ABSTRACT

Title of Document:	USERS' SPATIAL ABILITIES AFFECT INTERFACE USABILITY OUTCOMES
	Susan Grace Campbell, Doctor of Philosophy, 2011
Directed By:	Professor Kent L. Norman, Department of Psychology

The usability of a system depends both on inherent characteristics of the system and on its users. This paper argues that a major source of differences among users is variations in spatial ability, but that variations in different types of spatial ability affect components of usability differently. In two experiments, I investigated a simple model of the relationships of spatial visualization ability, spatial orientation ability, and spatial working memory with the usability constructs of efficiency and effectiveness. Both experiments used Wikipedia search as a representative information search task. The first experiment used a desktop computer interface, and the second experiment used a pair of mobile devices with widely different screen sizes. Better spatial orientation ability corresponded to faster performance on efficiency across devices. Better spatial visualization ability corresponded to slower performance on larger screens, but faster performance on smaller screens. Better spatial visualization ability also predicted better effectiveness in both experiments.

These results suggest that spatial ability is a useful way to characterize users and to improve usability testing, and that its effects vary in systematic ways depending on characteristics of the tested interface and on which metrics are chosen.

USERS' SPATIAL ABILITIES AFFECT INTERFACE USABILITY OUTCOMES

By

Susan Grace Campbell

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Kent L. Norman, Chair
Professor Paul J. Hanges
Professor Mo Wang
Professor Michael F. Bunting
Professor Kenneth R. Fleischmann

Dedication

To everyone who assured me that finishing was a matter of when, not if.

Acknowledgements

I am deeply grateful to my entire committee for their comments and suggestions throughout the process of writing this dissertation. I am especially grateful to my advisor, Kent Norman, for his support and guidance. Special thanks go to Kenneth Fleischmann, who stepped up in a difficult situation and made it possible for me to finish. I also thank Jennifer Preece, who provided great feedback and support as a member of my committee until circumstances intervened.

The research assistants in the Laboratory for Automation Psychology put in many hours and much thought making sure this research was carried out to a high standard. During the first experiment, our lab assistants were Victor Quach, Dustin Richardson, and Raz Goldin. During the second experiment, they were Johnny Wu and Mihai Sirbu. Other lab assistants provided great feedback and suggestions along the way. My colleagues in the Laboratory for Automation Psychology, Benjamin Smith and Walky Rivadeneira, were always there to help, and all of the students and faculty of the Cognitive Area provided useful feedback at one stage or another that made this work better. My colleagues at the Center for Advanced Study of Language were wonderfully supportive and understanding of my wandering off at odd times to work on my dissertation, and I owe them a great deal as well. Finally, I thank my family and friends, and especially John, for being there for me and keeping me sane throughout this process.

I could not have done this alone, and I appreciate all of the help and support I have received over the past few years.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Whether an object is usable depends on two things: the object itself, and the person attempting to use it. An interface is only "usable" in the sense that a particular user can use it, and what may be usable for one person may be incomprehensible to another. Usability testing adds two complications that affect the testing outcome: the tasks the person is trying to accomplish and the metrics used to assess performance. Thus, an interface is only usable if a particular user can use it to fulfill particular goals, and assessment depends on the metrics used. Because the tasks and metrics are the same across users, however, the most variability within a particular usability test will come from the users themselves. The implication for testing interfaces is that the users we test with can determine the success or failure of the test.

What determines whether a particular person can use an interface? Different answers have different implications for the design of interfaces and of usability tests. I intend to argue that the most important individual differences in interface usability are cognitive and experiential; specifically, the best predictors of performance with interfaces are spatial ability and experience with related interfaces. In order to make those differences explicit, however, I first need to explain what I mean by usability testing, and how the assessment of usability is performed. Then I will discuss non-cognitive individual differences that may affect usability and will explain why these are not sufficient to explain differences in performance. Furthermore, I focus on objective measures of performance rather than subjective measures of satisfaction, which may be more influenced by non-cognitive factors. Finally, I will discuss the

1

cognitive individual differences I pursued in this series of experiments and the theoretical reason I chose these particular differences.

Though several of these relationships have been investigated before, I plan to integrate the disparate results to provide a framework for considering the role of cognitive abilities in different aspects of the constructs that comprise usability.

Usability Testing

Software does not spring into the world fully formed, and any system that must interface with a human user requires testing to determine whether it is usable. Can the person for whom the system is supposed to work actually use it to accomplish his or her goals? Usability depends on a multitude of factors, including the target population, the goal of the software, and the circumstances of use. Tests, therefore, must try to replicate or simulate these factors in order to make the best estimate of future performance.

User testing

There are multiple ways of testing any given interface, and the selection of a particular method depends on the goal of the testing as well as the logistical constraints of the testing environment. User testing is one such method, in which representative users interact with the system in some controlled way (Nielsen, 1992).

Other ways of testing interfaces include heuristic evaluation (Nielsen & Molich, 1990), questionnaires, and log analyses (Norman & Panizzi, 2006). These testing methods do not produce the same kind of performance data as a user test, though they may be valuable for evaluating system usability in different phases of development. Heuristic evaluation has the advantage of being quick and generally

less expensive to perform than user testing, but the outcomes appear to depend heavily on the experience level of the testers, as John and Marks found in 1997 when they compared experienced evaluators to graduate students who were given the same instruction as novice evaluators. Both questionnaires and log file analyses, by themselves, oversample frequent users and tend to lack information about the users' goals (Norman & Panizzi, 2006).

A standard user test involves recruiting people who are as representative as is practically possible of the target population of users and asking them to use the software. In this kind of testing, the system is evaluated rather than the users, and all user errors are considered to be problems with the system (Nielsen, 1992). User tests tend to involve both performance measures, like whether a particular task can be accomplished using the interface, and qualitative measures, like verbal protocol data on how a particular user accomplished the task. Logs and questionnaires may be used as part of a user test, in order to increase the amount and type of information gleaned.

User Selection

In a user test where the performance of actual users is required, those users must be selected in some way. Ideally, when one is designing an interface to be used by a limited population (such as Air Traffic Controllers or teachers in a particular school district), one would sample participants for the user test directly from that population (Nielsen, 1992). For applications designed for a general population, however, or an ill-defined group (like photographers or college students), that sampling method is not practical.

There are many ways to choose users for a user test. Sometimes, if a particular set of disabilities is the focus, a panel of disabled users may be constructed (e.g. Petrie, Hamilton, & King, 2004). A second approach might include finding pre-existing groups of people, like credit union members or older adults in community centers, and treating the target population as composed of such groups (Dearden, Lauener, Slack, Roast, & Cassidy, 2006). In other cases, the selection is generally based on scenario factors, like what a user is likely to want to use the interface for, or demographic factors, like age and gender (Kujala & Kauppinen, 2004).

There is little evidence that cognitive abilities, rather than cognitive disabilities (where appropriate), are taken into account when composing user panels, though this may be a practice that is common in industry but not published on. If this is the case, there is room to improve the predictive power of such panels by including individual difference measures for selection and analysis.

Metrics for Assessing Performance

The outcomes of a usability test for software are broken into three broad categories in the International Standards Organization (ISO) standard 924: satisfaction, effectiveness, and efficiency (Abran, Khelifi, Suryn, & Seffah, 2003). According to the standard, these constructs are defined only for the tests and users that were actually performed, and are not guaranteed to generalize to the general user population. In practice, however, one expects the results to give some indication of the general usability of the system.

Satisfaction is a subjective measure of the user's opinion about a particular interaction or interface, while effectiveness and efficiency are based on the

measurement of user performance. Effectiveness refers to the user's ability to achieve goals using the software, while efficiency refers to the time and effort it takes to achieve goals (Abran et al., 2003; Bevan & MacLeod, 1994). The measurement and even definitions of these constructs vary, however.

Satisfaction. The construct of satisfaction refers to the subjective feeling of enjoyment or accomplishment that a person gets from an interaction. Satisfaction is generally measured using survey instruments, such as the Questionnaire for User Interface Satisfaction (QUIS; Norman, Shneiderman, & Harper, 2003). The QUIS uses a hierarchically organized set of Likert scale type questions to assess a user's satisfaction with a given interface or interaction. Competitors to the QUIS include the Software Usability Measurement Inventory (SUMI), which does not have a hierarchical structure like the QUIS and is billed as an inventory rather than a questionnaire (Kirakowski, 1996) and the System Usability Scale (SUS), which is a very short inventory designed for quick assessment of system usability (Brooke, 1996). The SUS is frequently used because of its short length and ease of administration, but there is no single measure of satisfaction that is universally accepted.

Abdinnour-Helm, Chaparro, and Farmer (2005) adapted a questionnaire instrument called the End User Computing Survey (EUCS) to measure satisfaction with websites. The main construct that the EUCS was designed to measure was satisfaction, but by including questions about perceived efficiency and effectiveness, the authors obtained correlations between the EUCS and success (r=0.33) and task duration (r=-0.28, where shorter durations are better). They did find, however, that

the factor structure of the test supported a unitary construct of site usability underlying responses on all parts of the questionnaire, suggesting that perhaps perceived efficiency and perceived effectiveness do correspond to some aspects of satisfaction.

Rozell and Gardner (2000) found that there were many predictors of satisfaction, but most of them depended on constructs such as causal attribution, computer anxiety, and perception of competence that depended on context rather than stable individual differences.

Frøkjaer, Hertzum, and Hornbaek (2000) measured satisfaction in a novel and practical way by determining which interface in a given set users preferred to continue using after the test was over. They further found that user satisfaction was greatest for the system which provided the most choice in interaction, rather than the system on which that user had performed the best objectively.

Kissel (1995) also found that objective measures of efficiency and subjective measures of satisfaction did not correspond to each other, though he suggested that more computer experience predicted a better correspondence between preference and objective measures of performance. His study was limited, however, by the fact that the best-performing interface may have been the one which was already the most familiar to users with more computer experience.

There is little evidence for systematic cognitive individual differences in satisfaction. In a cross-cultural study of website satisfaction, Simon (2001) found that women reported lower levels of satisfaction than men, especially in the United States and other more individualistic cultures. This finding may be related to Knight and

6

Pearson's (2005) finding that women reported higher levels of computer anxiety than men, though the genders did not differ on other measures including performance. The difference may also be related to Rozell and Gardner's (2000) observation that women perceive their competence with computers more negatively than men do; Schumacher and Morahan-Martin (2001) found that women rated their own competence lower than men did even when there appeared to be no differences in objective measures between the genders.

Effectiveness. The construct of effectiveness refers to the ability of a user to accomplish whatever goal they had in using the system. Measures of effectiveness are measures of accuracy and the ability of users to accomplish their goals; they also tend to be based on qualitative performance differences, rather than speed differences. Like satisfaction, effectiveness has no single universally accepted measure. Bevan and MacLeod (1994) calculated effectiveness by multiplying the percentage of subgoals completed by the percentage correctness of the completed goals. Frøkjaer, Hertzum, and Hornbaek (2000), however, defined effectiveness as the quality of the solution users reached.

Van Welie, van der Veer, and Eliëns (1999) suggested that the construct of effectiveness could be broken down into memorability and rate of user error, bringing the definition closer to Nielsen's (1992) five-part set of usability goals for ideal usability engineering, along with satisfaction, efficiency of use, and learnability.

Effectiveness is not the most common measure of WWW usability; efficiency is assessed much more frequently. This difference may have something to do with the construct of effectiveness being more difficult to define, or web interfaces being

considered so simple that no difference in effectiveness is expected between interfaces or between users.

If effectiveness is the measure of power and accuracy, then it follows that efficiency is generally the measure of speed and complexity.

Efficiency. The construct of efficiency refers to the amount of resources (including time) required to accomplish a particular task. Measures of efficiency tend to depend on the number of steps required to perform a particular task, or the amount of time it takes a user to perform that task. These measures are roughly analogous to measures of response time in behavioral research.

Bevan and MacLeod (1994) calculate efficiency, however, by dividing their effectiveness measure (percentage of tasks completed multiplied by percent of those tasks completed correctly) by the total time required. This approach makes it difficult to dissociate efficiency and effectiveness by making the measures directly dependent on each other, even though the constructs are supposed to be separable characteristics of the interface.

Frøkjaer, Hertzum, and Hornbaek (2000), on the other hand, measured efficiency simply as the time it took users to complete the task. Van Welie, van der Veer, and Eliëns (1999), in a survey of usability constructs, break efficiency down into completion time and learning time. These two constructs are the final two parts of Nielsen's (1992) set of system goals for usability engineering, but learnability seems only to weakly correspond to efficiency of use.

This disagreement about the best way to measure usability constructs leads to disagreements about whether particular measures actually correspond to the

constructs that they are intended to measure. For instance, Smith (1996) uses accuracy as a measure of efficiency, along with a measure of revisiting nodes and a measure of deviations from an optimal path. However, in order to be more directly comparable to the measures presented above, accuracy would be an effectiveness measure.

The three usability constructs are separable, but there is some disagreement about how related they are. Nielsen and Levy (1994) found in a meta-analysis that there was a positive association between objective measures like efficiency or effectiveness and subjective measures like satisfaction, but that the two sets of constructs were not close to identical (r=0.46). Frøkjaer, Hertzum, and Hornbaek (2000), however, found that the correlations among the three measures were small. Abdinnour-Helm, Chaparro, and Farmer (2005) again found significant correlations between their satisfaction instrument (which included questions about perceived efficiency and perceived effectiveness) and objective performance measures. In a more recent meta-analysis, Hornbaek and Law (2007) found that the correlation between efficiency and effectiveness tended to be around r=0.25, while the correlations between either efficiency or effectiveness and satisfaction tended to be slightly lower, suggesting that subjective and objective measures of usability are related to each other, but that they are not interchangeable.

The main reason that the relationships among the three usability constructs are difficult to define is that there is no consensus about what the constructs mean or how to measure them.

Non-cognitive Individual Differences Among Users

Users differ along a theoretically infinite set of dimensions, many of which are not relevant to their performance with any particular interface. In order to decide which dimensions were relevant for predicting performance, I looked at several non-cognitive ways that users have been documented to differ in their interaction with interfaces. These differences include level of experience with computers, gender, cognitive styles, and personality.

Experience

A person's previous experience with computers will affect how he or she interacts with computer systems in the future. This effect may be due to attitudinal changes, or it may be due to strategic learning.

Attitudes toward computers are entwined with computer experience, but it is unclear which is the cause and which the effect. Do people who like computers use them more, or do people grow less anxious around computers as they use them? A survey of college students addressed to this question in the late 1990s could differentiate objective computer experience (time spent using computers) from subjective computer experience (valence of experiences), but could not disentangle those subjective experiences from attitudes (Smith, Caputi, and Rawstone, 2000).

Further complicating the picture is the finding that computer experience may be a moderator of the relationship among usability constructs. In a small study with college students in the early 1990s, Kissel (1995) found that more computer experience predicted more correspondence between satisfaction with a system and

performance with that system. This finding suggests that computer experience may improve one's self-evaluation of performance.

Anxiety may also mediate the relationship between computer experience and usability ratings. Hackbarth and colleagues found that experience with a spreadsheet program significantly predicted ratings of ease-of-use for that program, but that computer anxiety fully mediated the relationship (Hackbarth, Grover, and Yi, 2003).

The best experiential predictor of performance with a particular interface is knowledge of the domain that interface works in; navigational aids in hypertext are more helpful if the person using the system can map them onto an existing knowledge structure. McDonald and Stevenson (1998A, 1998B) found that the best predictor of performance with their hypertext systems was knowledge of the domain, and the navigational aids they included had considerably smaller impact.

As people use computer systems (or anything, really), they learn strategies for dealing with the domain. These learned strategies may differ among people with different experiences, and may be adaptive or non-adaptive in new contexts.

For example, in a small study using graduate students, predominantly using hypermedia applications outside the study led to more willingness to jump around a hypertext system, rather than a more linear stepwise exploration pursued by those who primarily used spreadsheets and word processing programs (Reed, Oughton, Ayersman, Ervin, & Giessler, 2000). In a study investigating strategy use and spatial abilities, however, Campbell and Norman (2007) found no evidence that spatial ability predicted which strategies participants used (though it did predict adherence to a particular strategy).

11

In this series of experiments, I have assessed computer experience with self-report questions. The focus, however, is on the cognitive abilities that affect the interactions making up that experience.

Personality

Whether attitudes towards computers or anxiety around computers are caused by particular experiences or are temperamental, they affect people's willingness to use computers and how they use computers.

State anxiety in the presence of computers or at the thought of using computers (as opposed to trait anxiety, which is a disposition to be anxious regardless of context), is commonly called computer anxiety (Chua, Chen, & Wong, 1999; Brosnan, 1998; Beckers & Schmidt, 2001). Unfortunately, the factor structures of instruments used to measure computer anxiety vary greatly, and it is unclear whether they all measure the same underlying construct (Chua, Chen, & Wong, 1999).

Despite the scale uncertainty, however, results generally suggest that certain types of computer experience cause computer anxiety. A more complex model put forth by Beckers & Schmidt (2001) suggested that computer literacy (which was correlated with general self-efficacy) predicted physical arousal and affective changes in the presence of computers, which jointly predicted negative and positive beliefs about computers. In a follow-up study, they found that, in general, computer experiences appeared to cause computer anxiety rather than the other way around (Beckers & Schmidt, 2003).

Another possibility is that computer anxiety comes from the perception of past experience rather than actual past experience. The higher a person's perceived

knowledge of computers, the less anxious that person will be around computers (Anderson, 1996).

Computer self-efficacy seems to be the opposite of computer anxiety; the perception that one can accomplish tasks effectively with computers drives out anxiety about performance. Brosnan (1998), analyzing a study on the use of database lookup tables, suggested that, rather than just being correlated with self-efficacy, computer anxiety is a predictor of computer self-efficacy, but that self-efficacy, not anxiety, then predicts performance, strategy, and completion time. On the opposite side, however, Hackbarth and colleagues (2003) found that the positive attitudes engendered by extensive computer use were not as important to ease-of-use ratings as anxiety was. Computer self-efficacy, then, may not be as powerful a predictor of success as anxiety.

Certain groups, such as women and older adults, seem to have higher levels of computer anxiety than young men. Laguna and Babcock (1997), in a small-sample survey, found that older adults had higher levels of computer anxiety than younger adults. Whether gender predicts anxiety depends on which computer anxiety scale one is using, however (Chua, Chen, & Wong, 1999). In general, when these results exist, they appear to be due to lower levels of experience in people who show higher levels of computer anxiety. Beckers and Schmidt (2003) found, when validating their model of computer anxiety, that higher levels of computer anxiety in women depended on those women having lower levels of computer experience.

Though computer anxiety appears to affect the relationship between performance and subjective usability, there is little evidence that it affects performance itself.

Cognitive Styles

Cognitive styles are theoretically different from cognitive abilities because they reflect preferences for certain types of processing rather than differences in ability to execute particular kinds of processing. In the domain of hypertext specifically, several different kinds of cognitive styles, as varied as field independence and the Myers-Briggs type indicators, have been used to predict performance and strategy. However, those differences that exist in the ability to navigate hypertext can also be explained by differences in cognitive ability.

In a meta-analysis, Chen and Rada (1996) found no statistically significant results when they looked at whether "active" cognitive styles (internal locus of control and field independence) predicted effectiveness and efficiency, though they found that there was a positive trend that higher levels of "active" cognitive style predicted better effectiveness. Interestingly, these differences were smaller than the differences due to spatial ability. There was no clear theoretical reason in that analysis, however, that internal locus of control, field independence, and active learning style should have been combined into a single construct. Additionally, differences in field independence appear to reflect differences in spatial ability, specifically in mental rotation ability (Ozer, 1987), so at least some of the variance that appeared to be due to cognitive styles may actually have been due to spatial ability.

Graff (2005), who used the Myers-Briggs Type Indicator (MBTI) to predict users' performance on a hypertext system, found no effect for the Analyst/Intuitive dimension (one of the four dimensions of the MBTI) on recall of the information presented, and also found ambiguous effects on subjective usability ratings.

Because it seems that cognitive abilities are more important for predicting performance with computer interfaces than cognitive styles, I chose not to assess any cognitive styles for the studies that follow.

Gender

In the early days of personal computing, computer use was prevalent mostly in male-dominated occupations and hobbies. Gender was thus a good predictor of performance with computers and attitudes towards computing. More recently, however, any gender gap that existed in computer use has narrowed, and gender effects are now generally only observed in anxiety.

Between 1990 and 1997, Schumacher and Morahan-Martin (2001) found that differences between incoming first-year college student women and men in computer experience decreased dramatically, though differences still existed in video game playing and computer programming. The other difference that still existed in 1997 was a difference in perceived competence with computers; women reported lower levels of competence and comfort with computers (Schumacher & Morahan-Martin, 2001). In the early 2000s, Abdinnour-Helm and colleagues did not find any gender differences in satisfaction or perceived usability when they validated the End-User Computing Survey (EUCS).

A survey on changing computer usage demographics in the workplace found in the early 2000s that there was no difference between men and women in any aspect of computer use except anxiety, where younger women reported higher levels of computer anxiety than men or older women (Knight & Pearson, 2005).

Though spatial visualization ability has frequently been linked to performance with computer systems, and though spatial visualization ability is generally higher in men than in women, there do not seem to be systematic differences between men and women in performance. Campbell and Norman (2007) found no difference between genders, for instance, though they did find a large spatial visualization effect and a difference in spatial visualization performance between genders. This finding suggests that observed gender differences might be due to differences in spatial ability.

In fact, Murphy and Lorenz (2001) found that spatial visualization predicted differences in responding to computerized alerts accurately, but that the prediction held only for men, suggesting that gender might moderate spatial visualization effects.

Due to these mixed results, gender is still an important demographic variable to report in usability, but it does not always predict performance with information systems. Furthermore, cognitive individual differences may be sufficient to explain any apparent gender effects which do appear.

Cognitive Individual Differences Among Users

Cognitive individual differences span an almost infinite number of dimensions of performance. One way to split cognitive abilities is to divide them into fluid

abilities (Gf), which involve adapting to situations, but tend to decline with age, and crystallized knowledge (Gc), which includes knowledge about strategies and content (Horn & Blankson, 2005). For reasons of space and parsimony, I have looked only at specific fluid intelligence constructs, such as spatial ability and working memory span. Crystallized intelligence in this domain would involve experience with the interface, which is measured elsewhere, and experience with the specialized vocabulary or knowledge related to a specific task.

Differences in spatial ability lead to differences in performance with computer interfaces. But spatial ability measures are frequently confused or conflated, leading to uncertainty about exactly which constructs are being measured.

Spatial Visualization Ability

Spatial abilities in general are distinct from verbal abilities and are not completely determined by a general intelligence factor (Carroll, 1993). One of the earliest distinctions among different kinds of mental abilities was between spatial and verbal abilities, and a distinction between spatial and mathematical abilities followed soon after (Pellegrino, Alderton, & Shute, 1984).

Even Carroll, however, could not make sense of the way that spatial abilities are separated within the general domain of spatial ability. Visualization and spatial relations are mathematically separable, but in different administrations, the same tests will load on different factors (Carroll, 1993). The accepted definition of spatial visualization tasks is that they involve the manipulation of independent parts of an imagined object (Michael, Guilford, Furchter, & Zimmerman, 1957). However, if tests that appear to measure this construct end up loading on a spatial orientation

17

factor, it is less clear what the systematic differences are. One must thus be careful when choosing spatial visualization measures to emphasize those which are least likely to be strongly related to spatial orientation ability.

Though the cognitive processes that make up spatial operations seem fairly well established, it has historically been difficult to determine how individual differences arise in spatial visualization. Salthouse and colleagues could not find reliable differences in any cognitive attributes like transformational efficiency or processing speed, leading them to hypothesize that spatial visualization differences were due to differences in the ability to perform concurrent storage and processing operations (Salthouse, Babcock, Mitchell, Palmon, & Skovronek, 1990). Considering that working memory is generally defined by the ability to perform concurrent storage and processing, differences in working memory capacity may lead to differences in spatial visualization ability. This conclusion fits with Miyake and colleagues' factor analysis findings, in which they hypothesized that differences among spatial ability constructs might depend on the degree of central executive involvement required (Miyake et al., 2001).

Further support for the idea that spatial visualization reflects reasoning ability comes from the types of computer interactions that spatial visualization ability affects. Vicente, Hayes, and Williges (1987) found that spatial visualization, followed by vocabulary, was the best predictor of efficiency in a hierarchical database search task. They postulated that this result was due to people who were worse at spatial visualization tasks getting lost in the database structure. Norman and Butler (1989) confirmed this finding and suggested that differences were due to a strategy shift;

low-visualization participants spent more time revisiting the home screen of the database than following lower-level connections.

Seagull and Walker (1992) took a different view of getting lost in hypertext, however; they found that differences in spatial visualization ability did not predict early mistakes or starting over, but rather than they became stronger as participants became more accustomed to the interface. They also found no interaction between visualization ability and level of hierarchy (which would be expected if participants were actually getting lost); high-visualization participants were just as much better at finding information across all of the levels of hierarchy they investigated. They suggested that differences were due to differences in processing speed or proceduralization, but those differences could also be accounted for by differences in reasoning ability.

Spatial visualization also seems not to predict performance when decisions about which path to follow are not required. Pak, Rogers, and Fisk (2006) found no effect of spatial visualization on navigating a hierarchical web structure, but they had provided participants with a detailed map and asked them to follow that map, rather than making decisions.

Strategy, however, may be related to spatial visualization ability. Campbell and Norman (2007) did not find that what strategy participants chose depended on their spatial visualization ability, but did find that strategy adherence depended on spatial visualization ability.

Most of these results show an advantage for users with higher spatial visualization ability on efficiency; users with better skills at spatial visualization also

tend to be more efficient at finding information. A few studies have also looked at whether people who are better at spatial visualization are also better at getting the right answer.

Downing, Moore, and Brown (2005) asked experts and non-experts in two fields to find articles relevant to a particular topic in those fields, and found that people with better spatial visualization were faster to find the first article. They also found, however, that experts with better spatial visualization ability found more articles than those with worse spatial visualization ability, while there was no difference in non-experts. Perhaps this finding reflects an inability of non-experts to reason meaningfully about a particular set of articles.

In an attempt to reduce the influence of spatial visualization ability on performance, Zhang and Salvendy (2001) tried to reduce the overhead of a web navigation system by putting all of the links on a single interactive menu on all pages (like a computer menu system). They found that, while spatial visualization ability predicted how many relevant pages a person found in the traditional design, it did not affect how many relevant pages he or she found in the new design. Findings on efficiency with this system were more difficult to interpret because the new design was so much faster than the traditional design for their chosen tasks.

Spatial visualization ability depends on reasoning and on some underlying spatial ability. This dependency initially led me to believe that these spatial visualization results could be explained using working memory capacity, which is associated with reasoning, and lower-level spatial orientation ability instead.

Spatial Orientation Ability

Spatial orientation ability is frequently called mental rotation ability for the simple reason that performing spatial orientation involves speeded mental rotation (Michael et al., 1957). These tasks are distinct from spatial visualization tasks because spatial visualization tasks involve manipulating the imagined parts of an object separately rather than as a unit; though both types of task require rapid mental visualization, spatial orientation tasks require less complicated reasoning about the structure of imagined objects than spatial visualization tasks (Michael et al., 1957).

In a study comparing internal, or abstract, and external, or navigational, spatial ability, Höök, Sjölinder, and Dahlbäck (1996) found that spatial orientation ability, which they classified as an internal ability, predicted efficiency of performance in a hypertext help system. They did not measure spatial visualization (though they referred to spatial orientation as spatial visualization), so it is unclear whether spatial visualization would have been a better or worse predictor of performance.

In a task which involved following an abstract map which was arranged in either a hierarchical or rotational fashion, Pak, Rogers, and Fisk (2006) found an effect for spatial orientation ability in the rotation condition. They found no effect of spatial visualization ability, but they also did not ask participants to make any decisions or to maintain their position in abstract space.

These few results do suggest, however, that spatial orientation ability can predict performance on speeded search tasks. They do not suggest that it has any effect on reasoning ability or navigational choices, however.

Working Memory Capacity

Better working memory capacity is associated with better control of attention, better ability to attend to a goal, and better ability to remember pertinent facts. Working memory in general refers to items in a person's memory which are currently being acted upon; generally speaking, better capacity means that more items can be actively held in memory. The usual way to measure working memory is to use two tasks, or one task that requires concurrent processing of new information and storage of recently used information.

Different models of working memory predict slight differences in the mechanism for individual variation in working memory, however. Baddeley's current model includes component systems with individual, separate capacities, and a central executive which coordinates those systems, sometimes supplementing their capacity as needed (Baddeley, 2001). Individual variation, then, is due to differences in the capacity of each system, or to differences in the efficiency of the central executive to coordinate them. Spatial ability in this model is the interaction of the visuospatial sketchpad, which stores mental images as they are being acted upon, and the central executive. Miyake and colleagues found that differences in spatial abilities (especially spatial visualization) appeared to depend on differences in central executive function, however, not in differences in the visuospatial sketchpad (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001).

In order to demonstrate that a task has a working memory load, two different types of investigation are possible. One is to show that performance on the task is better in people with higher working memory capacity, and the other is to show that

performance of a concurrent task that taxes a particular kind of working memory capacity decreases performance on the target task.

Working memory has not been consistently shown to affect hypertext performance. Chavez (2002) found that better working memory capacity improved recall, but not navigation, in a web search task. DeStefano and LeFevre (2007) suggested that decision-making load related to the presence of embedded links in hypertext increased cognitive load and caused hypertext to require more working memory resources than linear text with no links. Wenger and Payne (1996) used concurrent digit and spatial loads during hypertext reading, but found inconclusive results about whether that affected performance.

If spatial abilities are the intersection of spatial representation and working memory, I would expect that working memory, being more related to reasoning, would account for the relationship between spatial visualization ability and performance on web search tasks.

Proposed Model

The next step, once one can determine that cognitive abilities do have an effect on performance with information search systems, is to create an explanatory model of how cognitive abilities affect each other and how they affect performance.

Structure

I expect that the relationships between cognitive abilities and measures of usability can be modeled in a linear fashion, and that they will interact with the characteristics of the system. Furthermore, I believe that the cognitive abilities described above are not independent predictors, and thus that they will be correlated

with each other. I also think that the components of usability will be related to each other.

The structure described above leads to the model shown in Figure 1. Spatial Orientation (SO) predicts efficiency because it involves efficient use of mental resources, while Spatial Visualization (SV) predicts both efficiency and effectiveness, because it involves both efficient use of mental resources and reasoning, and Working Memory (WM) predicts only effectiveness because it relates only to goal-maintenance and reasoning. Spatial orientation should correlated with spatial visualization because both are spatial tasks which require quick and accurate mental visualization, but not with working memory, while working memory is also correlated with spatial visualization but not with spatial orientation because both working memory and spatial visualization involve reasoning and goal maintenance. Computer experience (exp) also predicts both efficiency and effectiveness.
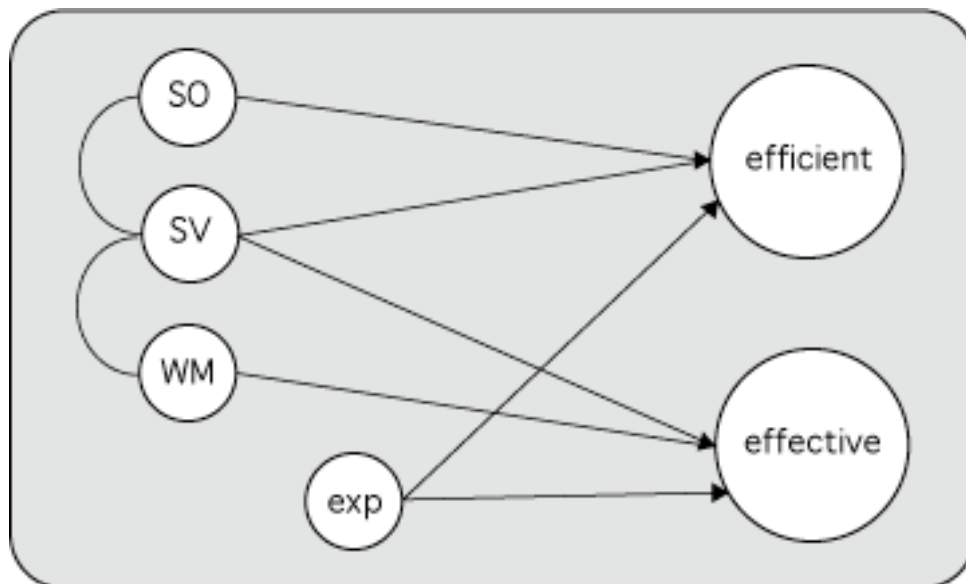
Figure 1. Theoretical relationships among constructs.

Hypotheses

Based on the evidence from the literature, it appeared that spatial visualization would predict performance on both efficiency and effectiveness, but the relationships among the constructs also suggest that there should be some redundancy between spatial visualization and the other two measures of spatial ability. I believed that the following hypotheses would account for the effects of spatial ability on performance.

The first hypothesis concerns the effects of spatial abilities on performance with information search. Because both have been implicated in performance efficiency in the past and because both require quick and accurate processing of information, I believed that both would predict efficiency, and that better performance on spatial tasks would predict more efficient performance with computer interfaces.

H1: Higher spatial visualization and spatial orientation scores (which will correlate with each other) will both predict better efficiency.

The second hypothesis concerns the effects of working memory capacity on effectiveness. Because working memory capacity is used for maintaining goals and reasoning about problems, I expected that better working memory capacity would lead to better performance, though not necessarily faster performance. Since spatial visualization ability has previously been shown to improve performance, I expected that it might have an effect as a mediator of the relationship between working memory or reasoning and effectiveness in information search.

H2: Higher working memory capacity will predict better effectiveness, and spatial visualization may mediate this relationship.

The third hypothesis concerns the previously confirmed finding that better spatial visualization ability leads to better performance on both efficiency and effectiveness. I believed that those differences would be accounted for by specific spatial abilities, measured by spatial orientation ability, and specific reasoning and working memory abilities, measured by working memory. H3: Spatial visualization effects will be accounted for by differences in spatial orientation and working memory.

Furthermore, I expected that experience would also predict performance, but that the type of experience would affect performance differently. Because computing experience has been shown to predict better performance overall, I expected that it would predict better efficiency, rather than effectiveness, which generally relates more to domain knowledge. Therefore, I expected, in the fourth hypothesis, that computing experience would predict better efficiency.

H4: More computing experience will predict better efficiency.

Because more experience with the specific system should lead to better reasoning about that system, I expected in the fifth hypothesis that more experience with the specific information system (in this case, Wikipedia) would lead to better reasoning using that system. I also expected that more experience with Wikipedia would lead to better representations of how knowledge is related within the Wikipedia framework, leading to better ability to evaluate information.

H5: More Wikipedia experience will predict better effectiveness.

Experiment 1 was designed to investigate these relationships by measuring the cognitive and experiential variables which seemed most likely to predict performance on Wikipedia, then giving participants the opportunity to look for information on Wikipedia. By also measuring the information search performance of participants, I could then relate their cognitive abilities to their performance.

Chapter 2: Experiment 1

The first experiment was designed to gather supporting data for the model of performance on a web-searching task. The hypotheses were as described in the preceding chapter.

Method

Participants were asked to find specific pieces of information within the Wikipedia website (http://www.wikipedia.org) and provide both the requested information and the URL of the specific page where they found that information, which requires even those participants who may have some idea of the answer to search for the correct page. They were then asked to perform a set of cognitive tasks to measure mental rotation ability, spatial visualization ability, and spatial working memory. The individual scores on these tasks were then used to test the proposed model of spatial ability and web search success.

Wikipedia was chosen as a site to search because it is familiar to participants but includes enough diverse kinds of information that participants cannot be expected to know the answers to all of the questions asked. It also has defined pages for each topic, which makes devising questions which require navigation as well as search to answer possible. Wikipedia pages also tend to have large numbers of embedded links, and as McDonald and Stevenson (1998A) found, the number of embedded links on a page decreases navigation performance, possibly by involving additional working memory load or general cognitive load.

Participants

All 108 participants were students at the University of Maryland, who received course credit for their participation. Four participants overrode the time limit on the cognitive tasks, and were excluded from all analyses. Two participants failed to get any answers correct, which rendered several measures as undefined, and thus those participants were also excluded from further analysis. Analyses are based on 102 participants, of whom 58 were female and 44 were male. The median age was 19. Because of data collection errors with the hardware and software used, some participants' data were missing a count of the number of clicks performed. Several other participants did not complete the Block Span task, and thus do not have a Block Span score. These participants are included in any analyses for which data existed for them, and the resulting counts for each analysis are given in Table 1. There is no reason to believe that these data are not missing completely at random, because the parts that are missing are clearly due to computer failure and not to participant behavior.

Measure	n
Spatial visualization, Spatial orientation	102
Number Correct	102
Time Spent	102
Working memory	97
Clicks	84
All measures	84

Table 1. Count of Participants available for each measure.

Materials

The materials for the first experiment consisted of a set of information-seeking questions, a set of cognitive ability measures, and several single-question experience assessments.

Wikipedia Questions. The information seeking questions were the same for every participant, though not all participants finished all questions. The questions were developed by the primary investigator and were tested by the research assistants, who advised on changes to the wording and verified that the questions were answerable as designed. Questions were designed to require more than one click after a successful Wikipedia search; further, questions were designed such that the title of the page on which the information was available was not in the text of the question so that browsing as well as searching was required (actual questions are listed in Appendix B). Participants were instructed not to skip questions, and the questions were presented in the same order for all participants. The Wikipedia search was performed in a separate browser window from the question, and participants were instructed on the quickest way to cut and paste URLs in the Macintosh operating system. Basic item analysis was performed on the questions and will be reported later, though a full Item Response Theory (IRT) analysis was not possible due to limited sample size. The complete list of questions can be found in Appendix B.

Cube Comparison (S-2), Form A. The Cube Comparison (S-2) test is a standard test of mental rotation ability which is included in the Kit of Factor Referenced Cognitive Tests (Ekstrom, French, and Harmon, 1976). In each item, participants were shown a drawing of a pair of blocks with letters written on the three

visible faces of each cube. They were then asked whether the cube on the right was a possible rotation of the cube on the left. Form A of this test is 21 items long and, in this version, was presented on a website on a single page in the same layout that was used on the printed version. Participants were given three minutes to complete as many of the 21 items as possible before moving on to the next task.

Paper Folding (VZ-2), Form A. The paper-folding (VZ-2) test is a standard test of spatial visualization ability which is included in the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, & Harmon, 1976). In each item, participants are shown a sequence of drawings depicting a square piece of paper being folded into a multilayered geometric shape. A hole is then depicted as being punched through the shape, and the participants are asked to determine where the resulting holes would be if the shape were unfolded back to a square. There are ten items, and participants were given three minutes to answer as many of the ten items as they could.

Block Span. The Block Span test is a new test of working memory ability which is based on the Corsi Block Tapping task (Harbison, Dougherty, & Bunting, 2008). In this task, participants are shown a 4x4 grid of gray squares. Sequences of squares of increasing length are shown, and participants must reproduce those sequences by clicking on the correct squares in the correct order. The test does not have a time limit.

Experience Questions. Computer, Internet, and Wikipedia experience were assessed using the unlabeled, single-question Likert-type scales pictured in Figure 2.

31

Please rate your experience with the following:										
	no experience									very experienced
Overall use of computers	○	○	○	○	○	○	○	○	○	○
Use of the World Wide Web	○	○	○	○	○	○	○	○	○	○
Use of Wikipedia	○	○	○	○	○	○	○	○	○	○

Figure 2. Screenshot of short self-report computer experience survey.

Procedure

Participants were seated at a Macintosh computer workstation with two monitors (both 15" diagonal LCD screens with a resolution of 1024 x 768 pixels) and were asked to follow the instructions presented in a Firefox (version 3.0) web browser on the screen. Clicks were recorded using TestGen4Web (version 1.0.0), a Firefox extension which records browser activity and pages visited. All procedures were briefly explained before informed consent was obtained (IRB approval letter and approved consent form are Appendix A). Participants first performed the information search task, then the S-2, then the VZ-2, and finally the Block Span task. If they did not finish the information search task in 40 minutes or less, the experimenter stopped them and asked them to continue to the next portion of the experiment. The entire experiment took approximately one hour. A short debriefing about the purpose of the experiment was provided after the participants had finished.

Results

All usability results were derived from the correctness, time to complete, and number of clicks required for each of the ten questions. Answers were coded as correct or incorrect based on answers obtained from Wikipedia before the experiment began; questions which were not attempted or not finished were left blank. Strange-

32

sounding answers were checked to determine whether the page had been changed during the experiment, but no changes appeared to affect the answers. Clicks were counted as the number of clicks and page loads recorded by TestGen4Web for a given question. Time was recorded as the time from when the page holding the question first loaded to the time the participant clicked the "Submit Answer" button.

Because the experimenters did not always give the same amount of time to all participants (due to tardiness or other differences), the cumulative time was computed, and the data were truncated such that steps taken and questions answered after more than 30 minutes had elapsed were eliminated from the analysis.

A table containing descriptive statistics for all measures used is given in Appendix C.

Usability Metrics

The two usability constructs assessed in this experiment were Effectiveness, which refers to the number of tasks completed or lack of errors, and Efficiency, which refers to the speed and number of steps required to perform tasks. Two measures of effectiveness and four measures of efficiency were computed.

The two measures of effectiveness were number of correct answers, which was the raw count of correct answers, and percent of answers given which were correct, which was the number of correct answers divided by the number of questions attempted. The Pearson correlation between number of correct answers and percent correct was 0.479 ($p<.01$). Percentage of attempted questions which were answered correctly is a purer metric for effectiveness because the number of correct answers depends on the number of questions completed, which a measure of efficiency.

Note that, though the reliability of the effectiveness scale across these ten tasks appears fairly high, its speededness may have led to overestimation of Cronbach's alpha (α=.629 for the 40 complete cases, or α=.703 for all cases with omitted responses scored as incorrect). Probably because the scale is not unidimensional (a Principal Components analysis extracted five factors with eigenvalues greater than one), the single-parameter item-response marginal reliability for the scale when omitted items are scored as not presented is only .214.

The four measures of efficiency which were computed for this experiment were as follows: Total Completed was the total number of questions attempted; Time (s) per Click was the average across questions of the number of seconds required for a particular question divided by the number of clicks performed for that question; Clicks per Correct Answer was the average number of clicks performed only on questions for which the participant answered correctly; and Time (s) per Correct Answer was the average time in seconds a participant spent on questions which they answered correctly. The correlation matrix among these four metrics is shown in Table 2. Time per click appears to be the purest measure of speed because it does not depend on the correctness of answers. It may be problematic, for instance, that the number of pages required differed between the easy and hard questions; if a participant only got the easy questions correct, their efficiency would seem higher than the efficiency of someone who also correctly answered more difficult (and thus more time-consuming) questions. Though it might be possible to construct a composite measure of efficiency using these disparate measures, concerns about interpretability and extensibility of results and the previously mentioned concerns

about bias rendered that approach impractical. Note also that choosing this measure due to its lack of correlation with measures of effectiveness precludes investigating how efficiency and effectiveness are correlated.

Efficiency	Total Completed	Time (s) per click	Clicks per correct answer
Time (s) per click	-0.407*		
Clicks per correct answer	-0.492*	-0.205	
Time (s) per correct answer	-0.657*	0.158	0.804*

* Correlation significantly different from 0 with p<.01
Table 2. Correlations among measures of efficiency.

Some measures of efficiency and effectiveness were correlated (as shown in Table 3), because they are derived from the same performance and may include some of the same terms. The metrics which were chosen as the main indicator for each construct (percent correct for effectiveness and time per click for efficiency), however, were those which were least related to the other construct.

	Total Completed	Time (s) per click	Clicks per correct answer	Time (s) per correct answer
Total Correct	0.796*	-0.367*	-0.377*	-0.473*
Percent Correct	-0.086	-0.163	0.226	0.172

* Correlation significantly different from 0 with p<.01
Table 3. Correlations between measures of effectiveness and measures of efficiency.

Non-cognitive individual differences

Computer, WWW, and Wikipedia experience were assessed using single-question self-report instruments, so internal consistency is not available. The correlations among them are shown in Table 4.

Experience	Computer	WWW
WWW	0.911*	
Wikipedia	0.555*	0.555*

* Correlation significantly different from 0 with p<.01

Table 4. Correlations among computer experience questions.

Wikipedia experience is not as strongly related to general computer experience as WWW experience is, and general computer experience and WWW experience appear to be nearly identical in this sample.

Cognitive predictors

Because the S-2 was a speeded test, Cronbach's alpha is an overestimate of its reliability, but all reliabilities were within the expected range. The VZ-2 is generally used as a composite of both forms, and only one form was given here, which leads to a lower Cronbach's alpha than is usually reported for that test. A correlation between S-2 and VZ-2 scores was expected, as both are measures of spatial ability and have sometimes been used interchangeably, but the correlation could not be distinguished from zero (even correcting the correlation for the measures' unreliability only produced a correlation of r=0.23). The VZ-2 and Block Span were expected to be correlated, and did show a small but detectable correlation (see Table 5).

Cognitive Predictors	Spatial visualization Score	Spatial orientation Score	Working memory
Spatial visualization Score	0.676		
Spatial orientation Score	0.164	0.778	
Working memory	0.266*	0.173	0.710

Diagonal values are Cronbach's alpha.
*Correlation significantly different from 0 with p<.01

Table 5. Correlations and reliabilities among cognitive predictors.

Mental rotation and speed

The significant correlation between mental rotation (S-2 score) and efficiency (time per click) may be accounted for by the number of problems attempted, rather than the correctness of those problems; the correlation between the S-2 score and number of S-2 items completed is 0.829 (p<.01), but while the correlation between the S-2 score and time spent per page is -0.3 (p<.01), the correlation between the number of S-2 items completed and the time spent per page is -0.369 (p<.01). This pattern suggests that it is possible there may be a general speed factor at work here, rather than a specific spatial component, or that people who are better at mental rotation are generally also faster at the task. The correlation between score and number of items attempted is due to the speededness of the S-2, as shown in Table 6.

	S-2 Score	S-2 Completed
S-2 Completed	.829*	
Time per Click (s)	-.300*	-.369*

* Correlation significantly different from 0 with p<.01

Table 6. Correlations between mental rotation and speed.

Spatial visualization and working memory

The expected pattern was that working memory (Block Span) would correlate with percent correct, and that spatial visualization (VZ-2) would correlate with efficiency. The obtained result was a correlation between spatial visualization and number correct, with no correlations of working memory to anything except spatial visualization, and no correlation of spatial visualization to efficiency (see Table 7). Though the zero-order correlation between spatial visualization and efficiency was 0.063 (ns), later results suggest it may be predictive in combination with other variables.

	Spatial visualization Score	Working memory	Total Completed
Working memory	0.266*		
Total Completed	-0.025	0.142	
Percent Correct	0.310*	0.104	-0.086

* Correlation significantly different from 0 with p<.01

Table 7. Correlations among spatial visualization, working memory, and measures of effectiveness.

Gender

Gender appeared to have an effect on spatial visualization and on percent correct. Men were more accurate at spatial visualization (t(101)=2.046, p<.05) and had a higher percentage of questions correct on the Wikipedia task (t(101)=2.02, p<.05). Gender did not significantly improve prediction in any of the regression models which follow, however.

Model testing

The initial model, in which spatial visualization and mental rotation predicted efficiency and working memory and spatial visualization predicted effectiveness was not supported.

Spatial visualization and specific (Wikipedia) experience predicted effectiveness measured by percent correct, $F(2,99)=7.236$, $p<.01$, *adj. $R^2=.110$*. As shown in Table 8 spatial visualization ($\beta=.252$) was a better predictor of effectiveness than Wikipedia experience ($\beta=.197$).

Parameter	Unstandardized Coefficients		Standardized Coefficients			
	B	Std. Error	beta	t	df	p
Intercept	.478	.083		5.788	99	<.001
Wiki experience	.021	.010	.197	2.027	99	.045
Spatial visualization	.022	.009	.252	2.589	99	.011

Table 8. Regression coefficients for predicting effectiveness from spatial visualization and Wikipedia experience.

The picture for efficiency was more complex, however. General WWW experience predicted (β=-.489) the amount of time that people spent per page; the more experience a person had, the less time they spent. As seen in Table 9, both mental rotation ability (β=-.283) and spatial visualization ability (β=.223) also predicted average time, but they did so in opposite directions; faster mental rotation predicted less time per page, while better spatial visualization predicted more time per page. The overall model was significantly predictive, $F(3,79)$=12.406, $p<.01$, *adj. R^2=.294*.

Parameter	Unstandardized Coefficients		Standardized Coefficients			
	B	Std. Error	beta	t	df	p
Intercept	43.582	3.946		11.044	79	<.001
Spatial orientation	-.475	.160	-.283	-2.974	79	.002
Spatial visualization	.725	.315	.223	2.300	79	.024
Internet experience	-2.179	.422	-.489	-5.170	79	<.001

Table 9. Regression coefficients for predicting efficiency with spatial visualization, mental rotation, and Internet experience.

A simpler model using only mental rotation and WWW experience predicted efficiency measured by time per click, $F(2,80)$=15.152, $p<.01$, *adj. R^2=.257*. As shown in Table 10, in this second model, as in the first, WWW experience (β=-.450) was a better predictor of efficiency than mental rotation ability (β=-.236).

Parameter	Unstandardized Coefficients		Standardized Coefficients			
	B	Std. Error	beta	t	df	p
Intercept	45.522	3.957		11.504	80	<.001
Spatial orientation	-.396	.160	-.236	-2.472	80	.016
Internet experience	-2.002	.425	-.450	-4.707	80	<.001

Table 10. Regression coefficients for predicting efficiency with mental rotation and Internet experience.

One outlier was removed from the regression analyses because the participant showed an extremely high (Z=7.87) time per click average and a low mental rotation score, leading to concerns that that participant was driving the mental rotation effect.

The obtained model, therefore, as presented in Figure 3, showed spatial

visualization (vz-2), spatial orientation (s-2), and experience (exp) predicting

efficiency (s/page), while spatial visualization and experience predicted effectiveness

(% correct). Working memory (bs) was correlated with spatial visualization, but did

not predict either outcome.
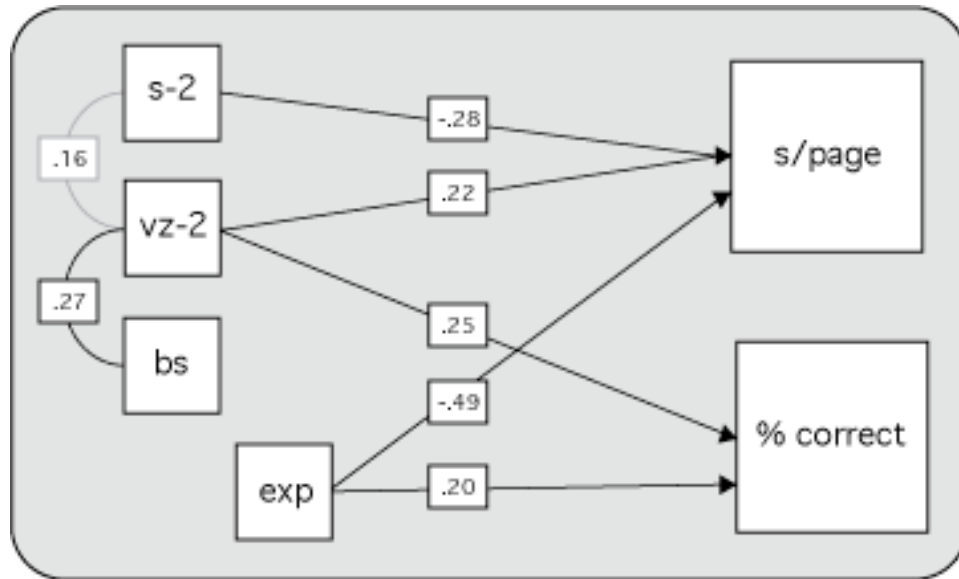
Figure 3. Obtained relationships in Experiment 1.

Discussion

The expected model was not supported, and two surprising findings will need

to be investigated further. The first was that I did not find the expected correlation

between spatial visualization and spatial orientation. This result may just be due to

random error, or could be due to some sort of task constraint; both tasks showed

acceptable reliability and no ceiling or floor effects. The other surprise was that

higher spatial visualization ability predicted slower performance on the search task

when spatial orientation and computer experience were taken into account, though the

zero order correlation between speed and visualization ability was small and there did not seem to be a tradeoff between speed and accuracy.

H1: Higher spatial visualization and spatial orientation scores (which will correlate with each other) will both predict better efficiency.

Hypothesis H1 was not completely supported. Higher spatial orientation scores did predict better efficiency, but they did not correlate with spatial visualization. Furthermore, higher spatial visualization scores predicted slower performance when spatial orientation performance was taken into account.

H2: Higher working memory capacity will predict better effectiveness, and spatial visualization may mediate this relationship.

Hypothesis H2 was not supported; higher spatial visualization scores did predict better effectiveness, but higher working memory scores showed no effect, and spatial visualization did not function as a mediator. One possible explanation for this finding is that the search task was too easy for differences in working memory capacity to be predictive, but the overall accuracy was around 75% and the average participant did not finish in the time allotted, so that explanation seems unlikely.

H3: Spatial visualization effects will be accounted for by differences in spatial orientation and working memory.

Hypothesis H3 was not supported at all. Spatial visualization appeared to provide unique explanatory power which was not only not accounted for by other measures of spatial ability and working memory, but also sometimes predicted in the opposite direction.

H4: More computing experience will predict better efficiency.

41

Hypothesis H4 was supported in this experiment, despite the limitations of the instrument. Higher self-rated computer experience predicted faster performance.

H5: More Wikipedia experience will predict better effectiveness.

Hypothesis H5 was also supported in this experiment, despite the limitations on its measurement. More Wikipedia experience predicted better performance, perhaps because users were more able to evaluate the links and answers.

The obtained results from this task suggest that spatial ability is important for predicting performance with computer interfaces, though different abilities are predictive of different outcome measures. Further, the experiment showed that computer experience of different kinds predicts different kinds of performance. The results, however, hold only for use of the World Wide Web on a stationary, large-screen computing device with multiple windows and with a keyboard and mouse. Experiment 2 was an attempt to replicate and extend these results by moving to a mobile platform, where working memory and experience should be more important.

Chapter 3: Experiment 2

Experiment 1 was concerned with the determinants of web search success on desktop computer systems. In Experiment 2, I examined whether the same determinants of success existed when people were using mobile devices, such as the iPod or iPad (shown in Figure 4).
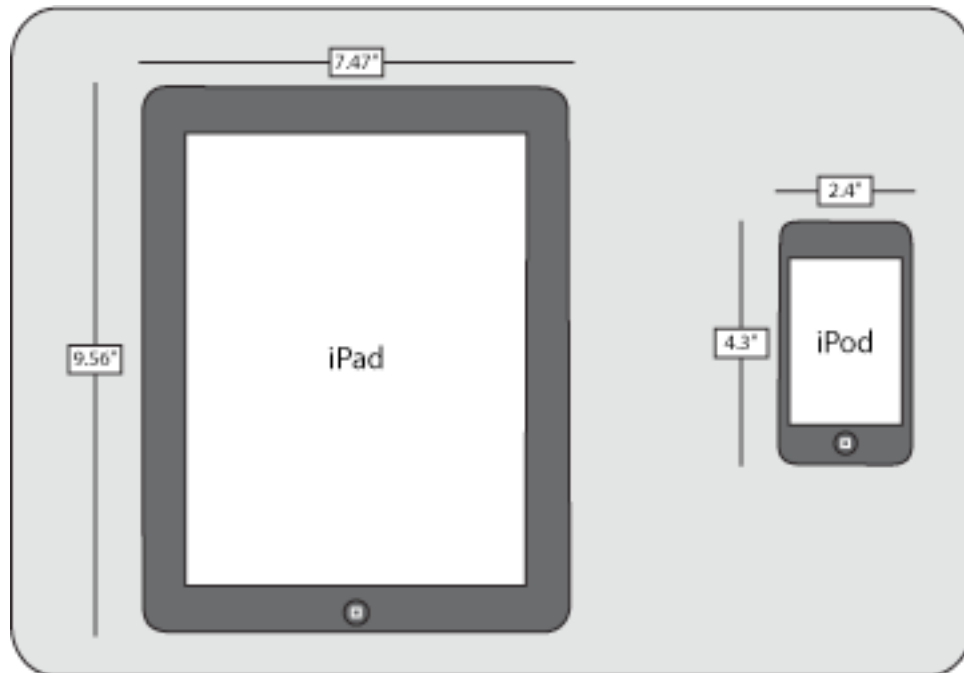
Figure 4. Size comparison of the iPad and iPod Touch.

Mobile devices have become ubiquitous over the past several years and are frequently used to access the Internet. Information seeking is a common task performed using the mobile Internet, and users in one survey report using Wikipedia almost exclusively to find information on mobile devices, largely due to the amount of information found there and the fact that users think the information on Wikipedia is generally reliable (Cui & Roto, 2008).

Though it is frequently asserted that mobile devices produce more of a load on working memory than desktop computers, this assertion appears to be based on logic and self-report, rather than a systematic investigation of the interaction of working memory and mobile devices. For instance, Ally (2008), in a survey article about using mobile devices for learning in a way which matches principles of cognitive psychology, asserts that mobile device screens should have no more than five to nine pieces of information on them, taking this admonition from Miller's (1956) estimate of span of apprehension. Reducing the amount of information on a screen, however, is only useful if one believes that screens should be processed in their entirety rather than read sequentially.

In Experiment 1, working memory appeared to have no effect on efficiency or effectiveness of information search. One goal of the second experiment is to determine if differences in working memory have an effect in mobile computing situations.

Another theory about mobile computing is that the same constraints which held in early computing, when screens were smaller, will hold now that screens have become small again. Another similarity between early computing devices and devices such as the original iPhone is that they only run a single application at a time in a single window. The similarities between modern small-screen devices and early computers would suggest that spatial visualization-related processes, which were considered important in the 1980s, will be important again in mobile computing contexts.

Because the first experiment did not vary the difficulty of the interfaces presented, only the difficulty of the measurement tasks, it is impossible to find real variance in usability. To rectify this situation, the second experiment should contrast at least two different interfaces for finding information. In order to contrast screen size without also contrasting method of interaction, I compared Wikipedia search on the iPod Touch and the iPad. Because the screen on the iPad is bigger, less scrolling and less cognitive overhead are required to maintain location within the page. Therefore, I predicted that information search should be easier and quicker on the iPad than on the iPod Touch.

Besides this prediction that information search should be easier and quicker on the iPad than the iPod Touch, I also predicted that the same model that held for large-screen devices should hold for small-screen devices, but that working memory might play more of a role because of the larger amount of cognitive overhead required to maintain one's goal in the face of more competing decisions.

Hypotheses

The second experiment included some of the same hypotheses as the first experiment, and some which were intended to replicate the findings of the first experiment. It also included new hypotheses related specifically to mobile devices.

The first hypothesis in the second experiment concerned device screen size; the bigger the screen, the easier information search should be, for all the reasons described above.

> H6: Participants using the iPad should have better effectiveness and efficiency than those using the iPod Touch.

In Experiment 1, only spatial visualization ability predicted effectiveness, but better working memory capacity should make holding the goal of the information search task in memory easier, increasing effectiveness. Though I did not find this result in the first experiment, in the second experiment using mobile devices, I expected it to emerge.

H7: Higher spatial visualization ability and working memory scores should predict better effectiveness.

In the first experiment, spatial orientation predicted efficiency. I expect the same result, and tried to confirm the surprising finding that higher spatial visualization ability predicted slower performance on a search task once spatial orientation was taken into account.

H8: Higher spatial orientation ability should predict higher efficiency, but higher spatial visualization ability should predict lower efficiency once spatial orientation ability is taken into account.

I expect that the size of the screen should affect how important spatial visualization is to performance; the smaller the screen, the more important spatial visualization will be to performance. I have no reason to believe that this will differ between efficiency and effectiveness.

H9: The device used should moderate the relationships between spatial visualization and efficiency and effectiveness.

Because smaller devices are supposed to induce more cognitive load than larger devices, I expected that the smaller device would cause the participants' working memory to better determine effectiveness than it did on the larger device.

H10: The device used should moderate the relationship between working memory and effectiveness.

The final two hypotheses were similar to H4 and H5 in Experiment 1, though H11 was not identical to H4 because the equipment used was slightly different in Experiment 2. I expected that the same relationships should hold between information system experience (Wikipedia experience) and effectiveness, as well as between device experience (computer experience in the first experiment, iPod/iPad experience in this experiment) and efficiency.

H11: Better mobile device experience should predict better efficiency.

H12: Better Wikipedia experience should predict better effectiveness.

Method

The second experiment replicated the first experiment with a subset of the original list of questions and with participants performing the information-seeking task using either an iPod Touch or an iPad. The list of questions was shorter because information seeking generally takes longer on a mobile device and because many participants did not finish in Experiment 1.

Participants

The participants in the second experiment were also students at the University of Maryland and received partial course credit for their participation. There were 106 participants, but several participants were excluded either because the server malfunctioned during their session (n=3) or because they failed to follow directions and used Google search instead of Wikipedia search to find the appropriate pages

(n=7), for a final analyzed sample size of 96 (67 female, 29 male; median age 19; mean age 19.4).

Materials

Like the materials for the first experiment, the materials for the second experiment consisted of a set of information-seeking questions, a set of cognitive ability measures, and several single-question experience assessments. The only difference between the two sets of materials was that a subset of the questions used in Experiment 1 was used in Experiment 2, and the use of mobile devices was added to the experience questions. In order to reduce the chances of participants failing to finish the Wikipedia question section in the allotted time, only five of the ten questions were selected for Experiment 2, and the questions chosen were those which more participants answered correctly in Experiment 1.

The complete list of Wikipedia questions used in Experiment 2 can be found in Appendix E.

The new computer experience section is shown in Figure 5. The only change was the addition of questions on the use of the iPod Touch/iPhone and the iPad.

Figure 5. Computer experience question section in Experiment 2.

Procedure

Participants were seated at a Macintosh computer workstation and were asked to follow the instructions presented in a Safari (version 5.0) web browser on the screen. Participants were assigned to use the iPad or iPod Touch based on a previously randomly generated list which was kept in the room. All procedures were briefly explained before informed consent was obtained (IRB approval letter and approved consent form are Appendix D). The mobile devices were set to use a local area proxy (included with Mac OS X Server 10.6), which recorded all HTTP requests sent from the device. Essentially, every time a device requested a page from the WWW, that request was recorded. Participants first performed the information search task using the mobile device, then the S-2 on the desktop computer, then the VZ-2 on the desktop computer, and finally the Block Span task on the desktop computer. If they did not finish the information search task in 40 minutes or less, the experimenter stopped them and asked them to continue to the next portion of the experiment. The entire experiment took approximately one hour. A short debriefing about the purpose of the experiment was provided after the participants had finished.

<div align="center">

Results

</div>

Three sources of information were integrated to produce the final data set. Question answers and timings were collected in a FileMaker 6.5 database, and the relevant records were directly exported for data analysis. Block Span results were collected in a separate file, which was matched to the ID numbers given in the database. Finally, the HTTP proxy log was parsed with a Perl script which returned a data file with only Wikipedia pages accessed by the devices used for the experiment

included, and those records were matched by date and time and device to the database records.

Of the participants, 54 used the iPad to complete the questions and 42 used the iPod Touch. The groups were not even because, in a few sessions, the experimenters could not get the iPod Touch to connect to any wireless network, so they substituted the iPad. A second iPod Touch was added as a backup later in the semester in order to reduce the number of technical issues.

To match the first experiment, where efficiency was measured in seconds per page and effectiveness was measured in percent correct, the same variables were derived again. Because the clicks needed to find an answer were not collected the same way, the number of seconds per page may not be comparable to the first experiment, but it is a measure of the same construct – time spent per page visited.

Descriptive Statistics

A complete table of descriptive statistics is available in Appendix F. The number of seconds per page visited is positively skewed, but all other variables appear to be of acceptable distribution.

The computer and device experience questions mostly clustered together, though a Principal Components Analysis with a Varimax rotation suggests that perhaps there are two separate factors, one for computer experience (general computer experience, internet experience, and Wikipedia experience) and one for mobile device experience (iPhone/iPod experience and iPad experience).

Differences between devices

There were no significant differences between the iPad users and the iPod Touch users in any of the cognitive or experiential variables, and the values for the cognitive variables are shown in Table 11.

	iPad	iPod Touch	t	df	p
Spatial Visualization (VZ-2)	5.49	5.90	-.813	90	.419
Spatial Orientation (S-2)	11.43	11.43	-.003	94	.997
Working Memory (Block Span)	52.46	52.50	-.014	90	.989

Table 11. Means of iPod Touch and iPad groups on cognitive individual differences variables.

The iPad and iPod Touch outcomes differed in aggregate measures of efficiency, but not in aggregate measures of effectiveness. The only difference in effectiveness was that the iPad users were less likely to get Question 3 correct. iPad users were significantly less likely to answer the question correctly, $X^2(1) = 5.17$, p=.02. There is no theoretical reason to believe that this is a meaningful result, and no other significant differences existed between devices on any measure of correctness.

The iPad users were more likely to take more steps (both searching and browsing) to find any answer, t(92)=3.01, p=.003. The variance in time spent per page among iPod Touch users was higher, F(1, 93)=7.52, p=.007, and iPad users were more likely to spend less time on each page, t(60.26)=5.19, p<.001.

About half of the iPod Touch users (22 of the 42) were redirected to the Wikipedia Mobile site, which shows a slightly simplified version of the same Wikipedia content. There were no quantitative differences in any of the measures between these two groups, however, except in the number of Search steps, so they will be treated as one group for the rest of the analyses. Participants using the Mobile version of the site took slightly more Search steps on average than participants using

51

the full version of the site, t(40)=2.337, p=.03. Instead of magnifying differences between the groups, this difference actually makes Wikipedia Mobile site users more similar to the iPad participants on that particular value.

Gender

There were a few gender differences in this sample. Frequently, the variance in measures differed between men and women, but this was likely due to the difference in sample size (there were about twice as many women as men in the sample).

Though differences in spatial ability are frequently found between men and women, in this sample, spatial orientation ability did not differ at all, t(94)=-0.74, p=.46. Spatial visualization ability also did not significantly differ between men and women, t(90)=-1.82, p=.07. The average working memory score was higher for men than for women, however, t(90)=-3.23, p=.002.

Men and women also did not significantly differ on any measure of computer experience (general experience, internet experience, Wikipedia experience, or iPhone/iPod experience), though men may have been more likely to have more experience with the iPad, t(94)=-1.93, p=.056.

Predicting Effectiveness

As seen above, which device a particular participant used did not affect their total percentage of questions answered correctly and further tests revealed that it did not interact with any measure of ability or experience to affect effectiveness. Both spatial visualization (r=.323, p=.002) and working memory (r=.272, p=.009) predicted percent correct individually, but working memory did not contribute

prediction over spatial visualization, as shown in Table 12. The overall prediction achieved in this regression is lower than that found in Experiment 1 (adj. R2=.094). Also, unlike in Experiment 1, none of the experience questions predicted the number of Wikipedia questions a participant would get correct.

Model		Unstandardized Coefficients		Standardized Coefficient			
		B	Std. Error	Beta	t	df	p
1	(Constant)	.604	.056		10.878	89	<.001
	Spatial visualization	.029	.009	.323	3.224	89	.002
2	(Constant)	.487	.091		5.349	88	<.001
	Spatial visualization	.023	.010	.257	2.390	88	.019
	Working memory	.003	.002	.175	1.623	88	.108

Table 12. Regression table for Percent Correct on spatial visualization and working memory.

Surprisingly, however, spatial orientation ability did predict the number of questions a participant would get correct. Given a particular spatial visualization score, the higher the person's spatial orientation score, the fewer questions they answered correctly. This regression model (*adj. R^2=.159*) is show in Table 13 (slight coefficient and degrees of freedom differences from the above table are due to a difference in the number of people finishing the Block Span task). This effect of spatial orientation ability may be due to participants with higher spatial orientation ability being more likely to go too fast.

Model		Unstandardized Coefficients		Standardized Coefficient			
		B	Std. Error	Beta	t	df	p
1	(Constant)	.606	.056		10.904	90	<.001
	Spatial visualization	.029	.009	.323	3.238	90	.002
2	(Constant)	.757	.076		9.972	89	<.001
	Spatial visualization	.035	.009	.383	3.89	89	<.001
	Spatial orientation	-.016	.006	-.277	-2.807	89	.006

Table 13. Regression table for Percent Correct on spatial visualization and spatial orientation.

Predicting Efficiency

Though which device participants used did not affect their success at answering the questions correctly, it did predict both how many pages they visited

(iPad users visited more pages) and how long they spent there (iPad users spent less time per page). The best ability predictor of how long a user spent on a page was spatial orientation ability; the better a participant did at the spatial orientation task, the less time they spent per page. In Experiment 1, higher spatial visualization ability caused a small amount of extra time to be spent per page, and in Experiment 2, that appeared to still be true in the iPad condition, but in the iPod Touch condition, the higher a participant's spatial visualization score, the less time he or she spent per page. So in addition to being a predictor on its own, screen size also moderated the relationship between spatial visualization and time spent on each page. Spatial visualization, which predicted in opposite directions in the two conditions, was not a significant predictor on its own. The device and ability measures together accounted for a moderate amount of variance (*adj. R^2=.301*), as shown in Table 14.

Parameter	Unstandardized Coefficients			df	p	Partial η^2
	B	Std. Error	t			
Intercept	74.403	7.078	10.512	85	<.001	.565
Device*	-31.795	7.641	-4.161	85	<.001	.169
Spatial orientation	-.976	.387	-2.519	85	.014	.069
Spatial visualization	-1.692	.999	-1.693	85	.094	.033
Spatial visualization x Device*	2.765	1.231	2.246	85	.027	.056

* Dummy coded such that the iPad=1, iPod Touch=0. In the interaction condition, this means that "Spatial visualization" by itself gives the coefficient for the iPod Touch condition, while the Spatial visualization x Device line gives the coefficient in the iPad condition.

Table 14. Regression of time spent per page on ability measures and screen size.

The difference in the spatial visualization correlation with time spent per page between the iPad (.203) and iPod Touch (-.256) conditions was significant, correlation test *z*=2.094, *p*=.02, though neither of the correlations was significantly different from zero.

Discussion

H6: Participants using the iPad should have better effectiveness and efficiency than those using the iPod Touch.

Hypothesis H6 was partially supported; participants using the iPad did show higher efficiency than those using the iPod Touch. However, there was no difference in effectiveness whatsoever. The lack of difference may be due to several factors, but the simplest culprit is that the questions may have been too easy to show real differences in effectiveness. Most participants answered most of the questions correctly (78% average percentage correct). Differences in individual questions did not seem to show a pattern.

H7: Higher spatial visualization ability and working memory scores should predict better effectiveness.

Hypothesis H7 was supported, though once spatial visualization ability was taken into account, working memory ceased to be a significant predictor of effectiveness (though the reverse was not true). This pattern suggests that, while the variance that is shared between working memory and spatial visualization predicts effectiveness, there is some additional variance contained in spatial visualization performance that contributes to effectiveness. This finding may be due to spatial visualization tests including more reasoning component than working memory tests.

H8: Higher spatial orientation ability should predict higher efficiency, but higher spatial visualization ability should predict lower efficiency once spatial orientation ability is taken into account.

The first part of hypothesis H8 was supported for both devices; better spatial orientation ability predicted faster performance on both, as shown in Figure 6. The second part was supported for the iPad, where better spatial visualization ability predicted slower performance, but not for the iPod Touch, where spatial visualization had the opposite effect. This finding suggests that there is some qualitative difference between performance on the iPad and the iPod Touch. Overall, there is no effect for spatial visualization ability on efficiency, but that is due to the effect existing in opposite directions for the two devices.
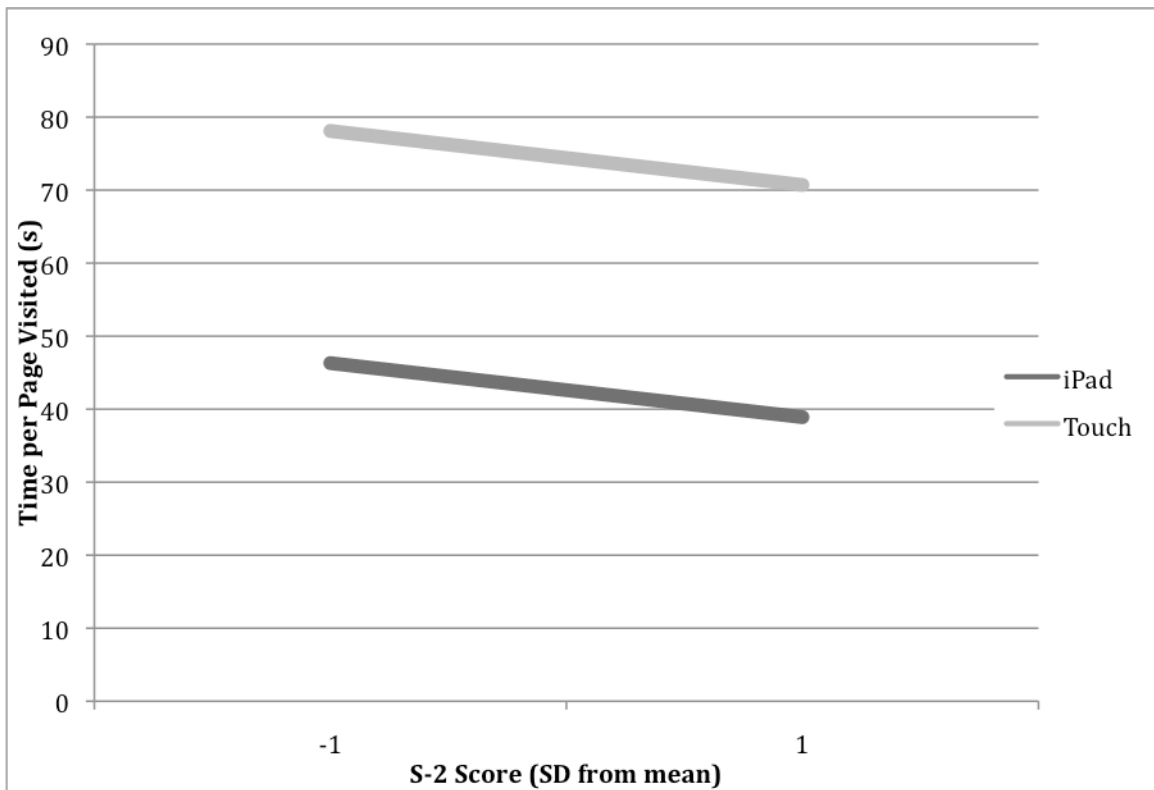
Figure 6. Interaction plot of the relationship between spatial orientation score and time per page.

H9: The device used should moderate the relationships between spatial

visualization and efficiency and effectiveness.

The problems with Hypothesis H8 are due to the success of H9. The

relationship between spatial visualization and efficiency is moderated by screen size,

though the relationship between spatial visualization and effectiveness is not. In fact,

the relationship between spatial visualization and efficiency reverses itself on the two

devices, with higher spatial visualization predicting better efficiency on the iPod

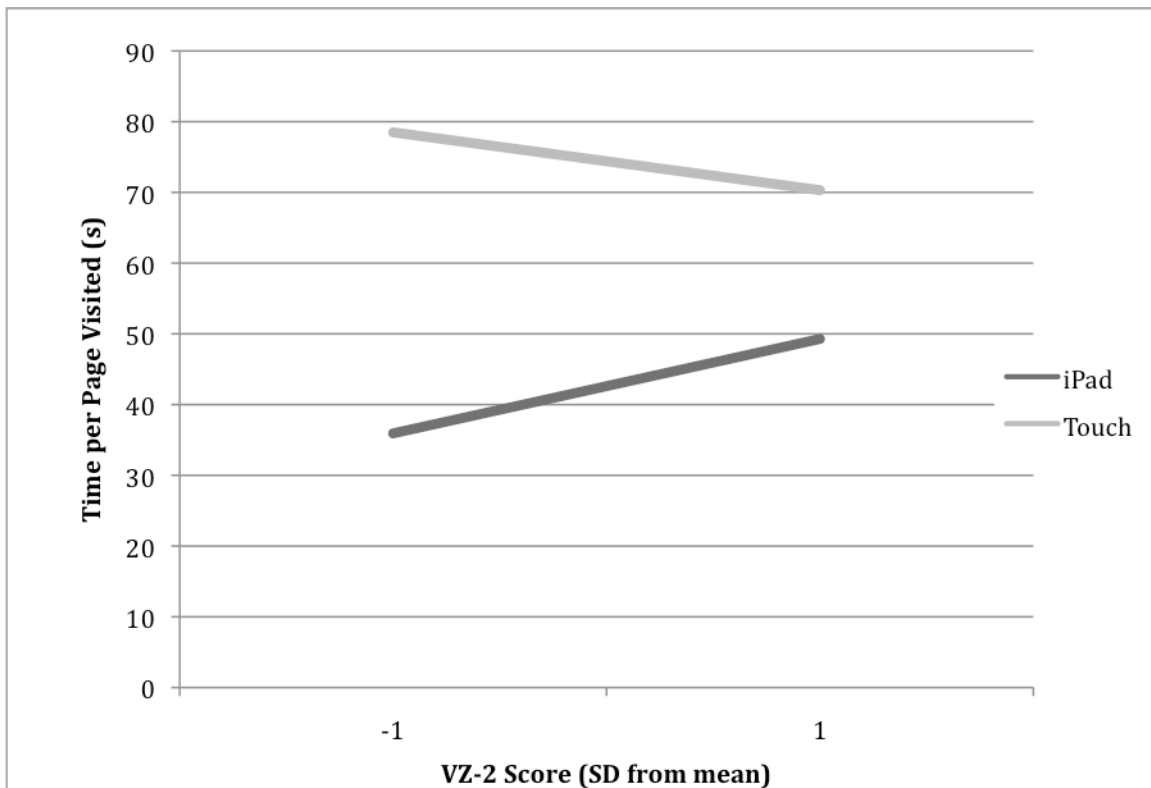Touch and worse efficiency on the iPad, as shown in Figure 7.

Figure 7. Interaction plot of the relationship between spatial visualization score and efficiency across devices.

H10: The device used should moderate the relationship between working memory and effectiveness.

Hypothesis H10 was not supported, though it is possible that future research may find that working memory is more predictive on the iPod Touch than the iPad, since working memory did predict slightly better on the iPod Touch than the iPad. The interaction did not approach significance, however.

H11: Better mobile device experience should predict better efficiency.

H12: Better Wikipedia experience should predict better effectiveness.

Neither Hypothesis H11 nor H12 was supported. None of the experience questions predicted anything about outcomes in Experiment 2, which was unexpected. Perhaps the questions were too easy for prediction of effectiveness in general. It is also possible that proficiency with mobile devices is more general than experience with the iPhone or iPad; the questions about mobile devices may have been too specific.

The users performed differently with the two differently sized devices regardless of their cognitive abilities, however. Though the devices did not differ on effectiveness, they did differ in efficiency; iPod Touch users did not make more mistakes than iPad users, but they spent more time on each page they visited. This pattern suggests that larger screens are easier to use for information search, though people are adaptable and can find information on smaller screens. The cognitive abilities required for performance on the smaller screens did not change from Experiment 1, suggesting that there is some stable amount of prediction that spatial visualization and spatial orientation can provide for information search regardless of

screen size. Another difference that should be taken into account for usability testing of an interface, then, is the size of device that the population will be using to access the interface. We can expect smaller-screen devices to be less efficient, all other things being equal, but based on these results, there is no reason to suspect that users will be less effective with them.

The second experiment showed that the cognitive abilities required for the use of mobile devices differ somewhat from the cognitive abilities required for the use of desktop computer systems, but only in the interaction between screen size and spatial visualization in predicting efficiency. In general, the framework of these particular abilities predicting performance was supported, though the exact details did not match my hypotheses.

Chapter 4: General Discussion and Conclusions

The two experiments just described investigated the effect of cognitive abilities on different kinds of performance. The first experiment looked at the relationship of spatial ability to performance using a desktop computer to search for information, while the second experiment compared information search on two different mobile devices. The dependent variables were components of usability: efficiency and effectiveness. In both cases, spatial visualization predicted effectiveness and spatial orientation predicted efficiency.

The results of the two experiments were complementary, but there were some particulars on which they did not match completely. Reconciling these differences allows a more complete picture to emerge and suggests future directions for research.

General Discussion

The two experiments showed generally consistent results for the relationship between spatial abilities and the outcomes on particular usability measures. The experience variables were not quite as consistent.

Computer experience was a better predictor in the first experiment than the second, but it is possible that the short list of questions was more applicable to desktop computers than mobile devices. It is also possible that the questions in the second experiment were too specific to Apple brand mobile devices, when the determining factor was actually experience with any small-screen mobile device.

Spatial Ability

Across the two experiments, patterns emerged in what abilities were necessary to perform information-search tasks. Higher spatial visualization scores predicted

more correct answers in both experiments. Higher spatial orientation scores predicted less time spent on each page in both experiments, and the effect size was similar between experiments.

Spatial working memory did not predict performance in the first experiment. In the second experiment, working memory was a significant predictor of effectiveness but was overshadowed by spatial visualization. The Block Span task did show the expected pattern of correlations with the other spatial ability measures, however, suggesting that there was not a problem with the measure itself. The idea that working memory is more important for mobile device use than desktop computer use was not strongly supported by these results, but it was certainly not disproved either. Working memory was in fact more predictive with the mobile devices than it was on the desktop computer task, though its effects were not strong in the mobile device experiment. It also did not show an expected difference in relationship to performance due to differences in screen size, however. Perhaps verbal working memory may be more predictive than spatial working memory.

Spatial visualization's relationship to efficiency was complicated, and making sense of it requires some conjecture. Studies from 20 years ago suggested that higher spatial ability predicted faster performance on desktop interfaces. Experiment 1 showed slightly slower performance for people with higher spatial visualization ability once spatial orientation ability was taken into account. The findings from Experiment 2 may shed some light on this disparity, however; the resolution of the iPod Touch is more similar to desktop computers of 20 or more years ago than the resolution of either the iPad or a modern desktop computer. Using the iPod Touch,

better spatial visualization magnified the effect of better spatial orientation on faster navigation, while on the iPad, the same pattern was observed as on the desktop interface. This finding could mean that spatial visualization is helpful on very small interfaces and counterproductive on larger screens. It could also suggest, however, that opposing processes are acting based on the complex nature of spatial visualization. Perhaps more thorough reasoning, which is one component of spatial visualization, produces longer dwell times on particular pages, while better maintenance of spatial representations, which is also a component of spatial visualization, predicts better maintenance of off-screen content in a scrolling environment.

One way of getting closer to the source of the small negative effect of spatial visualization on efficiency would be to investigate whether spatial visualization increased the time spent on specific pages. The effect does not seem to be a speed-accuracy tradeoff because speed and accuracy were not detectably related. Because the effect only exists when spatial orientation is taken into account, perhaps an investigation of the non-spatial components of spatial visualization ability, such as reasoning or central executive involvement, would better separate both the different effects and the different kinds of spatial ability.

Measuring Usability

In these two experiments, it was clear that effectiveness and efficiency, as I measured them, were not strongly related. There were no significant tradeoffs between them, and different abilities predicted different outcomes. This dissociation, however, depends on particular ways of measuring the constructs of efficiency and

effectiveness; while seconds per page and percent of questions answered correctly were unrelated, total number of questions answered and total number of questions answered correctly were strongly related.

The different possible measures of effectiveness and efficiency will yield different outcomes, so a consensus is necessary about what the appropriate measures look like. I argue that the measures I chose here are the most sensitive to differences between people and between interfaces. As an example, however, a comparison of two different interfaces that rely on data networks or devices of different speeds at the time of testing may require measurements less sensitive to time differences.

Possible Future Models

The models that I used here were refined across the two experiments. It is unclear what part experience plays in performance, but the cognitive ability measures were remarkably consistent in their parts across the two experiments (see Figure 8 and Figure 9). In both experiments, spatial orientation ability (s-2) predicted
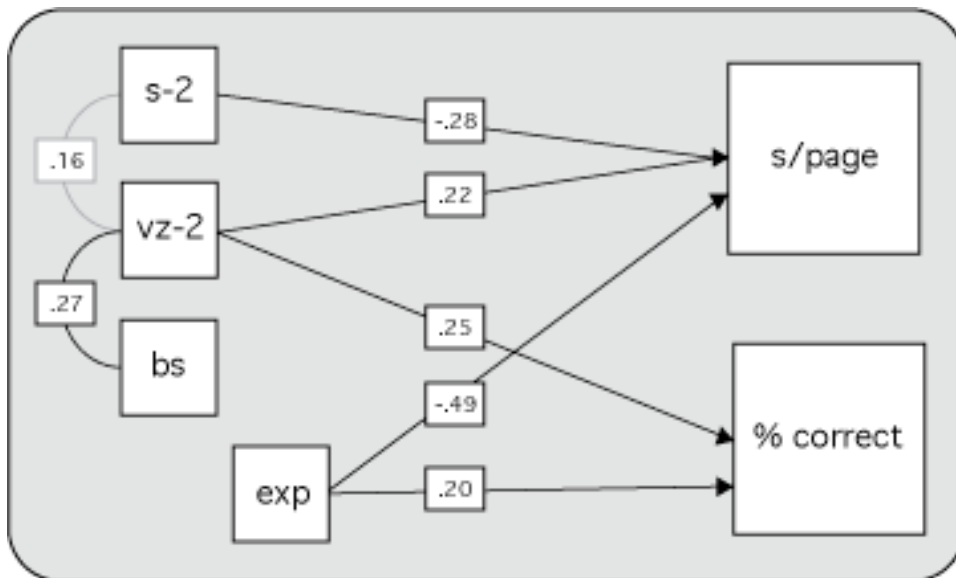
Figure 8. Obtained relationships from Experiment 1.
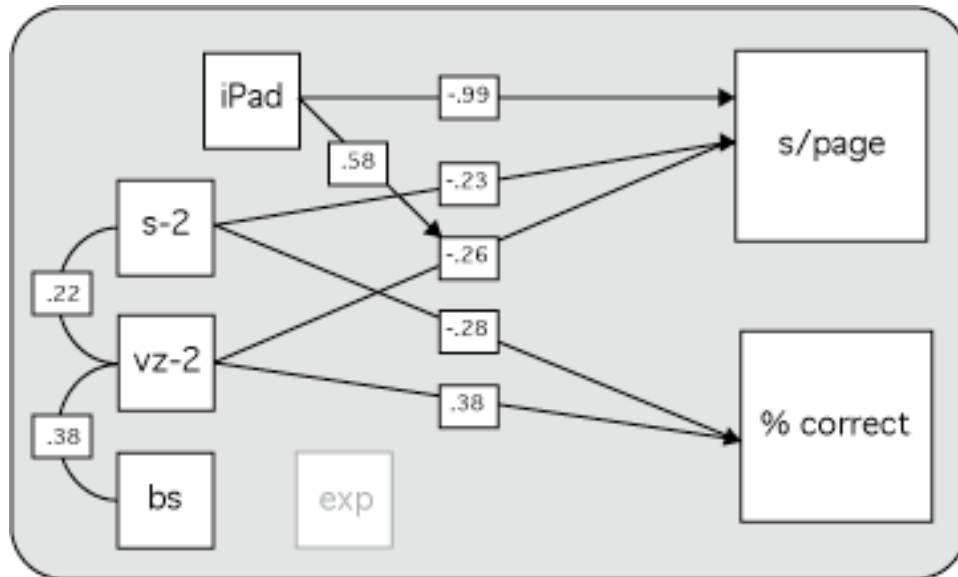
Figure 9. Obtained relationships from Experiment 2.

efficiency (s/page), while spatial visualization (vz-2) predicted effectiveness. On larger screens, spatial visualization (vz-2) also predicted slower performance, while on smaller screens, spatial visualization predicted faster performance. In the first experiment, experience played a large role, while in the second experiment it did not.

The resulting model, then, should suggest theoretically that spatial visualization plays a part in predicting effectiveness. That relationship may be moderated by question difficulty, though that effect was not investigated in the experiments presented here. The model should also show that spatial orientation ability predicts efficiency. That effect does not seem to be affected by screen size, though the effect of spatial visualization ability on efficiency is moderated by screen size. Since no evidence was found that working memory provides additional prediction above these other predictors on either outcome measure, it has been relegated to a minor player which may predict effectiveness on small screens (though the effect did not seem to be moderated by screen size, unlike the spatial visualization effect). The summary of these relationships is shown in Figure 10.

64
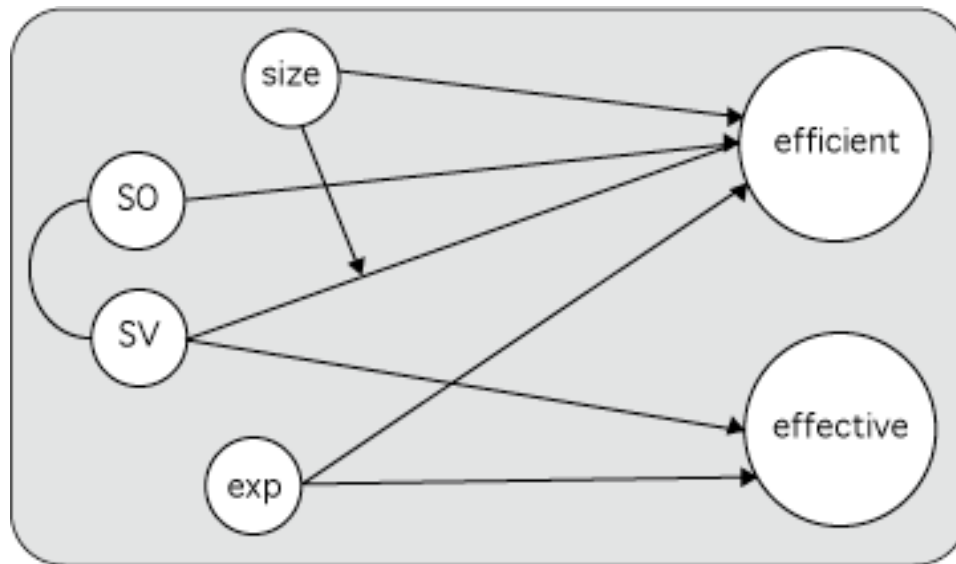
Figure 10. Combined theoretical diagram.

This model instantiates the idea that different measures of cognitive ability, or in this case spatial ability, affect different measures of usability differently. The constructs that make up usability performance relate to different kinds of spatial ability, suggesting that they depend on different cognitive processes. It also suggests that screen size affects not just the efficiency of performance, but the processes used to create that efficiency; not only do smaller screens slow performance, they also increase the importance of spatial visualization over larger screens.

Conclusions

In these experiments, I found that the spatial abilities of participants predicted their performance on tasks like those used to measure interface usability. In general, different abilities predicted different aspects of usability; spatial visualization predicted effectiveness, while spatial orientation predicted efficiency. Spatial visualization had mixed effects on efficiency depending on the screen size of the

device being evaluated. I did not find evidence that working memory added to the prediction of either effectiveness or efficiency.

Limitations

The two experiments were conducted using an undergraduate population, so specific generalizations about the larger population of computer users are difficult to make. The lack of educational and age-related diversity in the sample, however, can also be a source of control. Differences in this sample are unlikely to be due to age of onset for computer use, for instance, or non-cognitive age differences.

Implications for Spatial Ability Measurement

The original idea that spatial visualization was a combination of spatial orientation and working memory was roundly unsupported. Spatial visualization provided a unique contribution above both other measures. In the case of working memory, there was some overlap between the two predictors, but spatial visualization added more unique prediction above working memory than working memory did above spatial visualization, suggesting that some common processes were at work, but not that spatial visualization was a subset of working memory. This finding suggests that differences in spatial ability are not attributable to working memory differences.

Because of the level of prediction that was due to spatial ability, I would recommend that usability testers try to assess spatial ability as part of the usability evaluation. The tests I used for these experiments are easy to administer and do not require any specialized equipment – in fact, the tests of spatial visualization and spatial orientation can be administered on paper instead of using a web-delivery

system if required. The most specialized was Block Span, which was the measure of working memory, which does require sequential visual presentation of stimuli. Because of this ease of administration, gathering data on spatial visualization and spatial orientation ability should be quick (both tests take less than three minutes to perform) and relatively straightforward for researchers who are interested in making statements about whether their interventions are stronger than the cognitive variation among individuals in the target population. Additionally, these tests could be used in addition to demographic or role factors to select users for usability testing who are representative of the cognitive abilities of the population who will use the software.

The interfaces I used in this experiment have more information available at one time and are more familiar than navigational or medical devices. It is quite possible that medical devices with displays considerably smaller than the iPod Touch would be sensitive to differences in spatial ability.

Note that spatial visualization and spatial orientation ability interacted with interface usability in this study, but that the interfaces were not unusable as a result. In fact, all of the interfaces were quite usable, and participants were successful in completing the usability tasks, but task performance still interacted with spatial ability. This finding suggests that interaction between ability and performance is normal and does not prevent devices from being usable by a broad population.

Implications for Usability Measurement

The philosophical problem with usability measurement is that dissociable constructs are not necessarily measured in dissociable ways. Effectiveness and efficiency of the same interface are different things, but the ways we measure them

may contain the same components. If we accept that the measurements are going to contain the same terms, we cannot state whether the constructs are empirically related; there will be a relationship, but there is no way to tell whether it is due to a relation between the constructs or is a mathematical artifact. If the same component is included in two measurements (such as the number of questions finished being both the numerator of the efficiency measure and the denominator of the effectiveness measure), those measurements cannot be independent of each other.

In this investigation, I avoided using any measures of efficiency or effectiveness that contained the same terms, such as total number of correct answers (a measure of effectiveness) and correct answers per minute (a measure of efficiency). However, I also avoided using measures of efficiency and effectiveness that were highly correlated with each other, such as number of questions correctly answered (a measure of effectiveness) and total number of questions answered (a measure of efficiency). Because, especially in Experiment 1, not all participants finished, the number of questions answered correctly depended on the total number of questions answered.

The results of both experiments suggested that there exist non-overlapping measures of efficiency and effectiveness. In the second experiment, measures of efficiency differed between devices, while measures of effectiveness did not. This finding suggests that any difference that exists between the iPad and iPod Touch is mainly in efficiency, not in effectiveness; users can find the same information on both devices.

I would advise future researchers to be clear on what measures of efficiency and effectiveness they are interested in, preferably before starting to experiment. Ideally, a standard set of metrics should be possible for comparing interfaces with similar characteristics. The current standard does not specify how the constructs it specifies should be measured, leaving researchers many options of varying and unknown quality.

Regardless of which construct one is interested in, spatial abilities make a difference in performance on usability tests. This finding implies that it is important to find a set of usability test participants who are diverse not just on background factors but also on cognitive abilities. Especially in cases where different screen sizes are involved, a wide range of spatial abilities in one's test population is important for quantitative estimates of performance. Since spatial ability may also affect which cognitive processes are used for interaction with a particular interface, it could also be important for qualitative measures of interaction strategy, for instance.

Even more basically, the difference in cognitive processes that appears to occur on smaller screens means that there may not be a linear relationship between screen size and performance; the usability of a web site on a small screen may not be indicated by its usability on a large screen, necessitating testing on a wide range of devices, platforms, and screen sizes.

Future Directions

There are both short-term and long-term future directions for this research and its implications. In the short run, extending the findings from navigation tasks to other

applications and devices is fairly straightforward, though the particulars of the model may change based on the characteristics of the interfaces and tasks being investigated.

In the long run, screen size may not be the most important consideration. Future interfaces may use different kinds of screen technology, and will thus not have a particular screen size. In the extreme case, some interfaces, like the third generation iPod Shuffle (http://www.apple.com/support/ipodshuffle/3rd_generation/), do not have any visual interface at all. In cases like this, it seems likely that navigation tasks would require more visualization resources. Many music players lack a visual interface beyond a few buttons, though few are quite that extremely physically featureless. The challenge of designing a mass-market device without a visual interface, of course, is to try to avoid navigation as much as possible. Generally, because navigation requires cognitive effort, that effort is mitigated by providing as much information about the current state of the interface as possible, or by avoiding navigation.

Other types of screen technology, like projection, may depend more on screen resolution, which will set the preferred display size, than on physical display characteristics. At some point, however, especially on fixed displays like cellular phones, the pixel density becomes high enough that more pixels will not be legible. Actual retinal displays would be able to supplant the entire visual field with artificial information, and would thus allow for a much greater range of detail and a more immersive experience.

In the short term, this model should be applied to a wider range of devices and populations to see if it generalizes beyond a college student population. In the long

run, the variation due to individual differences among users should be taken into

account both when one designs an interface and when one designs a usability test.

Appendix A: Experiment 1 IRB Approval Information

Project Title: Cognitive Abilities and Usability
Primary Investigator: Kent L. Norman
Student Investigator: Susan G. Campbell
IRB approval number: 08-0621
PAS reference number: 2220
Date approved: December 1, 2008
Expiration Date: December 1, 2009

Consent Form

Cognitive Abilities and Usability	
Why is this research being done?	*This is a research project being conducted by Dr. Kent Norman and Susan G. Campbell at the University of Maryland, College Park. We are inviting you to participate in this research project because you are over 18. The purpose of this research project is to understand how people navigate websites on the internet.*
What will I be asked to do?	*In this experiment, you will be asked to answer a series of factual questions based on information that is available on Wikipedia. You will then be asked to take some simple cognitive tests that look at your ability to do spatial tasks.*
What about confidentiality?	*We will do our best to keep your personal information confidential. To help protect your confidentiality, we will store this consent form separately from the data we collect during the experiment, in such a way that no one could use our data to identify you or your answers. All of the information we collect on paper is kept in a locked room, and all digital information we collect is kept on a secure server.* *If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.*
What are the risks of this research?	*There are no known risks associated with participating in this research project.*
What are the benefits of this research?	*This research is not designed to help you personally, but the results may help the investigator learn more about how people use the internet and how spatial abilities interact. These results could help designers make better websites.*
Do I have to be in this research? May I stop participating at any time?	*Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.*
What if I have questions?	*This research is being conducted by Dr. Kent Norman of the Department of Psychology at the University of Maryland, College Park. If you have any questions about the research study itself, please contact Dr. Norman by phone: 1-301-405-5924, or by email: kent_norman@lap.umd.edu, or contact his office:* **Biology-Psychology room 3123F** **University of Maryland** **College Park, MD 20742** **(e-mail) klnorman@umd.edu** **(telephone) 301-405-5924** *If you have questions about your rights as a research subject or wish to report a research-related injury, please contact:*

73

	Institutional Review Board Office, **University of Maryland,** **College Park, Maryland, 20742** **(e-mail) irb@deans.umd.edu** **(telephone) 301-405-0678** *This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.*
Statement of Age of Subject and Consent	☐ I am 18 or over and have freely volunteered to participate in this project. ☐ I have been informed in advance as to what my tasks would be and what procedure would be followed. ☐ I understand that there are no known risks to my participation of this research, and that this research is not designed to help me personally. ☐ I am aware that I have the right at any time to withdraw consent and discontinue participation at any time. ☐ I am aware that I may contact the experimenter or the Institutional Review Board office if I have any questions about this experiment or my rights as a subject.
Name and Date	First Name: [ ] Last Name: [ ] Today's date: [ ]

[Agree and Submit Information to Continue] [I Do Not Agree]

L

A

P

Appendix B: Experiment 1 Question List

1. What kind of bird may only the Seigneur of Sark keep on that island?

2. What was the name of the project that Wikipedia was spun off from?

3. In what war did a faction in England use *rosa gallica officinalis* as its symbol?

4. Who was both elected chairman of the British Interplanetary Society and named Sri Lankabhimanya by the government of Sri Lanka?

5. What discipline does the Principle of Predictive Aiding refer to?

6. Which of the four provinces (not counties) of modern Ireland is the medieval kingdom of Mide part of?

7. What group of people in Malaysia were originally believed to be the only tribe to use lucid dreaming to ensure mental health?

8. What is encoding used for replacement of Cyrillic letters with Roman letters to do text messaging in Russian called, after a word that "sounds funny" in Russian?

9. Which hero of Russia commanded foot soldiers against German knights in a battle fought on a frozen lake?

10. What company brought the first major lawsuit against Facebook?

Appendix C: Experiment 1 Descriptive Statistics

	N	Mean	Min	Max	SD	Skew	Kurtosis
S-2 Score	103	11.223	2	20	4.13	0.153	-0.860
VZ-2 Score	103	6.01	1	10	2.23	-0.328	-0.722
Block Span	97	51.04	30	81	11.44	0.170	-0.276
Computer Experience	102	8.069	4	10	1.59	-0.788	0.056
WWW Experience	103	8.447	4	10	1.47	-0.962	0.551
Wikipedia Experience	103	7.350	1	10	1.86	-0.667	0.163
Average Correct Time	103	251.3	77	971	141.66	2.706	10.664
Average Correct Clicks	84	12.54	1	44.67	6.61	2.007	6.345
Average Time per Click	84	24.69	11.92	76.95	8.93	2.664	13.320
Number Completed	103	7.04	2	10	2.65	-0.419	-1.059
Number Correct	103	5.34	1	10	2.32	0.006	-0.631
Percent Correct	103	0.765	0.2	1	0.20	-0.562	-0.500

Appendix D: Experiment 2 IRB Approval Information

Project Title: Cognitive Abilities and Usability
Primary Investigator: Kent L. Norman
Student Investigator: Susan G. Campbell
IRB approval number: 10-0009
PAS reference number: 2835
Date approved: January 11, 2010
Expiration Date: January 11, 2013

Consent Form

Cognitive Abilities and Usability	
Why is this research being done?	*This is a research project being conducted by Dr. Kent Norman and Susan G. Campbell at the University of Maryland, College Park. We are inviting you to participate in this research project because you are over 18. The purpose of this research project is to understand how people navigate websites on the internet.*
What will I be asked to do?	*In this experiment, you will be asked to answer a series of factual questions based on information that is available on Wikipedia. You will also be asked to take some simple cognitive tests that look at your ability to do spatial tasks.*
What about confidentiality?	*We will do our best to keep your personal information confidential. To help protect your confidentiality, we will store this consent form separately from the data we collect during the experiment, in such a way that no one could use our data to identify you or your answers. All of the information we collect on paper is kept in a locked room, and all digital information we collect is kept on a secure server.* *If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.*
What are the risks of this research?	*There are no known risks associated with participating in this research project.*
What are the benefits of this research?	*This research is not designed to help you personally, but the results may help the investigator learn more about how people use the internet and how spatial abilities interact. These results could help designers make better websites.*
Do I have to be in this research? May I stop participating at any time?	*Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.*
What if I have questions?	*This research is being conducted by Dr. Kent Norman of the Department of Psychology at the University of Maryland, College Park. If you have any questions about the research study itself, please contact Dr. Norman by phone: 1-301-405-5924, or by email: kent_norman@lap.umd.edu, or contact his office:* **Biology-Psychology room 3123F** **University of Maryland** **College Park, MD 20742** **(e-mail)** klnorman@umd.edu **(telephone) 301-405-5924** *If you have questions about your rights as a research subject or wish to report a research-related injury, please contact:* **Institutional Review Board Office,** **University of Maryland,** **College Park, Maryland, 20742** **(e-mail)** irb@deans.umd.edu **(telephone) 301-405-0678** *This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.*

Statement of Age of Subject and Consent	☐ I am 18 or over and have freely volunteered to participate in this project. ☐ I have been informed in advance as to what my tasks would be and what procedure would be followed. ☐ I understand that there are no known risks to my participation of this research, and that this research is not designed to help me personally. ☐ I am aware that I have the right at any time to withdraw consent and discontinue participation at any time. ☐ I am aware that I may contact the experimenter or the Institutional Review Board office if I have any questions about this experiment or my rights as a subject.
Name and Date	First Name: _____ Last Name: _____ Today's date: _____

Agree and Submit Information to Continue | I Do Not Agree

Appendix E: Experiment 2 Question List

1. What company brought the first major lawsuit against Facebook?

2. What kind of bird may only the Seigneur of Sark keep on that island?

3. What was the name of the project that Wikipedia was spun off from?

4. What group of people in Malaysia were originally believed to be the only tribe to use lucid dreaming to ensure mental health?

5. In what war did a faction in England use *rosa gallica officinalis* as its symbol?

Appendix F: Experiment 2 Descriptive Statistics

	N	Min	Max	Mean	St. Dev.	Skew	Kurtosis
Computer Experience	96	5	10	8.56	1.33	-0.77	0.09
Internet Experience	96	5	10	8.76	1.30	-1.11	0.95
Wikipedia Experience	96	2	10	8.10	1.61	-0.94	1.25
iPod/iPhone Experience	96	1	10	6.64	2.66	-0.56	-0.73
iPad Experience	96	1	9	2.94	2.20	0.92	-0.11
Spatial orientation	96	3	20	11.43	3.79	0.09	-0.54
Spatial visualization	92	0	10	5.67	2.41	-0.01	-0.90
Working memory	92	14	93	52.47	13.11	0.11	0.59
Total Steps	94	11	67	28.89	11.36	0.64	0.22
Percent Correct	96	.20	1.00	0.78	0.21	-0.72	0.49
Time per Page (s)	94	19.78	123.29	44.47	15.90	1.75	5.72

References

Abdinnour-Helm, S.F., Chaparro, B.S., & Farmer, S.M. (2005). Using the End-User Computing Satisfaction (EUCS) instrument to measure satisfaction with a web site. *Decision Sciences 36*(2), 341-364. doi:10.1111/j.1540-5414.2005.00076.x

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal 11*(4), 325-338. doi:10.1023/A:1025869312943

Ally, M. Using learning theories to design instruction for mobile learning devices. In *Mobile learning anytime everywhere.* Attewell, J., & Savill-Smith, C., Eds. London: Learning and Skills Development Agency, 5-8.

Anderson, A.A. (1996). Predictors of computer anxiety and performance in information systems. *Computers in Human Behavior, 12*(1), 61-77. doi:10.1016/0747-5632(95)00019-4

Baddeley, A.D. (2001). Is working memory still working? *American Psychologist 56*(11), 851-864. doi:10.1037/0003-066X.56.11.851

Beckers, J.J., & Schmidt, H.G. (2001). The structure of computer anxiety: A six-factor model. *Computers in Human Behavior, 17*, 35-49. doi:10.1016/S0747-5632(00)00036-4

Beckers, J.J., & Schmidt, H.G. (2003). Computer experience and computer anxiety. *Computers in Human Behavior, 19*, 785-797. doi:10.1016/S0747-5632(03)00005-0

Bevan, N., & MacLeod, M. (1994). Usability measurement in context. *Behaviour and Information Technology 13*, 132-145. doi: 10.1080/01449299408914592

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In *Usability measurement in industry*. Jordan, P.W., Thomas, B., Weerdmeester, B.A., & McClelland, I., Eds. London: Taylor and Francis, 189-206.

Brosnan, M.J. (1998). The impact of computer anxiety and self-efficacy upon performance. *Journal of Computer Assisted Learning, 14*, 223-234. doi:10.1046/j.1365-2729.1998.143059.x

Campbell, S.G., & Norman, K.L. (2007). Strategy, spatial visualization ability, and performance on website navigation. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 832-836.

Carroll, J.B. (1993). *Human Cognitive Abilities: A survey of factor- analytic studies*. New York: Cambridge University Press.

Chavez, N.M. (2002). Individual differences in verbal working memory, visuo-spatial working memory, and metacognition: Learning from text in a hypertext environment. Unpublished dissertation, Texas Tech University, 2002.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta- analysis of experimental studies. *Human-Computer Interaction 11* (2), 125-156. doi:10.1207/s15327051hci1102_2

Chua, S.L., Chen, D.T., & Wong, A.F.L. (1999). Computer anxiety and its correlates: A meta-analysis. *Computers in Human Behavior, 15*, 609-623. doi:10.1016/S0747-5632(99)00039-4

Cui, Y., & Roto, V. (2008, April). How people use the web on mobile devices. *WWW 2008*, Beijing, China, 905-914.

Dearden, A., Lauener, A., Slack, F., Roast, C., & Cassidy, S. (2006). Make it so! Jean-Luc Picard, Bart Simpson and the Design of E-Public Services. PDC 2006 - Proceedings of the ninth Participatory Design Conference 2006, 67-76. doi: 10.1145/1147261.1147272

DeStefano, D., & LeFevre, J.A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior 23*, 1616-1641. doi:10.1016/j.chb.2005.08.012

Downing, R.E., Moore, J.L., & Brown, S.W. (2006). The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behavior 21*, 195-209.

Dünser, A., & Jirasko, M. (2005). Interaction of hypertext forms and global versus sequential learning styles. *Journal of Educational Computing Research 32*(1), 79-91. doi:10.2190/1J25-LWQF-PQ3W-LABM

Ekstrom, R.B., French, J.W., & Harman, H.H. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Frøkjaer, E., Hertzum, M., & Hornbaek, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? CHI Letters, 2(1), 345-352.

Graff, M. (2005). Differences in concept mapping, hypertext architecture, and the analyst-intuition dimension of cognitive style. *Educational Psychology 25*(4), 409-422. doi: 10.1080/01443410500041813

Hackbarth, G., Grover, V., & Yi, M.Y. (2003). Computer playfulness and anxiety: positive and negative mediators of system experience effect on perceived ease of use. *Information and Management 40*, 221-232. doi:10.1016/S0378-7206(02)00006-X

Harbison, J.I., Dougherty, M.R., & Bunting, M.F. (2008). Block Span: A new automated working memory span task. Poster presented at the 49th Annual Meeting of the Psychonomic Society, November 2008.

Höök, K., Sjölinder, M., & Dahlbäck, N. (1996). *Individual differences and navigation in hypermedia*. SICS Research Report R96:01. Swedish Institute of Computer Science ISSN 0283-3638.

Horn, J.L., & Blankson, N. (2005). Foundations for better understanding cognitive abilities. In *Contemporary Intellectual Assessment*, 2nd Ed., Flanagan, D.P. & Harrison, P.L. (Eds.), New York: Guilford Press, 41-68.

Hornbaek, K., & Law, E.L. (2007). Meta-analysis of correlations among usability measures. *CHI 2007 Proceedings - Empirical Models,* 617-626.

John, B.E., & Marks, S.J. (1997). Tracking the effectiveness of usability evaluation methods. *Behaviour and Information Technology 16*(4/5), 188-202. doi:10.1080/014492997119789

Kirakowski, J. (1996). The software usability measurement inventory: Background and usage. In *Usability measurement in industry*. Jordan, P.W., Thomas, B., Weerdmeester, B.A., & McClelland, I., Eds. London: Taylor and Francis, 168-188.

Kissel, G.V. (1995). The effect of computer experience on subjective and objective software usability measures. *CHI 1995 Proceedings*, 284-285.

Knight, M.B., & Pearson, J.M. (2005). The changing demographics: The diminishing role of age and gender in computer usage. *Journal of Organizational and End-User Computing 17*(4), 49-65. doi:10.4018/joeuc.2005100103

Kujala, S., & Kauppinen, M. (2004). Identifying and selecting users for user-centered design. *NordiCHI '04 Proceedings of the third Nordic conference on Human-computer interaction*, 297–303. doi: 10.1145/1028014.1028060

Laguna, K., & Babcock, R.L. (1997). Computer anxiety in young and older adults: Implications for human-computer interactions in older populations. *Computers in Human Behavior, 13*(3), 317-326. doi:10.1016/S0747-5632(97)00012-5

McDonald, S., & Stevenson, R.J. (1998a). Effects of text structure and prior knowledge of the learner on text navigation in hypertext. *Human Factors 40*(1), 18-27. doi:10.1518/001872098779480541

McDonald, S., & Stevenson, R.J. (1998b). Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers 10*, 129-142. doi:10.1016/ S0953-5438(98)00017-4

Michael, W.B., Guilford, J.P., Furchter, B., & Zimmerman, W.S. (1957). The description of spatial-visualization abilities. *Educational and Psychological Measurement, 17*, 185-199. doi: 10.1177/001316445701700202

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97. doi:10.1037/h0043158

Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General 130*(4), 621-640. doi:10.1037/0096-3445.130.4.621

Murphy, E.D., & Lorenz, B. (2001). Is spatial-visualization ability a stronger predictor of performance for males than for females on computer-based tasks? *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting - 2001*, 931-935.

Nielsen, J. (1992). The usability engineering life cycle. *Computer 25* (3), 12-22. doi:10.1109/2.121503

Nielsen, J., & Levy, J. (1994). Measuring usability: Preference vs. performance. *Communications of the ACM 37*(4), 66-76.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *CHI '90 Proceedings*, 249-256.

Norman, K.L., & Butler, S.A. (1989). Search by uncertainty: Menu selection by target probability. University of Maryland Center for Automation Research Technical Report CAR-TR-432.

Norman, K.L., & Panizzi, E. (2006). Levels of automation and user participation in usability testing. *Interacting with Computers 18*, 246-264. doi:10.1016/j.intcom.2005.06.002

Norman, K.L., Shneiderman, B., & Harper, B. (2003). QUIS: The Questionnaire for User Interaction Satisfaction, version 7. University of Maryland Human Computer Interaction Laboratory.

Ozer, D.J. (1987). Personality, intelligence, and spatial visualization: Correlates of mental rotations test performance. *Journal of Personality and Social Psychology 53*(1), 129-134. doi: 10.1037/0022-3514.53.1.129

Pak, R., Rogers, W.A., and Fisk, A.D. (2006). Spatial ability subfactors and their influences on a computer-based information search task. *Human Factors 48*(1), 154-165. doi: 10.1518/001872006776412180

Pellegrino, J.W., Alderton, D.L., and Shute, V.J. (1984). Understanding spatial ability. *Educational Psychologist 19*(4), 239-253. doi:10.1080/00461528409529300

Petrie, H., Hamilton, F., & King, N. (2004). Tension, what tension? Website accessibility and visual design. *WWW2004*, 13-18. doi: 10.1145/990657.990660

Reed, W.M., Oughton, J.M., Ayersman, D.J., Ervin Jr., J.R., Giessler, S.F. (2000). Computer experience, learning style, and hypermedia navigation. *Computers in Human Behavior 16*, 609-628. doi:10.1016/ S0747-5632(00)00026-1

Rozell, E.J., & Gardner, W. L. (2000). Cognitive, motivation, and affective processes associated with computer-related performance: A path analysis. *Computers in Human Behavior 16*, 199-222. doi:10.1016/S0747-5632(99)00054-0

Salthouse, T.A., Babcock, R.L., Mitchell, D.R.D., Palmon, R., & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence 14*, 187-230. doi: 10.1016/0160-2896(90)90004-D

Schumacher, P., & Morahan-Martin, J. (2001). Gender, internet and computer attitudes and experiences. *Computers in Human Behavior 17*, 95-110. doi:10.1016/S0747-5632(00)00032-7

Seagull, F.J., & Walker, N. (1992). The effects of hierarchical structure and visualization

ability on computerized information retrieval. *International Journal of Human-Computer Interaction 4* (4), 369-385. doi:10.1080/10447319209526049

Seffah, A., Donyaee, M., Kline, R.B., & Padda, H.K. (2006). Usability measurement and metrics: A consolidated model. *Software Quality Journal 14*, 159-178. doi:10.1007/s11219-006-7600-8

Simon, S.J. (2001). The impact of culture and gender on web sites: An empirical study. *The DATA BASE for Advances in Information Systems 32*(1), 18-37. doi:10.1145/506740.506744

Smith, B., Caputi, P., & Rawstorne, P. (2000). Differentiating computer experience and attitudes toward computers: an empirical investigation. *Computers in Human Behavior 16*, 59-81. doi: 10.1016/S0747-5632(99)00052-7

Smith, P.A. (1996). Towards a practical measure of hypertext usability. *Interacting with Computers 8*(4), 365-381. doi:10.1016/ S0953-5438(97)83779-4

Van Welie, M., van der Veer, G.C., & Eliëns, A. (1999). Breaking down usability. In *Proceedings of Interact'99* (Edinburgh, 30th August - 3rd September 1999). IOS Press, 613–620.

Vicente, K.J., Hayes, B.C., & Williges, R.C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors 29*(3), 349-359.

Wenger, M.J., & Payne, D.G. (1996). Comprehension and retention of nonlinear text: Considerations of working memory and material- appropriate processing. *The American Journal of Psychology, 109* (1), 93-130. doi:10.2307/1422929

Zhang, H., & Salvendy, G. (2001). The implications of visualization ability and structure preview design for web information search tasks. *International Journal of Human-Computer Interaction 13*(1), 75-95. doi:10.1207/S15327590IJHC1301_5