*Grant Final Report*

**Grant ID: R21HS018831**

# Improving Meaningful Access of Internet Health Information for Older Adults

**Inclusive Project Dates: 04/01/11 – 09/30/13**

**Principal Investigator**:
Sara J. Czaja

**Team Members:**
Joseph Sharit
Peter Pirolli
Mario A. Hernández
Les Nelson
Peggy González

**Performing Organization:**
University of Miami Miller School of Medicine

**Federal Project Officer:**
Rebecca Roper

**Submitted to:**
**The Agency for Healthcare Research and Quality (AHRQ)**
**U.S. Department of Health and Human Services**
**540 Gaither Road**
**Rockville, MD 20850**
**www.ahrq.gov**

# Abstract

**Purpose:** The purpose of this study was to refine a set of software aiding tools so they could be used by health consumers, particularly older adults, to complete Internet-based health management tasks and to evaluate their usability and impact on the performance of the tasks.

**Scope:** Although the Internet provides an efficient means to access a vast amount of information many consumers such as older adults often have difficulty finding, filtering, integrating and comprehending the health information they find online. Phase 1 involved 23 older adults aged 65+ years; Phase 2 involved 80 adults, aged 30 to 85 years. Both groups had computer and Internet experience.

**Methods:** Phase 1 involved a task analysis of the tools and focus groups. In Phase 2, participants were randomly assigned to one of four conditions, two of which involved the aiding tools, and performed Internet-based health information seeking tasks.

**Results:** The findings from the evaluation study indicated that overall the participants perceived the tools as useful and of potential value in aiding their information seeking activities. However, performance on the information seeking problems was generally lower among those who used the tools.

**Key Words:** Internet; search engine; health, information seeking; sensemaking; user-centered design

# Final Report

## Scope

The Internet is increasingly becoming the focal e-health application by which consumers find health information (Pew Internet & American Life Project, 2013). Efficient access to this vast repository of information is a particularly important consideration for older adults as many of them have multiple health issues. These users are thus likely to greatly benefit from information that can inform better health decision-making. However, the process of online gathering of health information can be an especially daunting task for many users as it requires seeking, filtering, and integrating relevant and valid sources of health information. In fact, evidence suggests that the effectiveness and utility of many consumer e-health tools that are available on the Internet is uneven across user groups. One group for whom this is particularly true is older adults, especially minority older adults or those from a lower socio-economic status. For example, older adults report more anxiety about technology adoption, less comfort with technology, and typically have more difficulty learning to use and operate new devices and systems (e.g., Czaja & Sharit, 1998; Czaja et al., 2006). Studies have also shown (e.g., Mead et al., 1997; Sharit et al., 2008) that older adults are generally less successful than younger adults in searching the Web for specific information and that they use less efficient search strategies (Czaja, Sharit, & Nair, 2008; Sharit, Hernandez, Czaja & Pirolli, 2008). Further, to date, most "e-health and e-government" applications have been designed without consideration of the needs and characteristics of diverse user groups and thus have increased the complexity of the demands of healthcare engagement (U.S. DHHS, 2006).

One strategy for confronting this problem is to offer users tools that could aid in extracting and filtering Internet health information, and which could provide better ways for visualizing, comprehending, and organizing the information in support of decision making related to health management. To date, a smattering of tools that help people find, interpret and organize information have been developed. Overall, these tools offer the potential for promoting cognitive abilities such as reasoning and knowledge that remain stable or increase with age, and also for compensating for abilities such as memory and attention that decline with age. However, they have only been evaluated with small samples of "experts" who are younger, have extensive Internet search experience, and for non-health-related searches (e.g., Chi, Hong, Gumbrecht, & Card, 2005; Hong, Chi, Budiu, Pirolli, & Nelson, 2008; and Budiu, Pirolli, & Hong, 2009). The objective of the proposed project was to evaluate, and refine these tools so that they are usable and useful for diverse older adult populations and enable them to effectively engage in Internet health-information seeking, which should benefit the larger community of users as well (Fisk et al., 2009). The outcomes of this exploratory study are intended to support the development of a large scale demonstration project to examine how availability of these software tools influences the use of Internet-based health applications and, in turn, impacts on more distal outcomes such as physician/patient relationships, use of health care services, adherence, and health self-efficacy.

## Overview of the Study Design

The study consisted of two phases, which correspond, respectively, to the two principal aims of the study: a tool refinement phase (Phase 1) in which the tools were tailored for health information seeking and health consumers through a user-centered iterative design process and a laboratory-based formal evaluation phase (Phase 2). Phase 1 involved adapting the software aiding tools (described below) developed at PARC to health information-seeking tasks through an iterative design process involving an expert panel of analysts with experience in human factors engineering, cognitive aging, cognitive engineering researchers, and older adults. These refined tools were then evaluated by a small and diverse group of older users for the purpose of early identification of design and usability problems. Phase 2 of the study was a formal evaluation of the tools. It involved a randomized trial where a sample of users used the tools to find and "make sense" of Internet-based health information to solve "ecologically valid" health problems. Their performance was compared to a sample of users who performed the study tasks unaided. Participants who used the tools were also asked to rate the potential usefulness and various aspects of usability of each tool.

# Phase 1—Focus Groups and Task Analysis

## Purpose

## Description of the Aiding Tools

A fundamental component of user-centered design is an iterative design approach where "users" are involved early in the design process. Thus, the primary purpose of the focus groups was to obtain data from a sample of older adults about perceived benefits and concerns associated with four different tools designed to aid their Internet health information-seeking processes. In addition, the research team performed a task analysis of the tools to identify potential usability problems. The four tools that were considered, which were either developed or refined by our collaborators at Palo Alto Research Corporation, included: Mr. Taggy Search Engine, SparTag.us Notebook, Automated Highlighting, and an Automated Glossary. Each of these tools is described below.

**Mr. Taggy Search Engine**. When using the Internet to find information about a health issue, an individual typically represents the problem in terms of a query and enters this query into a search box. Once the search process is initiated the individual is confronted with a listing of websites that they need to filter and select to obtain information. Mr. Taggy provides a means for users to identify content relevant to their search problem as well as associated websites by providing a list of search results together with a side list of search tags. The tags allow users to refine their search by indicating "thumbs up or thumbs down" to the tags. This helps them filter search results and enables more rapid exploration of the search space. This tool also provides

recommendations for related tags and search terms and helps users learn the vocabulary for a particular domain, which can help them refine later searches.

**SparTag.us Notebook**. This tool allows a user to build a "notebook." As users browse through web pages and find material of interest, they click on a word within that paragraph. A box then appears and the entire paragraph is placed into the box. The user then labels that box with a category name. Essentially, the box becomes a "notebook" of information related to a topic, which can be used to collect, organize and save material of interest within web pages, including the source website, and build a collective knowledge space on a particular topic. It also helps users organize and navigate material that is useful and related to a particular topic. Such functionality provides a kind of external working memory aid to support remembering information and arranging it into more personally meaningful organizational categories. It thus can potentially improve the ability of users, and particularly older users, to integrate information in support of decision-making activities.

**Automated Highlighting**. Identification of relevant information within a webpage requires selective attention and the ability to suppress irrelevant information. The automated highlighting tool supports this process by highlighting words and phrases that are related to the information interests of the user. As the user proceeds to scan web pages returned from a search, this tool will automatically highlight, in color, information within these web pages that includes not only the phrase entered into the tool's search box but also text and phrases related to this term, thus helping the user to skim through the webpage. The user can also move this information into their SparTag.us notebook.

**Pop-Up Glossary.** One problem consumers confront when they access many health websites is overly technical language. The automated glossary tool helps to enhance comprehension of technical language by providing simple common language translations of medical terminology or "medicalese" when the person places the cursor over the word or concept that is unfamiliar or difficult to understand.

# Methods

## Focus Groups

The sample included twenty-three older adults ($M$ = 73.1 yrs., $SD$ = 8.1). All participants were English speaking and had computer and Internet experience. The sample consisted of nine males and 14 females. With respect to education, two of the participants had a high school degree, nine had some college and 12 had a college degree or beyond. The majority (n =21) reported that they were in good to very good health. Three separate focus groups were conducted with eight, eight and seven participants respectively. All participants were compensated $30.00 for their participation. All participants initially read and signed an informed consent form that was approved by the University's Institutional Review Board, and which indicated that all discussions, whether group or individual, would be audio-taped.

Within each of the three focus groups there was a leader and two facilitators. The facilitators helped with the administration of questionnaires and took notes for the session. Each focus group consisted of the following components:

1. Introduction to the focus group.

2. Short group discussion.

3. Demonstration and overview of the tools.

4. Individual discussions with participants and questionnaire administration.

5. Integration of comments from the discussions.

6. Final group discussion.

In the Introduction, the group of participants was given an overview by the leader about the current role of the Internet in finding health information in terms of satisfying a wide range of health information needs. The issue concerning the difficulty users might encounter integrating and making sense of information they might find across a wide range of websites was then addressed.

Next, participants were informed that the primary purpose of the focus group was to solicit their initial impressions about some software tools that will be modified to target older users, such as themselves, in terms of helping them find, understand, and organize the information that users confront when searching or browsing the Internet. Participants were encouraged to offer any comments about the tools' features or suggestions for redesign.

The short group discussion was guided by the following questions:

1. Have you searched for health information on the Internet? If yes:

   (a) Where do you go to find health information?

   (b) What kind of health information do you usually look for?

2. Have you encountered problems while searching for health information? If yes:

   (a) What kinds of problems?

   (b) Do you generally have trouble remembering where you found information that you want to keep handy or understanding the information you find?

The facilitator, who was very knowledgeable of the software tools, then provided the participants with a PowerPoint (PP) presentation of the four tools, one by one. The 28 slides provided information (including snapshots, in figure form) about how each of the tools works to demonstrate their purported benefits. Figure 1 illustrates one of the Mr. Taggy slides, demonstrating features related to tags and the ability to refine searches using thumbs up/thumbs down mechanisms. With the automated highlighting tool (Figure 2) participants were shown how keyword search results are highlighted and could include related or associated text in addition to their keywords.

During the individual discussions that followed the PP presentation, those who were not being interviewed were administered three questionnaires: a Demographics questionnaire, a Technology, Computer and Internet Experience questionnaire (Czaja et al., 2006), and a Health Information-Seeking questionnaire.

In the one-on-one individual discussions with the participants, which were conducted by either the leader or one of the two facilitators, a brief overview of each aiding tool was provided along with printouts of the corresponding PP slides. The tools were presented one at a time, in the same order as they were presented in the PP demonstration. For each tool, participants were asked to respond to the following four questions:

1. Would you find the tool useful?

2. How do you think the tool could be helpful to you? Would you use the tool? If yes, why, and if not, why not?

3. Do you think this tool would be easy to use? If not, why not?

4. Do you think it would be hard to learn to use this tool or that it would make searching/finding health information more difficult? Do you think you would forget how to use this tool?

Following this discussion, the individual interviews concluded by asking the participants what other types of tools or aids they thought would be useful to them when searching the Internet for health information.

During these discussions the interviewer took notes, in addition to audio-taping the discussion. Following the integration of the comments of the participants, the leader presented a summary of the comments to the entire group of participants as a basis for the final group discussion, which lasted about one-half hour.

**Figure 1. The Mr. Taggy Search Browser Tool.**

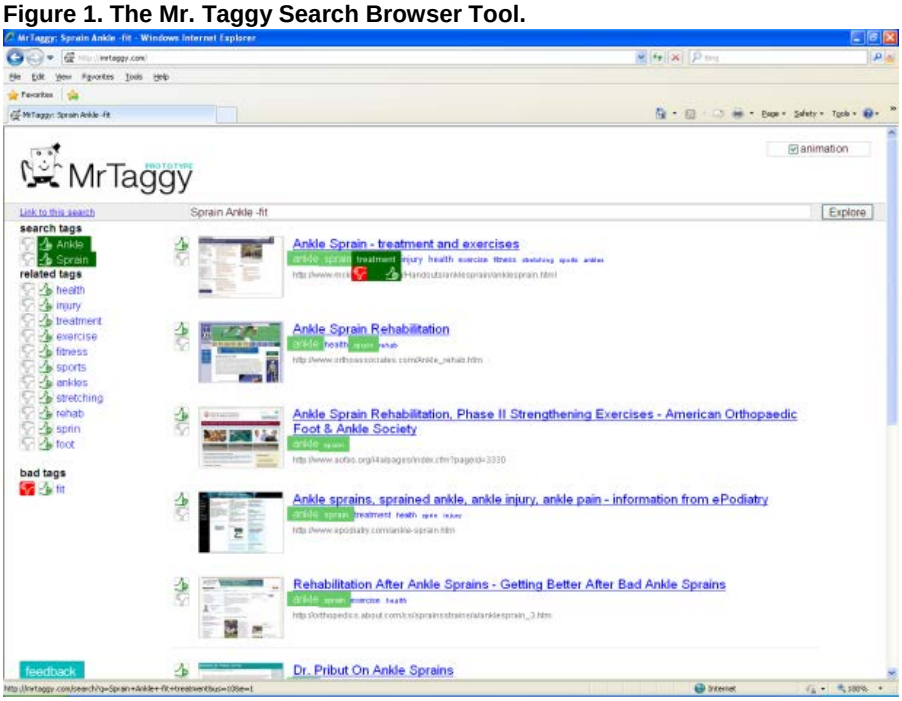
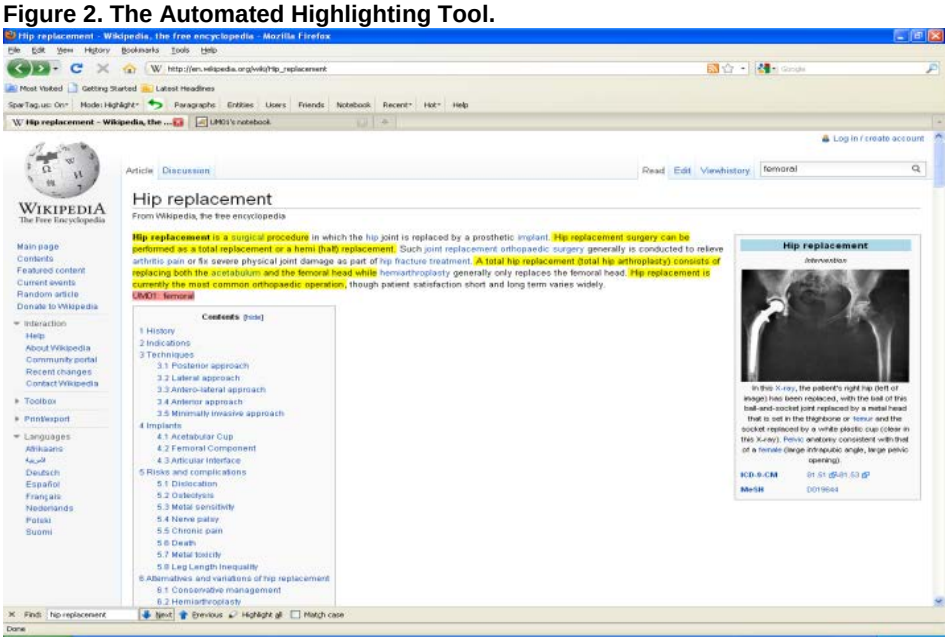


**Figure 2. The Automated Highlighting Tool.**




# Results

## Questionnaire Responses

The results from the Computer and Internet Experience questionnaire indicated that 16 of the 23 participants used a computer as frequently as several times a day; 18 of the participants used the Internet as frequently as several times a week; and that all 23 participants had access to the Internet in their homes. The Internet was used most frequently (i.e., several times a week) for communicating with family or friends ($n$ = 16), getting news ($n$ = 16), and for finding information about community events or resources ($n$ = 10).

The Internet was used more rarely (i.e., less than a few times per month) for activities such as looking for new people to meet ($n$ = 21), ), using an online social networking site ($n$ = 19), looking for religious or spiritual information ($n$ = 19), searching for information about employment ($n$ = 18), searching for "how-to-do-it" information ($n$ = 14), learning something new such as a language ($n$ = 13), making a travel reservation ($n$ = 13), bill paying ($n$ = 12), visiting a local, state, or federal government website (n = 10), and playing games and pursuing hobbies ($n$ = 10). With respect to using the Internet to look for advice about healthcare, 4, 15, and 4 of the participants said they used the Internet for this purpose several times a week, a few times a month, and rarely, respectively.

The majority of the participants indicated that they felt confident scrolling around a web page ($n$ = 21), using a search engine ($n$ = 21), selecting the right words for an Internet search ($n$ = 18), and getting to desired web pages using links ($n$ = 17). However, only three of the 23 participants agreed that they felt confident participating in online chats or discussions and using social networking sites such as Facebook or Twitter.

From the Health-Information Seeking questionnaire, of 11 sources of health information considered (e.g., popular books, friends or family, newspapers), the only two sources that the

participants indicated that they used most of the time or always were the Internet (*n* = 13) and their doctors or other providers (*n* = 14). Most importantly, in response to the question "In general, how difficult is it for you to find the health information that you need," only one participant indicated that it was not or just a little difficult; five participants indicated it was moderately difficult and 17 participants indicated that it was quite or extremely difficult.

When using the Internet as a source of health information, the participants' searches encompassed a wide variety of topics (e.g., nutrition, *n* = 18; illnesses or medical conditions, *n* = 16; exercise, *n* = 15; information about a doctor, hospital, or other health care provider, *n* = 15; news about health policy issues, *n* = 11; and health or medical products, *n* = 11). Most participants felt that the best virtues of Internet health information were that it could be obtained quickly (*n* = 20) and from numerous sources (*n* =20), and that the information made them more informed at doctor visits (*n* = 20), improved communication with their doctors (*n* = 15), and increased their ability to take better care of themselves (*n* =19).

## Focus Group Responses

The following preliminary data derive from the transcriptions of the group and individual discussions. The initial focus group discussion revealed a number of problems participants encountered that were particularly related to searching and making sense of health information. For example, one participant noted "I was researching back a while ago on shingles and if you research that you get all kinds of ambivalent information which doesn't necessarily equal with each other." Another participant emphasized the sheer volume of information that may need to be negotiated: "Sometimes when you're on the computer you get too much information before you get to where you want to go…for one thing, it may actually take 5 minutes. Two hours later you're still looking for it. That's the only thing bad about the Internet, there's so much information there. How much time do you have to look at it? You don't have too much time."

A summary of the results from the participants in response to the questions posed to them in the individual discussion portion of the focus group is presented in Table 1. Collectively, these data indicate that the participants were enthusiastic about the potential these tools offered in support of health-information seeking.

**Table 1. Summary of participants' responses to the questions posed in the individual discussions (*N* = 23)**

| **Individual Discussion Questions** | | M | S | A | G |
|---|---|---|---|---|---|
| Would you find the tool useful? | Yes | 18 | 17 | 18 | 22 |
| Would you find the tool useful? | No | 3 | 6 | 5 | 1 |
| Would you find the tool useful? | Not Sure | 2 | 0 | 0 | 0 |
| Would you use this tool? | Yes | 19 | 16 | 18 | 22 |
| Would you use this tool? | No | 4 | 7 | 4 | 0 |
| Would you use this tool? | Not Sure | 0 | 0 | 0 | 1 |
| Do you think this tool would be easy to use? | Yes | 19 | 20 | 19 | 23 |
| Do you think this tool would be easy to use? | No | 3 | 2 | 3 | 0 |
| Do you think this tool would be easy to use? | Not Sure | 1 | 1 | 1 | 0 |

| **Individual Discussion Questions** | | M | S | A | G |
|---|---|---|---|---|---|
| Do you think it would be hard to learn to use this tool? | Yes | 1 | 1 | 0 | 1 |
| Do you think it would be hard to learn to use this tool? | No | 19 | 22 | 23 | 22 |
| Do you think it would be hard to learn to use this tool? | Not Sure | 2 | 0 | 0 | 0 |
| Do you think this tool would make searching/finding health information more difficult? | Yes | 1 | 0 | 1 | 0 |
| Do you think this tool would make searching/finding health information more difficult? | No | 21 | 22 | 22 | 23 |
| Do you think this tool would make searching/finding health information more difficult? | Not Sure | 1 | 1 | 0 | 0 |
| Are you concerned that you would forget to use this tool or how to use it? | Yes | 2 | 4 | 2 | 1 |
| Are you concerned that you would forget to use this tool or how to use it? | No | 21 | 19 | 21 | 22 |
| Are you concerned that you would forget to use this tool or how to use it? | Not Sure | 0 | 0 | 0 | 0 |

M = Mr. Taggy Search Engine; S = SparTag.us Notebook; A = Automated Highlighting; G = Pop-up Glossary

However, their comments during the individual discussions also provided insights into some of the limitations, in addition to benefits, associated with the current versions of these tools, which pointed to initial design intervention considerations. For example, with regard to Mr. Taggy, on the positive side participants noted: "It would give you a variety of options of what to zero in on what you are looking for," "It's helpful because of the thumbs up and thumbs down mechanism," "It's more specific; subject can be found faster than Yahoo or Google," and "It gives you words (tags)—that I liked." In contrast, examples of negative comments directed at this tool were: "It's not helpful at all. Found tags confusing; why would you want all those tags? Prefer Google to Mr. Taggy," It doesn't seem helpful…any different than what I'm already doing," It's too involved; I want something simpler," "It's clunky; it's not very user friendly. It makes information seeking more difficult because of the way it goes about it," and "It's too cute, the cartoon element, I don't like it."

In response to the SparTag.us tool, examples of positive comments included: "It would enable me to integrate information," "You can grab all the information at once and then go back leisurely to read it," "I look at a dozen sites. Often, there are a few items of interest in each website (and) the idea of putting that all into one place is good," and "It organizes all the information I'm interested in." Some of the negative comments were: "I have never needed to save information in this way," "I prefer printing information," and "I don't quite understand how to get started using it."

For the automated highlighting tool, the positive remarks were not surprising, such as: "It would make searches easier," and "Saves you from reading every word in the article." However,

the negative comments pointed to concerns for overload and the need for training: “I may wind up with a page with a lot of yellow highlighting in it,” “It might highlight a lot of superfluous information,” and “I would need specific instructions (on how to use it).” For the Pop-up Glossary tool, advantages pointed out were: “It would be easier (to use) than a dictionary,” and “It seems more updated.” One participant did note that “It might not put in enough information to bring up exactly what I want.”

Finally, the participants provided recommendations for features they wanted to see in Internet health information-seeking tools. These included: a feature that provided pronunciation (and where relevant, pictures or animations) along with definitions of medical terms; the provision of links to medications associated with searched medical conditions; a feature that indicates how credible the medical information is; a feature that enables ads to be distinguished from (“actual”) information; and a tool that returns possible diseases in response to user-entered symptoms of a condition.

Overall, the results from this Phase of the project provided very useful data on the potential usability and usefulness of the tools. In particular, these data and provided guidance for the redesign of these tools, prior to the large-scale and more formal usability testing with older adults.

## Task Analysis and Heuristic Evaluation

Hierarchical task analyses (HTA) was conducted on the use of the two main tools, Mr. Taggy and the Spartag.us Notebook for performing information-seeking tasks. The HTA revealed the nature of the functionalities associated with these tools and how these functionalities are used to meet the information-seeking objectives. Three investigators performed heuristic evaluations of the Mr. Taggy search engine website and the Spartag.us Notebook in accordance with established human-computer interface guidelines for adults and older adults to establish a set of criteria needed for improving these tools in accordance with human factors design guidelines for older adults. A clinician at the Center on Aging, who specializes in cognitive aging and memory disorders, also reviewed the tools. These evaluations were discussed by the investigators and ultimately and the results from the focus groups were combined and sent to PARC who then selected a sample of high and medium priority items for implementation. The new interfaces were delivered and then pilot tested. After a few iterations, any remaining issues were resolved, and the new interfaces were approved and accepted for the formal evaluation study.

Recommendations for the Mr. Taggy search engine were divided into issues that could be addressed and resolved by user interface modifications and other issues that could be addressed by appropriate training. User interface modification issues primarily focused on assuring adequate readability of text. For example, text that appeared in small typefaces was enlarged and text that was written in gray (and not black) was enhanced to increase contrast and legibility. Other issues included adding user interface elements to the search results page that would highlight particularly salient results for a given search request. Low contrast thumbs up and thumbs down buttons were modified and button size was increased to provide a larger target area suitable for older adults. In addition, the search engine was thoroughly tested and all known software bugs were identified and eliminated.

Recommendations for the Spartag.us Notebook were similarly divided. An example of a Notebook user interface modification included adding a text box below each clipping (each text entry) so that participants could include comments and insights and thus aid working memory.

Another issue concerned Notebook clippings which could all too easily be accidentally deleted by a user. Therefore a safety dialog box was created which would alert the user to the deletion and provide an opportunity to cancel the action. Again, any software "bugs" in the Spartag.us Notebook were identified and corrected.

# Phase 2—Formal Evaluation Study

## Purpose

### Overview of the Study Design

The primary purpose of the Phase 2 formal evaluation study was to examine the impact of the aiding tools on information-seeking and sensemaking performance. Based on the task analyses, user evaluations in Phase 1 and technical considerations we selected three tools for the evaluation: Mr. Taggy, Automated Glossary, and Spartagus notebook tool. Given the current constraints associated with the tools, in particular Mr. Taggy and the Notebook, separate task problems were developed specifically for each of the tools: the Diabetes Scenario for the Mr. Taggy Search Engine and the Automated Glossary and the Multiple Sclerosis (MS) Scenario for the Spartagus notebook tool and the Automated Glossary tool. A sample of 80 adults was included and randomized into one of four conditions: 1) Diabetes Scenario aided; 2) Diabetes Scenario unaided; 3) MS Scenario aided; and 4) MS Scenario aided (described below). There were twenty participants in each of the 4 conditions. Participants in all four conditions used the Internet to perform the task problems, which involved searching, integrating and comprehending health information on the Internet. Those in the two unaided conditions performed the same tasks as those in the aided condition, using a search engine, but without the availability (and corresponding training) on the aiding tools.

### Scope

**Sample.** Eighty adults, all English speaking, 35+ years of age with a mean age of 57.3 years (SD = 10.8) (range of 34 to 82 years) who had with computer and Internet experience completed Phase 2 of the study. Twenty-three participants were males and fifty-seven were females. The total sample included eighty-nine adults. However, one person was a "no-show", five people failed the Internet criterion task and 3 people withdrew from the study. A total of 105 adults were initially screened for the study. See Table 2 for Sample characteristics.

**Table 2. Sample Characteristics**

| Demographics | Diabetes Scenario Unaided | | Diabetes Scenario Aided | | MS Scenario Unaided | | MS Scenario Aided | |
|---|---|---|---|---|---|---|---|---|
| Number | 20 | | 20 | | 20 | | 20 | |
| Age (M, SD) | 58.25 | 10.62 | 55.75 | 10.16 | 59.95 | 12.35 | 55.40 | 10.07 |
| Gender, Male, n,% | 8 | *40* | 3 | *15* | 4 | *20* | 8 | *40* |
| Gender, Female, n,% | 12 | *60* | 17 | *85* | 16 | *80* | 12 | *60* |
| Ethnicity, n,% | | | | | | | | |
| Hispanic | 3 | *15* | 4 | *20* | 1 | *5* | 9 | *45* |
| Non-Hispanic White | 10 | *50* | 10 | *50* | 15 | *75* | 7 | *35* |
| Non-Hispanic Black | 6 | *30* | 5 | *25* | 4 | *20* | 4 | *20* |
| Non-Hispanic Other* | 1 | *5* | 1 | *5* | 0 | *0* | 0 | *0* |
| Education, n,% | | | | | | | | |
| High School or less | 4 | *20* | 4 | *20* | 4 | *20* | 3 | *15* |
| Some College | 9 | *45* | 9 | *45* | 9 | *45* | 8 | *40* |
| College Graduate/Postgraduate | 7 | *35* | 7 | *35* | 7 | *35* | 9 | *45* |
| Yearly Household Income n,% | | | | | | | | |
| Less than $20,000 | 8 | *40* | 11 | *55* | 6 | *30* | 7 | *35* |
| $20,000 to $49,999 | 7 | *35* | 8 | *40* | 7 | *35* | 8 | *40* |
| More than $49,999 | 4 | *20* | 0 | *0* | 4 | *20* | 4 | *20* |
| Unknown/Did Not Answer | 1 | *5* | 1 | *5* | 3 | *15* | 1 | *5* |
| Occupational Status n,% | | | | | | | | |
| Work Full Time | 0 | *0* | 2 | *10* | 1 | *5* | 1 | *5* |
| Work Part time | 1 | *5* | 2 | *10* | 7 | *35* | 3 | *15* |
| Seeking Employment, Laid off | 7 | *35* | 4 | *20* | 1 | *5* | 10 | *50* |
| Retired | 9 | *45* | 5 | *25* | 9 | *45* | 2 | *10* |
| Other | 3 | *15* | 7 | *35* | 2 | *10* | 4 | *20* |
| General Health Status n,% | | | | | | | | |
| Poor | 0 | *0* | 1 | *5* | 0 | *0* | 0 | *0* |
| Fair | 2 | *10* | 2 | *10* | 1 | *5* | 1 | *5* |
| Good | 9 | *45* | 9 | *45* | 11 | *55* | 7 | *35* |
| Very Good | 7 | *35* | 6 | *30* | 4 | *20* | 7 | *35* |
| Excellent | 1 | *5* | 2 | *10* | 4 | *20* | 4 | *20* |
| Other | 0 | *0* | 0 | *0* | 0 | *5* | 1 | *5* |
| Length of time using Internet n, % | | | | | | | | |
| Less than 6 months | 0 | *0* | 0 | *0* | 0 | *0* | 0 | *0* |
| Between 6 months and 1 year | 0 | *0* | 0 | *0* | 2 | *10* | 0 | *0* |
| More than 1 year, but less than 5 years | 2 | *10* | 3 | *15* | 3 | *15* | 5 | *25* |
| 5 years or more | 18 | *90* | 17 | *85* | 15 | *75* | 15 | *75* |
| Hours/week using Internet n, % | | | | | | | | |
| Less than 1 h | 1 | *5* | 1 | *5* | 0 | *0* | 0 | *0* |
| Between 1 h and 5 h | 7 | *35* | 5 | *25* | 7 | *35* | 4 | *20* |
| More than 5 h, but less than 10 h | 2 | *10* | 8 | *40* | 3 | *15* | 5 | *25* |
| 10 h or more | 10 | *50* | 6 | *30* | 10 | *50* | 11 | *55* |
| Computer Proficiency (M, SD)*** | 138.4 | *21.05* | 131.1 | *22.91* | 138.9 | *21.42* | 146.4 | *23.38* |
| Health Literacy (M, SD)** | 4.65 | *1.309* | 3.70 | *1.342* | 3.75 | *1.209* | 3.95 | *1.731* |
| Reported Diabetes lifetime/now n, % | 2 | *10* | 4 | *20* | 4 | *20* | 1 | *5* |
| Reported M.S. lifetime/now n, % | 0 | *0* | 0 | *0* | 0 | *0* | 0 | *0* |

*Includes: No primary group, Multi-racial, and Other, ** T-test showed a significant difference between Diabetes Scenario groups, T=2.267, p=.029, with unaided being higher that aided. No significant difference for MS Scenario.*** No significant differences between groups.

# Methods

## Settings and Equipment

The study was conducted in the laboratory space of the Center on Aging at the University of Miami Miller School of Medicine during the period beginning February 01, 2013 and ending September 30, 2013. Two separate rooms were used allowing participants to be run concurrently. Each room was equipped with a Lenovo ThinkCenter A58 (7522-P1U) computer with 2GB RAM memory and a 300 GB Hard Drive. Additionally, each computer was running Microsoft Windows 7 Professional (2009) operating system and contained the following software programs: Techsmith Morae Recorder (v. 3.2.2) for recording all activity and interactions on-screen, and two web browsers, Internet Explorer 9.0 and Mozilla Firefox 3.6.X.

## Measures

**Demographic and Background Questionnaire.** This questionnaire consisted of the Background questionnaire (containing 8 items related to educational level, marital status, income, etc.), Occupational Status questionnaire, Health Information questionnaire (5 items addressing various aspects of overall health, activities of daily living and health conditions), Computer questionnaire (15 items related to perceptions about computer use, learning about computers, etc.), Computer Experience questionnaire (4 items) and Internet Experience questionnaire (5 items) which address length of computer/Internet use, frequency of use, etc.), Computer Proficiency questionnaire (34 items related to ease of use of computers, features of computers, software and software features divided into 6 groups: Computer Basics, Printer, Communications, Internet, Calendar, and Entertainment), and finally a test of health literacy, The Newest Vital Sign (NVS) (Weiss et all, 2005) which is a 6 item instrument that tests participant's ability to read and process information contained in nutrition labels.

**Cognitive Battery.** The cognitive battery consisted of the following tests: California Verbal Learning Test (CVLT)-Immediate (Delis, Kramer, Kaplan, & Ober, 1987), Paper Folding Test (Ekstrom, French, Harman, & Derman, 1976), Nelson-Denny Reading Comprehension Test (Brown, Fischo, & Hanna, 1993), Digit Symbol Substitution Test (Wechsler, 1997), Digit Symbol Recall Test (Wechsler, 1997), Shipley Vocabulary Test (Shipley, W. C., 1986). and the Inference Test (Ekstrom, French, Harman, & Derman, 1976).

**Disease Knowledge Questionnaire (Pre-Task, Post-Task).** The Disease Knowledge Questionnaire contained 32 items in the form of declarative statements (e.g., "31. Polyunsaturated fats are healthier for the heart than saturated fats.") dealing with a variety of medical issues including diseases such as Multiple Sclerosis, Diabetes, and Heart Disease and other issues and symptoms including High Blood Pressure, Cholesterol, and general diet and nutrition questions. Participants responded to these statements by checking Yes, No, or I Don't Know. The questionnaire was prepared after an exhaustive search was conducted for existing questionnaires dealing with the same topics. Relevant items were extracted from these questionnaires that were thought to serve the aims of the study. Additional questions were also

prepared and added to the questionnaire. Also, questions were intermixed to prevent participants from guessing the subject of the task they might be asked to complete.

**Sensemaking Interview.** The Sensemaking Interview was a semi-structured interview specifically tailored for each of the conditions, ranging from 5 to 12 questions, regarding the task problems and focused on how they proceeded through the problem, their satisfaction with the task and their work, and their conclusion about the medical status of the patient featured in the story (i.e., whether the patient had Diabetes or Multiple Sclerosis) that was at the core of each story task. Aided condition interviews included additional questions dealing with the tools and its features. The interview was recorded on a hand-held recording device.

**General Sensemaking Process Questionnaire.** The General Sensemaking Process Questionnaire contained 13 items in the form of declarative statements (e.g., "5. It was difficult for me to find information concerning Daniel's diet and lifestyle on the Internet."), which required participants to indicate their degree of agreement on a five point Likert scale (i.e., Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree). These items were self-appraisals of various information-seeking skills, performance on the task, and satisfaction with their information-seeking and sensemaking efforts. Two otherwise identical versions of this instrument were composed; one for the Diabetes Scenario, where the phrase "Daniels's Story" appears and another for the Multiple Sclerosis scenario, where the phrase "Jennifer's Story" appears.

**Usability Questionnaires.** Three different Usability Questionnaires were prepared for this study. Unaided condition participants were asked to rate 2 items, ease of use and helpfulness, of the Google search engine they used to conduct their searches on a five point Likert scale (i.e., Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree).

Diabetes Scenario aided participants received a questionnaire asking them to rate 6 elements or features of the Mr. Taggy search engine (i.e., the home page, tag clouds, the search results page, thumbs up and thumbs down buttons, tag sidebar) and then to provide overall ratings of the search engine itself, the find tool and the automated glossary. Participants rated these on usefulness, learnability, helpfulness and for some of the items, readability. This questionnaire contained 29 items in total.

MS Scenario aided participants received a questionnaire asking them to rate 4 elements or features of the Spartag.us Notebook (i.e., the tag tool, note box, notebook search tool, and notebook tag area) and then to provide overall ratings of the notebook itself, the find tool and the automated glossary. Participants rated these on usefulness, learnability, helpfulness and for some of the items, readability and organization. This questionnaire contained 25 items in total.

**Usability Interview.** The Usability Interview was a semi-structured interview conducted with participants in both aided conditions and consisted of questions regarding aspects, features, interface elements and screens of the tools they were asked to work with. For each tool, they were asked whether they found it useful, if there was anything they did not like about the tool, and also any features of the tool they would change or any features they would like to add to the tool that were not already there. They were also asked whether they had received enough training and whether they would use the tools if they were commercially available. The Diabetes

Scenario interview contained 15 items while the MS Scenario interview contained 17 items. The interview was also recorded on a hand-held recording device.

**Procedure.** Participants received an overview of the study and read and signed, the Informed Consent. An audio consent form was also read and signed for those who could not understand the written form. The study was approved by the Institutional Review Board of the University of Miami.

Participants then completed the Demographic and Background Questionnaire, Computer Experience and Internet Experience Questionnaires, Computer Proficiency Questionnaire and the Newest Vital Sign (NVS) health literacy questionnaire.

Following a brief break, a cognitive battery was administered, consisting of the California Verbal Learning Test-Immediate, Paper Folding Test, Nelson-Denny Reading Comprehension Test, Digit Symbol Substitution Test, Digit Symbol Recall Test , Shipley Vocabulary Test and the Inference Test.

Upon completion of the cognitive battery, participants performed two short Internet exercises. If they completed the exercises correctly they were able to proceed to the second part of the study and had a 1-hour lunch break. If they did not qualify, they were dismissed from the study and given $40.00 plus a parking voucher card.

Qualified participants were then randomized into one of the four study groups: Diabetes Scenario aided or unaided, MS Scenario, aided or unaided. Participant in the Diabetes Scenario unaided condition used the Google search engine to complete the Daniel Story task, while those in the aided condition used a suite of tools consisting of the Mr. Taggy search engine, the Mozilla browser find tool and the automated glossary tool to work on the same task. In a similar manner, participants in the MS Scenario unaided condition used the Google search engine to complete the task while participants in the aided condition used the Spartag.us Notebook, the Mozilla browser find tool and the automated glossary tool to complete the same task.

Participants in all four conditions received training by a trained interventionist corresponding to their study condition. Each training module was scripted and included examples and criterion tasks. Each participant had to complete each criterion task without any help whatsoever from the experimenter. A description of each training module is provided below.

Following training, participants were asked to complete the Disease Knowledge Questionnaire. They then proceeded to the task problem and had 45 minutes (Diabetes Scenario) or one hour (MS Scenario) to complete the problem on their own. They then completed the 1) Sensemaking Interview, 2) General Sensemaking Process Questionnaire, 3) Disease Knowledge Questionnaire (post-task), and 4) Usability Questionnaire. Those in the aided condition also completed the Usability Interview. All participants participated on an individual basis and were compensated $80.00 plus parking for their participation.

## Training

**Diabetes Scenario and MS Scenario Unaided Condition Training.** Training for the two unaided conditions were identical and consisted of providing participants with a very brief refresher of basic computer/window and Internet skills. This training consisted of pointing out essential features of the web browser such as the back button, forward button and address bar, among other items. Participants were also instructed on the basics of using tabs and were then asked to practice the skill on their own by creating a new tab and then closing the tab. Although

all participants had Internet experience (a basic level of competence was proved by the criterion Internet tasks administered in the morning session), some participants were not familiar, or not very familiar, with using tabs (either creating them or closing them).

Next, participants were shown the basics of conducting a Google search and were then asked to conduct a search on their own. Finally, they were then shown the different parts of the Google search results page. This concluded both unaided condition trainings, after which they spent 30 minutes using the Internet, as described above. These elements of training were incorporated into the more extensive Diabetes Scenario and MS Scenario aided training modules. Following this brief refresher training participants were instructed to spend exactly 30 minutes using the Internet. During this time, they could check email, conduct personal searches for information, play games, read news items and so on.

**Diabetes Scenario Aided Condition Training.** The Diabetes Scenario aided condition was developed in order to examine the use of the Mr. Taggy search engine. To this end, training was designed to enable participants to be reasonably proficient in the use of four software tools: 1) the Firefox web browser; 2) the automated glossary; 3) the Mozilla Firefox web browser Find Tool; and 4) the Mr. Taggy search engine. Both aided training modules were developed according to recommendations in the 4C/ID Instructional Model (van Merrienboer, Clark, de Croock, 2002). The model stipulates the incorporation of learning tasks and part-task practice tasks, in addition to the use of supportive and just-in-time information to create effective training.

Training began with a brief overview and then the essentials of the Mozilla Firefox web browser. For the purposes of this study, the Firefox web browser was considered to be functionally identical to the Internet Explorer web browser and training focused on identifying and locating essential navigation buttons including the back button, forward button, address bar, and tabs.

The Find tool was then accessed and described. Later, the participant practiced the use of the Find tool on a webpage with special emphasis on several features of the tool including the Highlight All button which highlights all of the instances of a word on a web page.

The Mr. Taggy search engine was accessed and a search was conducted. From the resulting search results page, a link was identified and clicked on.

Next, the Automated Glossary tool was demonstrated and the participant accessed one the many a pop-up definitions available on the system.

A unique feature of the Mr. Taggy search engine home page is the use of "tag clouds" to narrow down and execute a keyword search and the participant had an opportunity to practice its use. On the following search results page, another unique feature—thumbs up and thumbs down buttons—were demonstrated and their use in refining and narrowing down search results was explained. The participant was then asked to perform several tasks using these buttons. A practice exercise was carried out which enabled the participant to rehearse all the essential skills needed to successfully conduct a search using the Mr. Taggy search engine. Afterwards a criterion exercise was given to demonstrate the participant's proficiency using the search engine. Upon successful completion of the criterion exercise (1A), the participant continued on to the Diabetes Scenario task. If the participant failed the criterion task, they proceeded to review the material and participate in more practice tasks after which they attempted a second criterion exercise (1B). If the participant failed this second criterion exercise, they were excluded from the study.

**MS Scenario Aided Condition Training.** The MS Scenario aided condition training allowed participants to become proficient in the use of the Spartagus Notebook tool to collect information from webpages, classify that information and then be able to retrieve it as needed. The Notebook is a unique software tool designed to allow the user to collect information from webpages in an effortless manner by "clipping" information from a webpage (i.e., pointing to the desired information and then clicking on it) and storing it in an online notebook In addition, tags are assigned to the information, allowing for more convenient retrieval once it is saved within the notebook. Notebook training was the most extensive and complex of all the training modules. Broadly, it consisted of the following parts. First, a brief introduction to saving clippings was given followed by training on the use of the Automated Glossary and the Mozilla Firefox web browser Find tool. Next, participants completed a more extensive set of clipping exercises, each focused on a particular aspect of the clipping task, culminating in a criterion task. If the participant completed the criterion task (1), they moved on to the Notebook phase of training. If not, they were given a refresher module, upon completion of which they would have a second criterion clipping task (1A) to perform.

The next training section dealt with the using the Spartagus Notebook which consisted of identifying all of the major elements of the Notebook interface such as the tag area, note box feature and the Notebook search tool (which is used exclusively to search for clippings within the Notebook). The participant was given a variety of exercises in using the Notebook search tool.

The participant then completed an extensive practice exercise involving finding and saving information on various over-the-counter (OTC) headache medications. This practice exercise provided ample opportunity for the participant to use the Notebook to repeatedly practice clipping skills. Participants were encouraged to make use of the Note Box feature to write personal insights that might help them make a more informed decision regarding OTC medications at the end of the practice exercise.

Finally, participants completed a very similar criterion task (2) involving OTC allergy medications. If they were successful, they moved on to the MS Scenario task. If not, they continued on to a review with practice tasks after which they completed a third criterion task (3). If they succeeded, they would move on to the MS Scenario task; if not, they were dismissed from further participation in the study. Participants had to demonstrate basic proficiency in clipping information from webpages and using the Notebook.

## Tasks

**Diabetes Scenario Task.** The Diabetes Scenario condition task was developed in order to examine participants' use of the Mr. Taggy search engine, along with the Mozilla Firefox find tool and the automated glossary. The development of this story took into account the current constraints associated with Mr. Taggy, specifically, the fact that it was based on information "tagged" by a large community of internet users but still limited in scope to search engines such as Google. Thus, it was imperative that a scenario be developed that required the collection of information that would be accessible or "reachable" by the Mr. Taggy tool.

A story was developed where the central character—Daniel—was introduced along with a description of his lifestyle, eating habits and family history of diabetes. After reading the story, the participant was asked to explore each aspect of Daniel's lifestyle and to decide whether or

not he was at risk of developing Diabetes. The participant was then asked to write a summary that explained this risk in the context of his (Daniel's) family history and lifestyle and to provide specific steps that he should take with regard to eating sugar, pasta, exercise, taking vitamin and mineral supplements, etc. Participants in the aided condition had to use the Mr. Taggy search engine to find information that would allow them to assess the risk and detail any steps he needed to take in case he was at risk of becoming a Diabetic. Participants had 45 minutes to complete this part of the task and all computer screen activity was captured using the Techsmith Morae Recorder.

Participants then had to complete 3 more questions using the Mr. Taggy search engine that dealt with blood glucose levels and the Glycemic Index (GI). For this part, participants were simply asked to work as quickly as possible with no specific time limit, although they were expected to work no more than 30 minutes. Participants recorded their responses in a response booklet.

**MS Scenario Task.** In a similar manner, the MS Scenario aided condition task was developed in order to examine participants' use of the Spartag.us Notebook tool, along with the Mozilla Firefox find tool and the automated glossary. This scenario was much more complex than the Diabetes scenario problem as participants would have access to Google to search the entire web for information, and thus were not restricted in the scope of information that could be collected.

In this scenario, the central character was a woman named Jennifer who lived in Alaska that might or might not have Multiple Sclerosis (MS). Participants were asked to cast themselves in the role of a doctor's special medical assistant charged with reading Jennifer's medical history, in narrative form, that provided a description of her current state of health, mother's medical history, and a chronology of medical and life events dating back to 2006 through to the present day. The chronology detailed a series of attacks of weakness and numbness to her extremities, a rock climbing accident, a blood test and her employment history. Instructions stressed the two main participant goals, namely 1) to copy all relevant information from the Internet to the Spartag.us Notebook using Google as their search engine and 2) writing a complete detailed summary using information they copied to the notebook. Participants were also asked to write a summary examining all of the relevant parts of the medical history, how they were related to Jennifer's current medical state, and provide a diagnosis, to whatever degree they were capable of, of her medical condition, i.e., whether she had MS or not or what other event, exposure, etc., could explain her current medical condition. This written summary would allow the fictional doctor to quickly assess facts and arrive at a professional medical conclusion as to Jennifer's condition. Participants had 60 minutes to complete this part of the task. As in the Diabetes Scenario, all computer screen activity was captured using the Techsmith Morae Recorder.

Before moving to the additional questions, participants were asked to answer the following question: "Do you think there is anything wrong with Jennifer?" to which they could respond "Yes", "No" or "I Don't Know". They also had to explain this answer. In addition, if they answered "Yes," they had to complete the sentence "Jennifer is suffering from…." and if the answered "No," they had to complete the sentence "Jennifer is currently experiencing weakness and numbness in her arms and legs because…." These questions were included in order to be certain that participants came to a conclusion regarding Jennifer's state of health, summarize an explanation as to the conclusion and also be explicit as to the disease or the explanation.

Participants also had to complete additional 4 questions dealing in direct manner with issues presented in the MS Scenario narrative. Again, for this part (3), participants were simply asked to work as quickly as possible with no specific time limit.

Given the more extensive and complicated nature of the Jennifer Story narrative and task, the response booklet was divided into three parts, enabling the participant to easily consult the story and write the summary without having to flip through pages back and forth.

# Results

## Cognitive Tests

Means and standard deviations of all cognitive measures by scenario and aided and unaided condition are listed in Table 3. There were no significant differences in any of the cognitive measures between the unaided and aided groups in either scenario. Correlations between task performance scores and cognitive measures were calculated using Spearman's correlation coefficient and are presented in the Table 4.

**Table 3. Means and standard deviations of cognitive measures and task performance scores.**

| Cognitive Measures* | Diabetes Scenario Unaided | | Diabetes Scenario Aided | | Multiple Sclerosis Scenario Unaided | | Multiple Sclerosis Scenario Aided | |
|---|---|---|---|---|---|---|---|---|
| CVLT **(M, SD)** | 27.15 | **5.659** | 26.55 | **6.270** | 28.40 | **5.548** | 26.65 | **6.037** |
| Paper Folding **(M, SD)** | 6.55 | **2.743** | 6.60 | **3.409** | 6.05 | **2.724** | 7.25 | **3.110** |
| Reading Comprehension **(M, SD)** | 21.30 | **9.985** | 18.60 | **7.612** | 23.35 | **7.734** | 21.20 | **10.17** |
| Digit Symbol **(M, SD)** | 54.95 | **10.15** | 53.40 | **9.405** | 53.65 | **7.576** | 56.75 | **13.05** |
| Digit Symbol Recall **(M, SD)** | 5.20 | **2.546** | 5.25 | **2.149** | 5.55 | **2.164** | 5.25 | **2.221** |
| Shipley **(M, SD)** | 31.75 | **3.370** | 29.85 | **5.294** | 32.25 | **3.959** | 30.95 | **5.530** |
| Inference Test **(M, SD)** | 10.00 | **4.155** | 8.75 | **3.323** | 10.05 | **4.058** | 10.10 | **4.833** |
| Task Performance Scores **(M, SD)*** | 8.10 | **2.453** | 5.850 | **3.265** | 11.63 | **5.650** | 7.050 | **6.477** |

*CVLT measures memory span. Paper Folding measures Spatial/Visualization Ability. Reading Comprehension and Shipley measure Verbal Ability. Digit Symbol measures Perceptual Speed. Digit Symbol Recall measures Working/Incidental Memory. Inference Test measures Reasoning and Inductive Ability** Diabetes Scenario Score Range (0-12), Multiple Sclerosis Scenario Score Range (0-21).

**Table 4. Correlations of various cognitive measures with task performance scores.**

| | Diabetes Scenario (Mr. Taggy Search Engine) | | | Multiple Sclerosis Scenario (Spartag.us Notebook) | | |
|---|---|---|---|---|---|---|
| Cognitive Test | Aided (n=20) | Unaided (n=20) | Overall (n=40) | Aided (n=20) | Unaided (n=20) | Overall (n=40) |
| CVLT - Immediate | .632*** | .195 | .403*** | .384* | .216 | .363** |
| Paper Folding | .180 | .265 | .185 | .384* | .057 | .099 |
| Reading Comprehension | .454** | .374 | .444*** | .482** | .046 | .323** |
| Digit Symbol Substitution | .333 | .462** | .427*** | .508** | .228 | .360** |
| Digit Symbol Recall | -.251 | .102 | -.038 | -.226 | -.115 | -.073 |
| Shipley Vocabulary | .701*** | .230 | .516*** | .458** | .189 | .344** |
| Inference Test | .499** | .328 | .462*** | .653*** | .062 | .387** |

Note: Significance levels are 2-tailed, *$p$ < = .1, **$p$ < = .05, ***$p$ < = .01.

As shown in Table 4 cognitive abilities were related to overall performance for both tasks, however this pattern changed when examining the relationships separately for those in the aided vs. unaided conditions suggesting that use of the aiding tools increased the cognitive demands of the tasks.

## Disease Knowledge Questionnaire (DKQ)

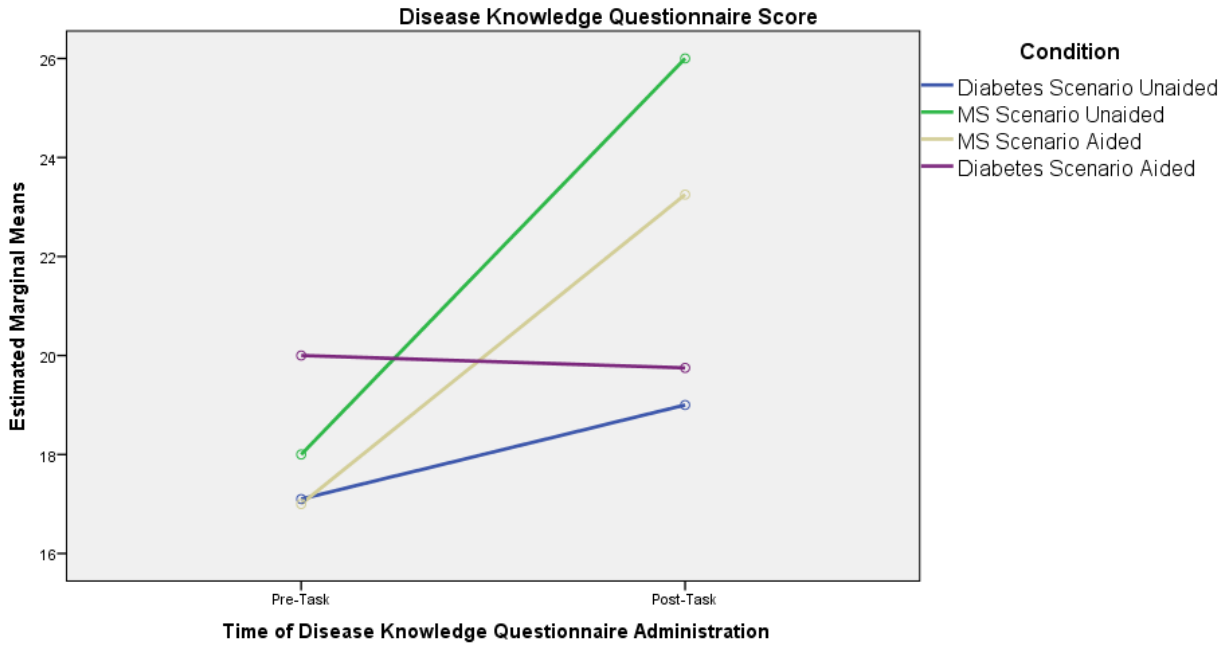
A two-way mixed ANOVA was conducted on the DKQ Score for both scenarios (Diabetes and MS) with time (Pre and Post) as the within-subjects variable and condition (Aided and Unaided) as the between-subjects variable.

In the Diabetes Scenario, there was no significant effect of time on the DKQ score, $F(1, 38)= 1.896$, p=.177. There was also no significant effect of Condition on the DKQ Score, $F(1. 38)=1.040$, p=.314. There was a no interaction effect of time and condition, $F(1, 38)=3.220$, p=.081 (Figure 3).

In the MS Scenario, there was a highly significant effect of time on the DKQ score, $F(1, 38)= 199.0$, p=.000. The DKQ score was higher when given after the task (Post) (M=24.63) than when given before the task (Pre) (M=17.50). There was no significant effect of Condition on the DKQ Score, $F(1, 38)=2.158$, p=.150. There was also no interaction effect of time and condition, $F(1,38)=3.001$, p=.091.

In order to examine further any differences in knowledge, paired t-tests were performed. Over the entire sample (n=80), a strongly significant difference occurred between pre and post Disease Knowledge Questionnaire scores, $t(79) = -7.44$, $p = .000$, with post-task scores being higher, $M = 22.00$ *(SD* = 5.66*)*, than pre-task scores, $M = 18.03$ *(SD* = 5.45*)*.

**Figure 3. Pre-Task and Post-Task Disease Knowledge Questionnaire Scores by Condition.**




For participants in the Diabetes Scenario condition (n=40), no significant difference was observed between pre and post Disease Knowledge Questionnaire scores, $t(39) = -1.339$, $p = .188$, with post-task scores being higher, $M = 19.38$ ($SD = 5.83$), than pre-task scores, $M = 18.55$ ($SD = 6.14$).

Participants in the Diabetes Scenario aided condition (those using the Mr. Taggy search engine (n = 20), no significant difference was obtained, $t(19) = .340$, $p = .738$, with post-task scores being fairly similar, $M = 19.75$ ($SD = 6.17$), to pre-task scores, $M = 20.0$ ($SD = 6.00$).

For participants in the Diabetes Scenario unaided condition (those using the Google search engine (n = 20), a significant difference (to the p = 0.1 level) was observed between pre and post Disease Knowledge Questionnaire scores, $t(19) = -2.01$, $p = .059$, with post-task scores being higher, $M = 19.00$ ($SD = 5.61$), than pre-task scores, $M = 17.10$ ($SD = 6.07$).

Overall, for the MS Scenario participants (n = 40) there was a strongly significant difference between pre and post Disease Knowledge Questionnaire scores, $t(39) = -13.758$, $p = .000$, with post-task scores being higher, $M = 24.63$ ($SD = 4.12$), than pre-task scores, $M = 17.50$ (SD = 4.70).

Participants in the MS Scenario aided condition (those using the Google search engine with the Notebook) (n = 20), a strongly significant difference was obtained, $t(19) = -8.038$, $p = .000$, with post-task scores being higher, $M = 23.25$ ($SD = 4.898$), than pre-task scores, $M = 17.0$ ($SD = 5.68$).

For participants in the MS Scenario unaided condition (those using only Google and pen and paper) (n = 20), a strongly significant difference was observed between pre and post Disease Knowledge Questionnaire scores, $t(19) = -12.41$, $p = .000$, with post-task scores being higher, $M = 26.0$ ($SD = 2.58$), than pre-task scores, $M = 18.0$ ($SD = 3.54$).

## Task

To examine differences in information-seeking and sensemaking performance between participants in the Diabetes Scenario condition that used the Mr. Taggy search engine (aided) versus those who used the Google search engine (unaided), a t-test was performed to determine if there was any significant difference.

For Part B questions, a significant difference occurred, $t(38) = 2.46$, $p = .018$, with the unaided group performing significantly better, $M = 8.10$ ($SD = 2.45$), than the aided group, $M = 5.850$ ($SD = 3.265$).

To examine differences in information-seeking and sensemaking performance between participants in the MS Scenario condition that used the Spartag.us Notebook (aided) versus those who used the Google search engine (unaided), a t-test was performed to determine if there was any significant difference.

For Part 3 questions, a significant difference occurred, $t(38) = -2.38$, $p = .022$, with the unaided group (participants not using the Spartag.us Notebook, just pen and paper) performing significantly better, $M = 11.63$ ($SD = 5.65$), than the aided group (participants using the Spartag.us Notebook), $M = 7.05$ ($SD = 6.48$). This somewhat surprising result can be partially explained by an unusually high number of poor scores for the MS Scenario aided condition (i.e., 8 cases in 20—40%—where the score was either 0.0 (1 case), 0.5 (1 case), 1.0 (5 cases), 2.0 (1 case) out of a possible 21 points),

## General Sensemaking Process Questionnaire (GSPQ)

In order to examine perceived information-seeking and sensemaking post-task between the unaided and aided conditions, t-tests were performed to determine if there were any significant differences.

For Diabetes Scenario participants (those testing the Mr. Taggy search engine (n = 40), a significant difference was observed between aided and unaided GSPQ scores, $t(38) = -2.06$, $p = .046$, with aided GSPQ scores being higher, $M = 29.50$ ($SD = 9.01$), than unaided GSPQ scores, $M = 24.55$ ($SD = 5.82$).

For MS Scenario participants (those testing the Spartag.us Notebook search engine (n = 40), no difference was observed between unaided and aided GSPQ scores, $t(38) = .683$, $p = .499$.

## Usability Questionnaires

**Google Usability.** Unaided condition participants (n = 40) in both Diabetes Scenario and MS Scenario groups were given a 2-item Google Usability Questionnaire to fill out. A t-test was performed to determine if significant differences were present. No significant difference was observed, $t(38) = .319$, $p = .752$, with Diabetes Scenario unaided scores, $M = 9.20$ ($SD = 1.01$), being virtually identical to MS Scenario unaided scores, $M = 9.30$ ($SD = .979$). Twenty-five, or 62.5% of respondents, reported that they strongly agreed with the statements “Overall, the Google search engine was easy to use” and “Overall, the Google search engine helped me find the information I needed,” while fifteen, or 37.5% of respondents, reported agreement with the same statements.

**Mr. Taggy Search Engine Usability.** Participants in the Diabetes Scenario aided condition used the Mr. Taggy search engine as an aid and they completed the Mr. Taggy Search Engine Usability Questionnaire. For each feature, such as the home page, a visual cue was included on the same page as the questions so the participant understood exactly what they were being asked to rate. In addition, the experimenter went over the entire questionnaire prior to being filled out. See Table 5 for a complete breakdown of responses.

**Table 5. Usability Measures for the Mr. Taggy Search Engine**

| Mr. Taggy Search Engine Usability Measures | Strongly Agree/ Agree | | Neither Agree nor Disagree | | Disagree/ Strongly Disagree | |
|---|---|---|---|---|---|---|
| | n | *%* | n | *%* | n | *%* |
| **Home Page** | | | | | | |
| …. was easy to learn. | 17 | *85* | 0 | *0* | 3 | *15* |
| …. was hard to use. | 2 | *10* | 2 | *10* | 16 | *80* |
| …. helped me find the information I needed. | 15 | *75* | 5 | *25* | 0 | *0* |
| **Tag Clouds** | | | | | | |
| …. were easy to learn. | 17 | *85* | 1 | *5* | 2 | *10* |
| …. were hard to use. | 1 | *5* | 2 | *10* | 17 | *85* |
| …. helped me find the information I needed. | 13 | *65* | 4 | *20* | 3 | *15* |
| **Search Results Page** | | | | | | |
| …. was hard to learn. | 2 | *10* | 7 | *35* | 11 | *55* |
| …. was easy to use. | 15 | *75* | 3 | *15* | 2 | *10* |
| …. helped me find the information I needed. | 15 | *75* | 2 | *10* | 3 | *15* |
| Type size of the search results page was too small | 1 | *5* | 8 | *40* | 11 | *55* |
| …. was well organized. | 16 | *80* | 4 | *20* | 0 | *0* |
| **Thumbs Up / Thumbs Down Buttons** | | | | | | |
| …. were easy to learn. | 17 | *85* | 2 | *10* | 1 | *5* |
| …. was easy to use. | 17 | *85* | 2 | *10* | 1 | *5* |
| …. did not help me find the information I needed. | 4 | *20* | 7 | *35* | 9 | *45* |
| The sizes of the buttons were too small. | 0 | *0* | 10 | *50* | 10 | *50* |
| **Tag Sidebar** | | | | | | |
| …. was hard to learn. | 3 | *15* | 3 | *15* | 14 | *70* |
| …. was easy to use. | 16 | *80* | 2 | *10* | 2 | *10* |
| …. helped me find the information I needed. | 12 | *60* | 5 | *25* | 3 | *15* |
| **Search Engine, Overall** | | | | | | |
| …. was easy to learn. | 16 | *80* | 2 | *10* | 2 | *10* |
| …. was hard to use. | 3 | *15* | 3 | *15* | 14 | *70* |
| …. helped me find the information I needed. | 13 | *65* | 7 | *35* | 0 | *0* |
| **Find Tool** | | | | | | |
| …. was easy to learn. | 16 | *80* | 1 | *5* | 3 | *15* |
| …. was easy to use. | 15 | *75* | 2 | *10* | 3 | *15* |
| …. helped me find the information I needed. | 10 | *50* | 8 | *40* | 2 | *10* |
| ….features and results were hard to read. | 3 | *15* | 6 | *30* | 11 | *55* |
| **Automated Glossary Feature, Overall** | | | | | | |
| …. were easy to learn. | 16 | *80* | 3 | *15* | 1 | *5* |
| …. were easy to use. | 16 | *80* | 3 | *15* | 1 | *5* |
| …. helped me find the information I needed. | 9 | *45* | 11 | *55* | 0 | *0* |
| Definitions were hard to read. | 2 | *10* | 8 | *40* | 10 | *50* |

**Spartag.us Notebook Usability.** Participants in the Multiple Sclerosis Scenario aided condition used the Spartag.us Notebook as an aid and they completed the Notebook Usability Questionnaire. For each feature, such as the tag tool, a visual cue (see Figure 1 in the Appendix)

was included on the same page as the questions so the participant understood exactly what they were being asked to rate. The experimenter went over the entire questionnaire prior to being filled out. See Table 6 for a complete breakdown of responses.

**Table 6. Usability Measures for the Spartag.us Notebook**

| Spartagus Notebook Usability Measures | Strongly Agree/ Agree | | Neither Agree nor Disagree | | Disagree/ Strongly Disagree | |
|---|---|---|---|---|---|---|
| | n | *%* | n | *%* | n | *%* |
| **Tag Tool Feature** | | | | | | |
| …. was hard to learn. | 1 | *5* | 2 | *10* | 17 | *85* |
| …. was easy to use. | 19 | *95* | 0 | *0* | 1 | *5* |
| …. helped me collect and organize the info. I needed. | 19 | *95* | 0 | *0* | 1 | *5* |
| **Note Box Feature** | | | | | | |
| …. were easy to learn. | 20 | *100* | 0 | *0* | 0 | *0* |
| …. were easy to use. | 20 | *100* | 0 | *0* | 0 | *0* |
| …. helped me write notes for each clipping. | 18 | *90* | 2 | *10* | 0 | *0* |
| **Notebook Search Tool** | | | | | | |
| …. was hard to learn. | 2 | *10* | 0 | *0* | 18 | *90* |
| …. was easy to use. | 20 | *100* | 0 | *0* | 0 | *0* |
| …. helped me find clippings and info, in Notebook. | 18 | *90* | 2 | *0* | 0 | *0* |
| **Notebook Tag Area Feature** | | | | | | |
| …. was hard to learn. | 3 | *15* | 0 | *0* | 17 | *85* |
| …. was easy to use. | 19 | *95* | 0 | *0* | 1 | *5* |
| …. helped me find clippings and info, in Notebook. | 19 | *95* | 1 | *5* | 0 | *0* |
| **Notebook, Overall** | | | | | | |
| …. was easy to learn. | 19 | *95* | 1 | *5* | 0 | *0* |
| …. was easy to use. | 18 | *90* | 1 | *5* | 1 | *5* |
| …. helped me find the information I wanted. | 19 | *95* | 1 | *5* | 0 | *0* |
| …. was hard to read. | 0 | *0* | 1 | *5* | 19 | *95* |
| Notebook and features were well organized. | 18 | *90* | 1 | *5* | 1 | *5* |
| **Find Tool** | | | | | | |
| …. was easy to learn. | 18 | *90* | 2 | *10* | 0 | *0* |
| …. was easy to use. | 18 | *90* | 2 | *10* | 0 | *0* |
| …. helped me find the information I wanted. | 14 | *70* | 6 | *30* | 0 | *0* |
| ….features and results were hard to read. | 0 | *0* | 2 | *10* | 18 | *90* |
| **Automated Glossary Feature, Overall** | | | | | | |
| …. were easy to learn. | 19 | *95* | 1 | *5* | 0 | *0* |
| …. were easy to use. | 19 | *95* | 1 | *5* | 0 | *0* |
| …. helped me find the information I wanted. | 12 | *60* | 6 | *30* | 2 | *10* |
| Definitions were hard to read. | 0 | *0* | 5 | *25* | 15 | *75* |

## Discussion

This study evaluated the usability of software aiding in terms of their impact on the performance of health information seeking among a sample of older adults. Overall, as shown in tables 5 and 6, the participants perceived the tools as useful and indicated that they helped them with the search tasks examined. Those who used the Mr. Taggy tool also had significantly higher sense making process knowledge post-task performance. In addition, all participants irrespective of task condition also showed an increase in health knowledge following completion of the health information seeking performance tasks. This finding is encouraging as it suggests that

being engaged in Internet information-seeking activities may in fact be beneficial in terms of enhancing an individual's health knowledge.

However, the data also indicated that those who used the tools actually performed worse on the health information seeking problems. There are several possible reasons for this finding. First, the Spartag.us Notebook tool was fairly complex and required a fair amount of training and concentrated effort. Training was designed to make sure participants were able use the tool to successfully create and tag clippings made from online information and to manipulate these clippings in the Notebook using the search, note boxes, etc. Thus training time was long for many participants, which may have influenced their subsequent performance. Additionally, and perhaps most importantly, the participants were required to perform the tasks immediately following training on these tools. Consequently, they were most likely to be allocating a significant amount of cognitive effort in recalling the features of the tools and how to use them. With repeated use of these tools, as would be the case if these participants had the opportunity to use these tools in their natural environment such as their home, these users would able to allocate most of their cognitive effort to the problem and not to use of the tool and thereby potentially make more effective use of these tools. In addition, these tools require additional software modifications, beyond those we were able to incorporate in this demonstration project, in order to make them easier to use. Instead, we overcame some of these limitations through training, which added to the learning burden on the participants. A larger longitudinal study with such changes incorporated would provide a truer test of the added benefits of working with these tools and especially the Spartag.us Notebook tool. Finally, as explained above, this group may have had an unusually high number of poor performers (8 cases were scores were between 0 and 2 points compared to only 2 cases in the unaided group).

In spite of their poorer performance, this group gave the Spartag.us Notebook system very high ratings overall and across all of its features (Table 6), and indicated that use of the tool helped them find information. Anecdotally, participants were very enthusiastic about the tool and its features, inquiring whether it was available as a commercial product (it is not). It's worth pointing out that this group benefitted in one important way over the unaided group; namely, they were left with work product in the form of the computerized notebook, complete with information, text, links, tags and notes that could be referred to and reused to extend their knowledge about the particular issue being researched. Many participants stressed the importance of having their research, links, comments and notes available in this form. A couple of people commented that they did something similar using standard word processing software, like Microsoft Word, but that the Spartag.us Notebook was a far superior and elegant solution. As implied above, a longitudinal study would be able to test the longer term benefits that some of these participants alluded to with regard to follow-up use of the tool

Participants who used the Mr. Taggy search engine also performed less well on the tasks than their unaided (Google) counterparts. This is likely to be attributed to the same reasons noted above, specifically, the fact that these participants were still likely trying to get comfortable with this tool while at the same time performing the task following the training they received on use of this tool. However, participants were generally positive about the tool's organization and features, ease of use and ease of learning and indicated it helped them find the information they needed (Table 5). Future research with the Mr. Taggy interface should not only be investigated under conditions of longer term use and thus acclimation to this tool, but also with search engine capabilities (for example, that can rival today's major search engines) that can access a much larger array of information than it is currently capable of doing due to the constraint that it

presently relies on tags from users of a particular social media website. In fact, the reason those participants working in the Diabetes Scenario unaided condition had post-task DKQ scores that were higher than those in the aided group may be due to the superior breadth of websites available to users of the Google search engine as compared to those available to those who used the Mr. Taggy search engine.

In general, as shown in Table 4, cognitive abilities were strongly predictive of performance in the aided conditions and for the most part not predictive of performance in the unaided conditions. As emphasized above, the tools likely increased the cognitive burden on the participants in the aided conditions, especially considering that it was their first opportunity to use the tools for relatively complex problem solving and thus these tools were still relatively unfamiliar to them during task performance. Thus the performance of these participants was expected to be especially sensitive to cognitive abilities as in fact the data in Table 4 shows. However, this finding also suggests that the tools need to be designed so that they are less complex, especially for this user group, and also that people need more extensive practice on use of the tools.

As for overall the highly significant gains in DKQ scores, it's worth noting that the results suggest an improvement in scores by virtue of simply doing the tasks. In particular, the MS Scenario narrative contained many items that needed to be explored, searched and processed and engaging in the task seemed to affect both aided and unaided groups in a similar positive manner. This suggests that further research should explore the process of search, in other words the setup and execution of search, which could be incorporated as a feature of the search engine or interface being used—a search guide.

Overall, our findings are encouraging and suggest that the tools examined in this study may be useful in helping users engage in Internet-based health information seeking tasks. The participants in our study were receptive to using the tools and perceived them as helpful. However, our data also suggest that the tools as currently designed may be too complex from a usability standpoint for older adults. As we noted, some of these features can be improved upon from a software standpoint, but we were required in a number of instances to overcome these issues through training which increased the cognitive complexity in the use of these tools. It is important to note that all of our participants had prior computer and Internet experience and, as an added precaution, were required to pass to two brief Internet search exercises, one easy and another more complex, before they could be part of the study. The findings also indicate that if these tools become available, training and instructional support materials must also be available. Our findings provide very useful and detailed information for the redesign of these tools. As indicated above, the refined tools could then be tested in a larger evaluation study with a larger sample that would use the tools over an extended period of time.

# List of Publications and Products


Budiu R, Pirolli P, Hong L. Remembrance of things tagged: How tagging effort affects tag production and human memory. *Proceedings of the ACM CHI*, 2009 April: 615-624.

Brown, JI, Fischo, VV, & Hanna, GS. Nelson-Denny Reading Test. Chicago: Riverside. 1993.

Chi E, Hong L, Gumbrecht, M, Card, S. (2005). Scent Highlights: Highlighting conceptually-related sentences during reading. *Proceedings of the Intelligent User Interfaces* (IUI), pp. 272-272, January.

Czaja S J, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, Sharit J. Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging*, 2006. *21*, 333-352.

Czaja SJ, Sharit J, Nair SN. Usability of the Medicare health web site. *JAMA* 2008 *300 (7)*: 790 - 792.

Delis, DC, Kramer, JH, Kaplan, E, Ober, BA. California Verbal Learning Test. San Antonio, TX: Psychological Corporation. 1987.

Ekstrom, RB, French, JW, Harman, HH, & Dermen, D. *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service. 1976.

Fisk AD, Rogers WA, Charness N, Czaja SJ, Sharit J. *Designing for older adults*: *Principles & creative human factors approaches, second edition*. Boca Raton: CRC Press. 2009.

Hong L, Chi E, Budiu R, Pirolli P, Nelson L. SparTag.us: A low cost tagging system for foraging of web content. *International Working Conference on Advanced Visual Interfaces* (AVI), 2008: 65-72.

Pew Internet & American Life Project. *Health Online 2013*. Available on-line at: http://pewinternet.org/Reports/2013/Health-online.aspx

Sharit J, Hernandez MA, Czaja SC, Pirolli P. Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. *ACM T Comput-Hum Int* 2008 *15*(*1*), 3:25.

Shipley, WC, *Shipley institute of living scale.* Los Angeles: Western Psychological Services. 1986.

Van Merrienbor JJG, Clark R, de Croock MBM. Blueprints for Complex Learning: The 4C/ID Model. *Educ Train Technol Res* 2002 50(2):39-64.

Wechsler, D. *Wechsler Adult Intelligence Scale*. 3rd. San Antonio, Tex: Psychological Corp 1997.

Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, Mockbee J, Hale FA. Quick assessment of literacy in primary care: the Newest Vital Sign. *Ann Fam Med* 2005; 3: 514–22.