Measuring Cognitive Load to Test the Usability of Web Sites

Janet Patton Tracy
Dr. Michael J. Albers

This paper provides an overview of how cognitive load theory applies to web site design and then considers three different methods of measuring cognitive load. NASA TLX is the standard method, but is a post-event measure. The Sternberg Memory Test and a tapping test both provide a means of measuring cognitive load while the person performs a usability test. By knowing the areas of high cognitive load, the designer can then consider what factors are causing the high load and potentially redesign the web site to reduce it.

INTRODUCTION

As more and more people facilitate web sites into their daily routines, it is important to create an interaction that is as easy as possible. As the web continues to evolve, more and more web site designs depend upon a complex, flexible structure (Reeves). However, the lack of understanding the major impediments to the user within that structure present a substantial problem to both designers and users. The designers lack a clear view of what areas to focus on to improve the interaction and users experience higher frustration levels and inability to find information. One method of determining the problem areas within a web site is to measure the user's cognitive load level. The designer can then take the results of the cognitive load measurements and work to redesign the problem areas.

This paper discusses how cognitive load effects web site usability and considers three different methods of measuring cognitive load. The Sternberg Memory Task, tapping task, and the NASA TLX can measure the levels of cognitive load. All three tests work at different granularity levels of their measurements and have different implementation issues.

COGNITIVE LOAD

A person has a limited amount of cognitive resources which they must allocate amount all mental tasks being concurrently performed. Cognitive load refers to the total amount of mental activity on working memory at an instance in time (Cooper). Short term memory is limited in the number of elements it can contain simultaneously (Sweller). If a design requires the user to hold too many items in short term memory, it will fail (Balogh). When a user's working memory is available to concentrate on the details of to-be-used information, usability is increased (Cooper).

Technical writers have discovered that readers appreciate brief chunks of information whether they are located online or printed on paper. Segments of information are gathered into groups, called chunks of information. Working memory is only capable of retaining five to nine chunks of information (Baddeley). It has become apparent that this theory is even truer for online documentation since computer screens only provide a limited view of long documentation.

An effective design allows a user to focus on the information of interest to them. While this statement sounds obvious, many web site architectures are an obstacle to be overcome. They impose a high cognitive load with poor navigation or forcing the user to figure out cryptic categories or link names. All of these draw cognitive resources away from the user's primary task of finding and comprehending the site's information.

Besides proper design factors which a designer or writer can control, many external factors also effect the amount of cognitive resources a user has available. Stress, time pressure, and lack of sleep are but a few of many factors which can decrease the resource pool. A design which tests fine in a usability lab may fail in the field because of these external factors. The normal stress of a job may push the user into cognitive overload although this wasn't seen during user testing.

As a user's cognitive load increases, their ability to perform effectively slowly decreases until they reach a point of cognitive overload. Since a person starts with a very limited pool of cognitive resources, a poor design can easily exhaust it. At that point, performance drops sharply and frustration and error rates sky rocket.

As web sites become more complex, user confusion becomes of greater concern, and confusion forces cognitive overload (Reeves). Successful user interface must respect the limits of human cognitive processing (Balogh). The overwhelmed/confused reaction users often experience is due to abundant options that increase cognitive load (Celt). Cognitive overload can be divided into two categories: saturation, when overload is too great to process; and pollution, when a user must process unnecessary flawed or poorly designed information. Inadequately planned websites force both, saturation and pollution, on a user (Albers).

Usability testing should measure cognitive load as soon as basic prototypes of the web site are available and problems can be fixed with minimal expense (Mowat). Successful measurements will require a working prototype, rather than paper mock ups or wire frames, but as pieces become available, they should be tested. Usability can flush out areas of high cognitive load and help obtain the goal of creating a web site that is as useful for the user as possible. The problem areas may include issues such as poor navigation categories, inconsistent design, or poorly written information.

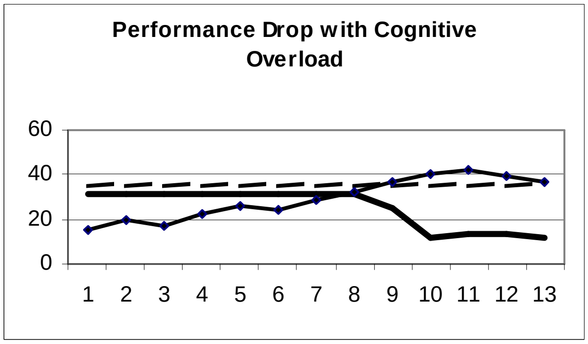

Figure 1. When the cognitive load level (increasing line) crosses the overload threshold (dotted line), the performance (heavy line) drops abruptly drops off.

Examining the levels of cognitive load when searching informational web sites is important in determining how the websites are interpreted. The easiest method of measuring cognitive load involves imposing a secondary task on the subject (Figure 2). Besides the primary task of finding information on a web site, the person has to perform a second task. When simultaneously performed, the primary and secondary tasks share the person's cognitive resources. Since a person has a very limited pool of cognitive resources, when primary task performance exhausts that pool. No longer having the amount of resources that they usually require, the primary or secondary tasks are affected and their performance is presumed to decrease". (Olive). They slough the second task to concentrate resources on the primary task. In other words, the person discards secondary tasks and proceeds with the primary tasks. As users become confused or lost they will drop the secondary task. At this point, we have some foundation to state that areas of usability are evident.

This sloughing of tasks in a primary benefit of imposing a secondary task load. The primary task may not have a high enough load level to exhaust a person's cognitive resources and thus make it hard to measure points of overload. However, the combined requirements of the primary and secondary task make causing overload more likely.

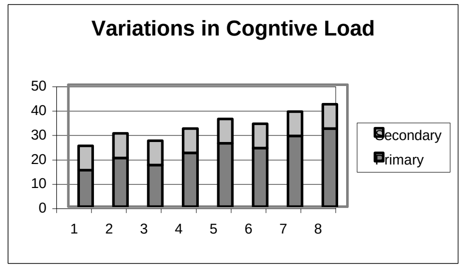

Figure 2. While the cognitive resources consumed by the primary task varies, the resources of the secondary task remain constant.

METHODS OF MEASURING COGNITIVE LOAD

There are many different methods of measuring a user's cognitive load, ranging from direct measurement of neuro-physiological response to post-event questions. While techniques measuring physical responses such as EEG and pulse rates are very accurate, they are also expensive and require special equipment and training. This paper discusses three methods which are easily applied within a simple usability test.

NASA TLX

The NASA TLX test was developed to measure the overall work load of equipment operation. "It can be used to assess workload in various human-machine environments such as aircraft cockpits; command, control, and communication (C3) workstations; supervisory and process control environments; simulations and laboratory tests" (Manning).

To assess the workload experienced while completing multiple tasks, six rating scales have been established: mental demand, physical demand, temporal demand, performance, effort, frustration (JTAP). The user rates each of these on a Likert scale and then has 15 questions that pair up two scales and ask the user to select the one which is the most important contributor to workload for the task.

The concept of the NASA-TLX is to allow the user time to access the workload situation once the testing is complete. The subscales provide detailed information, not just one-answer questions. We must consider that it measures workload as an afterthought. NASA-TLX does not test the user while they are in the process of completing the task. The user is forced to rely on what they remember and provide and opinion based on memory.

NASA TLX is designed for work loads which often had a higher physical component than computer mouse operations. On the other hand, for the information seeking activities which are the norm on web sites, the cognitive load a person experiences tends to be directly related to the work load.

An electronic version of the test can be downloaded at http://www.nrl.navy.mil/aic/ide/NASATLX.php.

Sternberg Memory Test

As originally designed, the Sternberg Memory Test was designed to measure how quickly people can search for and retrieve information from short-term memory. People were give a small set of number (1-6) to memorize and were then give a sequence of probe numbers. The subjects had to respond yes/no to whether the probe number was one of the numbers they had memorized. For example, you memorize 3 and 6. The probe number is 8, you say "no." The probe number is 6, you say "yes." The part of the website using high cognitive resources can be determined because the yes/no response will take longer.

According to Sternberg, several theories of short-term memory can be tested by altering the number of items on the list. Sternberg found that as the theory set increased, reaction times increased, and whether the probe was or was not committed to memory, did not alter the reaction time (Cog Lab Wadsworth).

While this test sounds very simple, the speed of response (saying yes/no) varies with the cognitive load. For high load situations, a user responds slower and if they are overloaded, there could be a substantial delay or out-right forgetting of some of the numbers which were memorized (Miyake).

Using the Sternberg Memory Test

At the start of the usability test, subjects are shown a short (1-6) list of numbers and are asked to commit the numbers to memory.

The subjects are then given a task to perform on a web site.

While the subject is performing the task, a tape recorder is running and a number is given at fixed intervals (30-45 seconds) and the person must respond if it is on the memorized list. It works best to have a tape recording of the numbers rather than the test administrator give the numbers. A second recorder (such as Camtasia) is needed to record both the probe number and the response, and the area of the web site the person is interacting with.

Post-test analysis measures the time interval between probe and response. For the above average response times (areas imposing high cognitive load), the designer should examine the specific area of the web site to determine what the user was doing at the time and if a redesign is appropriate.

Tapping Test

Using tapping is a simple way of imposing a secondary load on the user. If the subject's concentration is focused at processing information other than tapping, it is hard for subjects to apply the concentration needed to execute the tapping task (Miyake).

While seemingly a trivial task, it does require cognitive resources to continue to rhythmically tap either a finger or foot. This imposes the additional load which helps push the user into cognitive overload.

Using the tapping test

At the start of the usability test, the subject is told to tap with their non-dominate hand. They should be instructed to tap with a steady rhythm. A speed of about 1 tap per second works well to keep the person's hand from getting tired during the test. If both hands are required during the test, such as for data entry work, toe tapping also works.

The subjects are then given a task to perform on a web site and reminded to tap the entire time.

A recorder (such as Camtasia) is needed to record both the tapping sound and the area of the web site the person is interacting with.

Post-test analysis examines the areas of the web site where tapping slowed down or stopped. Areas with slow or unrhythmical tapping are imposing an extra load. Areas where tapping stopped have pushed the user into cognitive overload and should be redesigned.

CONCLUSION

NASA TLX is a post-event test which captures how people thought about or remember the interaction; whereas, the Sternberg Memory task and the tapping task provide the ability to collect the users response throughout the interaction. As such, they provide a means of finding the specific areas of a web site which are causing users problems.

The recordings of both the tapping test and the Sternberg Memory Test reveal the areas of high cognitive load and thus areas of increased user frustration within a given site.

The Sternberg Memory Task is not as fine grained as the tapping task, but is easier for the user. There is a chance the high load areas occur between probe numbers. Also, the person is forced to suspend their web site action and respond to the number probe. The speed and ease at which a person can make this shift is also a measure of cognitive load.

All three of these measure provide an additional level of usability testing besides the normal method of watching a user interact with a site. Both tapping and the Sternberg Memory Test provide a clear indicator of high load situation whereas with just observation, it is possible the user has simply paused and relaxed for a moment while doing the task.

REFERENCES

"A Neuroimaging Study of Automatic Movements." National Institute of Neurological Disorders and Strokes, 2004. NIH: Clinical Research Studies 11 Aug. 2004 <http://www.clinicaltrialssearch.org/imaging_study_of_automatic_movements.html>.

"Evaluating multimedia packages." *Educational Technology* 4 Aug. 2004 < http://www.csu.edu.au/division/celt/edtech/multimedia/mmeval.htm>.

"JTAP Project 305 Human-Computer Interface Aspects of Virtual Design Environments for Engineering Education: Application of MUSIC to the Evaluation of dVISE Implemented on an Immersive Display Platform." 1998. Joint Information Systems Committee of the Higher Education Funding Councils. Aug. 1997 < http://www.avrrc.lboro.ac.uk/jtap305/JTAP305expt1.html>.

"Sternberg Search". Wadsworth CogLab Online Laboratory Cognitive Psychology Online Laboratory 11 Oct. 2005 < http://coglab.wadsworth.com/index.html>.

Albers, Michael, and Loel Kim. "Information Design for the Small-screen Interface: An Overview of Web Design Issues for Personal Digital Assistants." Technical Communication 49.1 (2002): 45-60+.

Allen, Bryce. Information Task: Toward a User-Centered Approach to Information systems. San Diego: Academic Press, 1996.

Baddeley, Alan. "Is Working Memory Still Working?" *European Psychologist* 22 Jan. 2005 <http://www.manningaffordability.com/S&tweb/HEResource/Tool/Shrtdesc/Sh_NASATLX.htm>.

Balogh, Jennifer, Michael Cohen, and James P. Giangola. "Voice User Interface Design: Minimizing Cognitive Load." *Addison Wesley Professional* 2004. 9 Feb. 2004 <http://www.awprofessional.com/articles/article.asp?p=170792>.

Carey, T., B. Nonnecke, D. Lungu, and J. Mitterer. "Access Methods for Online Information: Cost/Benefit Approach to Users' Choices." Paper presented at the annual meeting of International Professional Communication Conference. Discataway, 1993.

Chandler, Paul, and John Sweller. "Cognitive Load While Learning to Use a Computer Program." *Applied Cognitive Psychology* 1996.University of New South Wales. 6 Jan 1999 <http:www3.interscience.wiley.com/cgi-bin/abstract/5000817/ABSTRACT>.

Cooper, Graham. "Research into Cognitive Load Theory and Instructional Design at UNSW." University of New South Wales. 9 Aug 2004 <http://www.google.com/scholar?hl=en&lr=&q=cache:BP2uyE_8R1EJ:www.uog.edu/coe/ed451/tHEORY/LoadTheory1.pdf+research+into+cognitive+load+theory+and+instructional+design+at+unsw>.

Dale, C.L., G. Gratton, and J. Gibbon. "Untitled." Neuroreport. 12.14 (2001): 3015-8+.

Manning. "NASA-TLX (NASA Task Load Index)." *Human Engineering Homepage*. 1998. 12 Dec. 2005 <http://www.manningaffordability.com/S&tweb/HEResource/Tool/Shrtdesc/Sh_NASATLX.htm>.

Miyake, Yoshihiro, Yohei Onishi, and Ernst Poppel. "Two Types of Anticipation in Synchronization Tapping." Acta Neurobiol 64 (2004): 415-426.

Mowat, Joanne. "Cognitive Walkthroughs: Where They Came From, What They Have Become, and Their Application to EPSS Design." The Herridge Group Inc. 19 Apr. 2002 <http://www.herridgegroup.com/pdfs/cognitive.pdf#search='joanne%20mowat%20cognitive%20walkthroughs'>.

MRSCI.com. 23 Jan. 2006 <http://www.mrsci.com/Neuroscience/Working_memory.php>.

NCARAI ~ Interface Design and Evaluation. http://www.nrl.navy.mil/aic/ide/NASATLX.php

Olive, Thierry. "Working Memory in Writing: Empirical Evidence From the Dual-Task Technique." *European Psychologist* 23 Jan. 2006 < http://cogprints.org/3246/01/Olive_Euro_Psy.pdf>.

Reeves, Thomas C., and Stephen W. Harmon. "User Interface Rating Tool for Interactive Multimedia." 5 Aug. 2004 < http://it.coe.uga.edu/~treeves/edit8350/UIRF.html>.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

Janet Patton Tracey
Technical Writer
Inventory Locator Service
Memphis TN USA
JTracy@ilsmart.com

Janet Patton Tracy is a Technical Writer for Inventory Locator Service (ILS) in Memphis, Tennessee. She holds a BA in English from the University of Memphis, and is completing a MA in Professional Writing and Editing from the University of Memphis.

Michael J. Albers
Associate Professor
University of Memphis
Memphis TN USA
malbers@acm.org

Michael J. Albers is an associate professor at the University of Memphis, where he teaches in the professional writing program. In 1999, he completed his PhD in technical communication and rhetoric from Texas Tech University. He has worked for 10 years as a technical communicator, writing software documentation and performing interface design. His research interests include designing documentation focused on answering real-world questions and online presentation of complex information.

