedings of the CHI'04 Extended Abstracts on Human Factors in Computing Systems*, Vienna, Austria.

Kuhlthau, C. C. (1997). The influence of uncertainty on the information seeking behavior of a securities analyst. *Proceedings of the International Conference on Information Seeking in Context*, Tampere, Finland.

Kuhlthau, C. C. (1999). The role of experience in the information search process of an early career information worker: Perceptions of uncertainty, complexity, construction, and sources. *Journal of the Association for Information Science and Technology, 50*(5), 399.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433.

Li, Y., Baldassi, M., Johnson, E. J., & Weber, E. U. (2013). Complementary cognitive capabilities, economic decision making, and aging. *Psychology and Aging, 28*(3), 595-613.

Lin, J. T., Bumcrot, C., Ulicny, T., Lusardi, A., Mottola, G., Kieffer, C., & Walsh, G. (2016). National Financial Capability Study (pp. 36). Washington, DC: FINRA Investor Education Foundation.

Littlejohn, S. W. (1996). *Theories of Human Communication*. Belmont: Wadsworth Publishing Company.

Lusardi, A., & Tufano, P. (2015). Debt literacy, financial experiences, and overindebtedness. *Journal of Pension Economics and Finance, 14*(Special Issue 04), 332-368.

MacFarlane, A., Albrair, A., Marshall, C., & Buchanan, G. (2012). Phonological working memory impacts on information searching: An investigation of dyslexia. *Proceedings of the Information Interaction in Context*, Amsterdam, The Netherlands.

Maidel, V. (2014). *Searching as thinking: The role of cues in query reformulation.* (Unpublished Dissertation), Syracuse University, Syracuse, NY.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. New York: Cambridge University Press.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM, 49*(4), 41-46.

Consumer protections in financial services: Past problems, future solutions, United States Senate, 111th Congress, 1st Sess. (2009).

Megaw, T. (2005). The definition and measurement of mental workload. In E. N. Corlett & J. R. Wilson (Eds.), *Evaluation of Human Work* (3rd ed., pp. 525-551). Boca Raton, FL: Taylor & Francis.

Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience, 8*, 1063.

Mezick, E. M. (2002). Investing wisely: Individual investors as information seekers. *Journal of Business & Finance Librarianship, 7*, 3-23.

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science, 48*(9), 810-832.

Morita, M., & Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland.

Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive strategies in web searching. *Proceedings of the 5th Conference on Human Factors & the Web*, Gaithersburg, MD.

O'Brien, H. L., Dickinson, R., & Askin, N. (2017). A scoping review of individual differences in information seeking behavior and retrieval research between 2000 and 2015. *Library & Information Science Research, 39*(3), 244-254.

O'Brien, H. L., & Toms, E. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology, 59*(6), 938-955.

Oh, S.-H., & Kim, M.-S. (2004). The role of spatial working memory in visual search efficiency. *Psychonomic Bulletin and Review, 11*(2), 275-281.

Ondrusek, A. L., Ren, X., & Yang, C. (2017). A content analysis of strategies and tactics observed among MLIS students in an online searching course. *Journal of Education for Library and Information Science, 58*(3), 141-159.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38*(1), 63-71.

Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows* (3rd ed.): McGraw-Hill Education.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication, 12*(3), 801-823.

Peterson, M. S., Beck, M., & Wong, J. H. (2008). Were you paying attention to where you looked? The role of executive working memory in visual search. *Psychonomic Bulletin & Review, 15*(2), 372-377.

Peterson, R. L. (2007). Affect and financial decision-making: How neuroscience can inform market participants. *The Journal of Behavioral Finance, 8*(2), 70-78.

Qvarfordt, P., Golovchinsky, G., Dunnigan, T., & Agapie, E. (2013). Looking ahead: Query preview in exploratory search. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland.

Rasch, D., & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science, 46*, 175-208.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rees, A. M., & Saracevic, T. (1966). *The Measurability of Relevance*. Western Reserve University: Center for Documentation & Communication Research.

Reeve, C. (2007). Cognitive Abilities. In S. Rogelberg (Ed.), *Encyclopedia of Industrial and Organizational Psychology* (pp. 77-80). Thousand Oaks, CA: SAGE Publications, Inc.

Rokhlenko, O., Golbandi, N., Lempel, R., & Leibovich, L. (2013). Engagement-based user attention distribution on web article pages. *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, Paris, France.

Rouet, J.-F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Interacting with Computers, 15*(3), 409-428.

Saldana, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: SAGE Publications.

Salthouse, T. A. (2017). Comparable consistency, coherence, and commonality of measures of cognitive functioning across adulthood. *Assessment*, 11.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science, 26*(6), 321-343.

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 1915-1933.

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 2126-2144.

Savolainen, R. (2016). Information seeking and searching strategies as plans and patterns of action: A conceptual analysis. *Journal of Documentation, 72*(6), 1154-1180.

Savolainen, R., & Kari, J. (2006). User-defined relevance criteria in web searching. *Journal of Documentation*, *62*(6), 685-707.

Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology, 29*, 3-48.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management*, *26*(6), 755-776.

Schwandt, T. A., Lincoln, Y. S., & Guba, E. G. (2007). Judging interpretations: But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation*(114), 11-25.

Shovon, M. H., Nandagopal, D., Du, J. T., Vijayalakshmi, R., & Cocks, B. (2015). Cognitive activity during web search. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, Chile.

Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum, 33*(1), 6-12.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201-292.

Steichen, B., Carenini, G., & Conati, C. (2013). User-adaptive information visualization: Using eye gaze data to infer visualization tasks and user cognitive abilities. *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, Santa Monica, CA.

Steichen, B., Conati, C., & Carenini, G. (2014). Inferring visualization task properties, user performance, and user cognitive abilities from eye gaze data. *ACM Transactions on Interactive Intelligent Systems, 4*(2), 1-29.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 23*, 257-285.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson/Allyn & Bacon.

Thatcher, A. (2006). Information-seeking behaviours and cognitive search strategies in different search tasks on the WWW. *International Journal of Industrial Ergonomics*, *36*(12), 1055-1068.

Thomas, P., Scholer, F., & Moffat, A. (2013). What users do: The eyes have it. *Proceedings of the 9th Asia Information Retrieval Societies Conference*, Singapore.

Tobii. (2016). Tobii Studio User's Manual (Vol. Version 3.4.5): Tobii AB.

Toker, D., Conati, C., Carenini, G., & Haraty, M. (2012). Towards adaptive information visualization: On the influence of user characteristics. *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization*, Montreal, Canada.

Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess Web pages for information seeking. *Journal of the American Society for Information Science and Technology, 56*(4), 327-344.

Toms, E. G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., Dawe, E., & MacNutt, A. (2008). Task effects on interactive search: The query factor. In N. Fuhr, J. Kamps, M. Lalmas & A. Trotman (Eds.), *Focused Access to XML Documents: 6th International Workshop of the Initiative for the Evaluation of XML Retrieval* (pp. 359-372). Dagstuhl, Germany: Springer Berlin.

Turpin, L., Kelly, D., & Arguello, J. (2016). To blend or not to blend? Perceptual speed, visual memory and aggregated search. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pisa, Italy.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics, 106*(4), 1039-1061.

Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. *Proceedings of the Content-Based Multimedia Information Access*, Paris, France.

Vakkari, P. (2003). Task-based information seeking. *Annual Review of Information Science and Technology, 37*, 413-464.

Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing & Management, 39*(3), 445-463.

van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25*, 325-331.

Veerasamy, A., & Heikes, R. (1997). Effectiveness of a graphical display of retrieval results. *SIGIR Forum, 31*, 236-245.

Villa, R., & Halvey, M. (2013). Is relevance hard work? Evaluating the effort of making relevant assessments. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval,* Dublin, Ireland.

Weinreich, H., Obendorf, H., Herder, E., & Mayer, M. (2008). Not quite the average: An empirical study of Web use. *ACM Transactions on the Web, 2*(1), 1-31.

White, R., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining,* Barcelona, Spain.

White, R. W., & Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, The Netherlands.

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior, 26*(3), 474-481.

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology, 55*(3), 246-258.

Wildemuth, B. M., & Freund, L. (2009). Search tasks and their role in studies of search behaviors. *Proceedings of the 3rd Annual Workshop on Human Computer Interaction and Information Retrieval*, Washington, DC.

Wildemuth, B. M., & Freund, L. (2012). Assigning search tasks designed to elicit exploratory search behaviors. *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*.

Wilson, M. L., Sharon, N., Maior, H. A., Midha, S., Craven, M. P., & Sharples, S. (2018). Mental workload as personal data: designing a cognitive activity tracker. *http://eprints.nottingham.ac.uk/50630/*.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*, 249-270.

Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, *14*(4-8), 808-830.

Xie, I., & Benoit, E. (2013). Search result list evaluation versus document evaluation: Similarities and differences. *Journal of Documentation, 69*(1), 49-80.

Xie, I., Benoit, E., & Zhang, H. (2010). How do users evaluate individual documents? An analysis of dimensions of evaluation activities. *Information Research: An International Electronic Journal, 15*(4), n4.

Xie, I., & Joo, S. (2010). Tales from the field: Search Strategies applied in Web searching. *Future Internet, 2*(3), 259.

Xu, Y., & Chen, Z. (2006). Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology, 57*(7), 961-973.

Yilmaz, E., Verma, M., Craswell, N., Radlinski, F., & Bailey, B. P. (2014). Relevance and effort: An analysis of document utility. *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management*, Shanghai, China.

Zhang, X., Anghelescu, H. G., & Yuan, X. (2005). Domain knowledge, search behaviour, and search effectiveness of engineering and science students: An exploratory study. *Information Research: An International Electronic Journal, 10*(2), n2.