Technological University Dublin
ARROW@TU Dublin

Conference papers | School of Computer Sciences

2012

The Importance of Human Mental Workload in Web Design

Luca Longo
Technological University Dublin, luca.longo@tudublin.ie

Fabio Rusconi
Trinity College Dublin, Ireland

Lucia Noce
Trinity College Dublin, Ireland

See next page for additional authors

Follow this and additional works at: https://arrow.tudublin.ie/scschcomcon

 Part of the Computer Sciences Commons

Recommended Citation

Longo, L., Rusconi, F., & Noce, L. (2012). The importance of human mental workload in web design. *WEBIST 2012: 8th International Conference on Web Information Systems and Technologies*, 18-21 April, Porto, Portugal. doi:10.5220/0003960204030409

This Conference Paper is brought to you for free and open access by the School of Computer Sciences at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie.



This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License

Authors
Luca Longo, Fabio Rusconi, Lucia Noce, and Stephen Barrett

This conference paper is available at ARROW@TU Dublin: https://arrow.tudublin.ie/scschcomcon/179

THE IMPORTANCE OF HUMAN MENTAL WORKLOAD IN WEB DESIGN

Luca Longo, Fabio Rusconi, Lucia Noce and Stephen Barrett
Department of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
{*llongo@scss.tcd.ie*}

Keywords: Human Mental Workload, Interaction Design, Web Design, Usability, Human Factors.

Abstract: The focus of this study is the introduction of the construct of Human Mental Workload (HMW) in Web design, aimed at supporting current interaction design practices. An experiment has been conducted using the original Wikipedia and Google web-interfaces, and using two slightly different versions. Three subjective psychological mental workload assessment techniques (NASA-TLX, Workload Profile and SWAT) with a well-established assessments usability tool (SUS) have been adopted. T-tests have been performed to study the statistical significance of the original and modified web-pages, in terms of workload required by typical tasks and perceived usability. Preliminary results show that, in one ideal case, increments of usability correspond to decrements of generated workload, confirming the negative impact of the structural changes on the interface. In another case, changes are significant in terms of usability but not in terms of generated workloads, thus raising research questions and underlying the importance of Human Mental Workload in Interaction Design.

1 INTRODUCTION

Human Mental Workload (HMW) is a multidimensional complex construct mainly applied in Cognitive Science and sporadically used in Human Computer Interaction. Although a well-established definition is absent in the literature, the goal of measuring mental workload is to quantify the mental cost of performing tasks and having estimates of system and operator performance. Intuitively, mental workload is the amount of mental work necessary for a person to complete a task over a period of time. To be more precise, the construct emerges from the interaction between the requirements of a given task, the circumstances under which it is performed, the context and the skills, behaviours, emotional state and perceptions of the operator (Kantowitz, 1988). In the field of Interaction Design, the application of the construct of Human Mental Workload may have important practical implications in the interface design process and in the evaluation of system usability. For instance, assessments of the mental workload of users on tasks performed on a web-site may be useful for evaluating both the behavior of end-users upon it and the usability of the web-site itself.

This research investigates the application of the construct of Human Mental Workload in the field of Web Design. In detail, the main objective is to compare different HMW assessments techniques over tasks performed on selected web-sites, along with the study of their correlation with a well-known subjective assessments tool of interface usability. The issues investigated are:

- analysis of the distribution of mental workload levels over selected web-tasks, produced by three subjective assessment procedures, along with the study of their inter-correlations;
- examination of the correlation of the outputs produced by the three HMW procedures with the one by a well-established usability assessment technique (SUS - Subjective Usability Scale)
- impact analysis of HMW in Web Design.

The reminder of this paper is organized as follows. We review related works in the field of Human Mental Workload and we describe the main subjective assessment tools used in the research. We analyse applications of HMW in Human-Computer Interaction. We then introduce the experiments conducted in line with the research questions showing the outcomes following a critical discussion. We conclude highlighting the reasons and advantages that HMW can potentially bring in the field of Interaction and Web Design.

2 RELATED WORK

Human Mental Workload (HMW) is a multifaceted complex construct mainly applied in psychology and other cognitive sciences. A plethora of definitions exists in the literature (Hancock and Meshkati, 1988; Wickens, 1987; Cain, 2007; Gopher and Donchin, 1986). Intuitively, mental workload or cognitive workload is the amount of mental work necessary for a person to complete a task over a given period of time. Generally, it is not an inherent property, rather it emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviors and perceptions of the operator. The operational and practical nature of the construct of human mental workload, in the last few decades, has been acquiring interest in Neuro-science, Physiology and even Computer science (Kramer and Sirevaag, 1987; Kantowitz, 1988; Donnell and Eggemeier, 1998; Young and Stanton, 2001). There is a wide application field (Donnell and Eggemeier, 1998; Tracy and Albers, 2006; Xie and Salvendy, 2000; Gwizdka, 2009a; Gwizdka, 2009b) and this new research domain may have an important impact in the future, above all in Human-Computer Interaction. The concept has become increasingly important since modern interactive systems and interfaces may impose severe requirements on mental workload or information-processing capabilities.

There exist three major types of mental workload measures: performance-based, subjective and physiological. The rationale behind performance-based measures is that performance on a selected secondary task will decrease as a function of the demands of a selected primary task. Subjective measures include self-assessments using uni-dimensional or multi-dimensional scales. The former consider a measures of overall mental workload, the latter take into consideration individual dimensions of mental workload, therefore being more accurate in determining the source of any potential workload problem. Physiological measures are based on the premise that mental workload will generate changes in the body such as pupil dilation, changes in skin conductance, body pressure and heart rate. Although they are accurate and can work on a continuous scale, the equipment they require is generally impractical for experiments as it requires trained staff.

In this paper we focus on subjective multi-dimensional measures and we use three well-established tools: The NASA Task Load Index (NASA-TLX) (Kantowitz, 1988); The Simplified Subjective Workload Assessment Technique (SWAT) (Luximon and Goonetilleke, 2001); The Workload Profile (WP) (Tsang and Velazquez, 1996). In the following paragraphs, we briefly describe each technique, introducing the formal models in section 3.

NASA-TLX (Hart, 2006) uses six dimensions to estimate mental workload: mental demand, physical demand, temporal demand, performance, effort and frustration. Each of these is in a scale from 0 to 100. The final mental workload index is a weighted average of the six areas that provide an overall score. The weights are obtained via a paired comparisons which requires the operator to choose which dimension, across all pairs with the six dimensions, is more relevant to mental workload. The number of times a dimension has been chosen by the operator represents the weight of that dimension scale, for a given task (Kantowitz, 1988).

SWAT is a subjective multi-dimensional rating procedure that uses three areas to evaluate mental workload: time load, mental effort load, psychological stress load, each of them in a three-levels scale. In this paper we have adopted a simplified version of the SWAT model, the Continuous SWAT dimensions with weight (Luximon and Goonetilleke, 2001) which uses a paired comparisons among the three dimensions exactly as in the NASA-TLX model. The final mental workload is the average of the weighted areas (Luximon and Goonetilleke, 2001).

WP is a subjective workload assessment technique, based on the Multiple Resource Theory (MRT) of Wickens (Wickens, 1987). In this procedure eight dimensions are considered: perceptual/central processing, response selection and execution, spatial processing, verbal processing, visual processing, auditory processing, manual output and speech output. The WP procedure asks the operators to provide the proportion of attentional resources, in the range 0 to 1, used after the execution of a task. The overall workload rating is computed summing each of the 8 scores.

The three subjective techniques have low implementation requirements along with low intrusiveness and high subject acceptability. These peculiarities have promoted new research in which the construct of Human Mental Workload has been adopted for evaluating alternative interfaces. Tracy and Albers adopted three different techniques for measuring mental workload applied to web-site design: NASA-TLX, The Sternberg Memory Test and a tapping test (Tracy and Albers, 2006) (Albers, 2011). They proposes a technique to individuate sub-areas of a web-site, in which end-users manifested higher mental workload during interaction. In turn, this allowed designers to modify those critical regions for enhancing their interface. Zhu and Hou (Zhu and Hou, 2009) noted how roles can be useful in interface design and proposed a role-

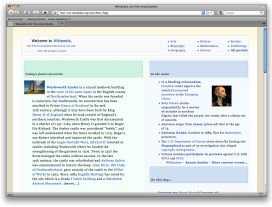
based method to measure the mental workload. This can be applied in the field of Human-Computer Interaction for dynamically adjusting workload levels of humans to enhance their performance of interaction.

3 METHODOLOGY

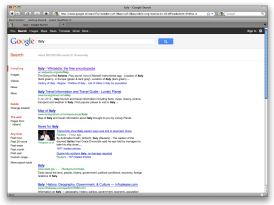
To investigate the research issues, we have designed four web-tasks as depicted in figure 1. Nineteen people aged between 19 and 35 years with different cultures, native language and ethnic background participated in the experiment. Web-tasks were designed on top of two major web-sites: wikipedia.com and google.com.



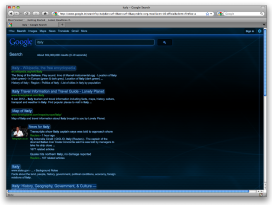
(a) W1 (Original Wikipedia)



(b) W2 (Mod. Wikipedia)



(c) G1 (Original Google)



(d) G2 (Mod. Google)

Figure 1: Web-interfaces used in experiment web-tasks.

Task W1 was designed to be performed on the original interface of Wikipedia (screenshot 1 a) while task W2 on a modified version (screenshot 1 b). Similarly task G1 was designed to be executed on the original google interface (screenshot 1 c) and task G2 on the modified version (screenshot 1 d). For the interface of picture 1 b, the layout has been modified, removing the left menu and the searching box from the original wikipedia along with the various tabs at the top of the page. For the interface of picture 1 d, the structural changes involved the removal of the left menu and the application of a different background. Volunteers needed to naturally interact with the web-browser and execute as best as they could the designed tasks. After the completion of each task, they were asked to fill in 4 questionnaires (random order):

- NASA-TLX questionnaire, as shown in table 2, along with the pair-wise comparisons among the 6 dimensions. The question related to the "physical demand" dimension (NTQ_2) was omitted due to the cognitive nature of tasks, thus its pair-wise comparison was automatically assigned to the other 5 dimensions (Answers scale: 0 [strongly disagree] to 100 [strongly agree]);
- WP questionnaire, as in table 3 (Answers scale: 0 [strongly disagree] to 100 [strongly agree]);
- SWAT questionnaire, as described in table4. Users were asked to choose, for each of the three areas (time load, mental effort load, psychological load) one of three possible levels;
- SUS questionnaire, as reported in (Brook, 2008) and shown in table 5, for evaluating the usability of the interface used in the performed task.

NASA-TLX. The NASA-TLX model is not based on a simple average of the 6 questions of Table 2 (NTQs), rather on a weighted aggregation.

These questions are formally expressed as:

$$NTQ_x : [0..100] \in \aleph, \qquad x_1 : \{M|PD|T|E|F\} \quad x_2 : \{P\}$$

The weight of each of the 6 considered areas needs to be computed. Users needed to decide, for each possible pair of the 6 areas (binomial coefficient), *'which of the two contributed more to their workload during the task'* (Example: Mental or Physical Demand? Physical Demand or Performance? and so forth). This cross tabulation generates 15 preferences:

$$\binom{6}{2} = \frac{6!}{2!(6-2)!} = 15$$

The weights are the number of preferences, for each area, in the 15 answers set (number of times that each factor was selected). They range from 0 (not relevant) to 5 (more important than any other factor). Formally: $NTW_x : [0..5] \in \aleph$. The final human mental workload score (HMW) of users is computed by multiplying the score of each NTQ question by its computed weight NTW (Note the Performance value is brought back to the original NASA-TLX scale):

$$NASA_{HMW} : [0..100] \in \Re$$

$$NASA_{HMW} = \left(\sum_{x_1} NTQ_x \cdot NTW_x + \sum_{x_2} (100 - NTQ_x) \cdot NTW_x \right) \frac{1}{15}$$

WP - Workload Profile. For each questions of the Workload Profile questionnaire (table 3) subjects provided a number between 0 and 100. A rating of 0 indicates that the task placed no demand on the dimension being rated, while a rating of 1 means that the task required full attention. Formally: $WPQ_i : [0..100] \in \mathbb{R}$ The ratings of each individual dimensions, that means the answer for each of the questions in table 3, are summed toward an overall mental workload rating.

Table 1: Experiment Tasks on Google and Wikipedia.

Task #	Description
$W1$	Using www.wikipedia.com find out how many people currently live in Sydney.
$W2$	Only using http://en.wikipedia.org/wiki/Main_Page find out how many people currently live in Sydney.
$G1$	Using www.google.com find how many years passed between the foundation of Apple Computer Inc. and the year of the 14th FIFA world cup.
$G2$	Using www.google.com find out how many years passed between the foundation of the Microsoft Corp. and the year of the 23rd Olympic games.

Table 2: Nasa Task Load Index (NASA-TLX) questionnaire.

Label	Question	Area
NTQ_1	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?	Mental Demand
NTQ_2	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?	Physical Demand
NTQ_3	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?	Temporal Demand
NTQ_4	How hard did you have to work (mentally and physically) to accomplish your level of performance?	Effort
NTQ_5	How successful do you think you were in accomplishing the goals, of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?	Performance
NTQ_6	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	Frustration

Table 3: Workload Profile (WP) questionnaire.

Label	Question	Area
WPQ_1	How much attention was required for activities like remembering, problem-solving decision-making, perceiving (detecting, recognizing and identifying objects)?	Perceiving / Remembering / Solving / Deciding
WPQ_2	How much attention was required for selecting the proper response channel (manual - keyboard/mouse, or speech - voice) and its execution?	Selection / Execution of Response
WPQ_3	How much attention was required for spatial processing (spatially pay attention around you)?	Task and Space
WPQ_4	How much attention was required for verbal material (eg. reading, processing linguistic material, listening to verbal conversations)?	Verbal Material
WPQ_5	How much attention was required for executing the task based on the information visually received (eyes)?	Visual Resources
WPQ_6	How much attention was required for executing the task based on the information auditorily received (ears)?	Auditory Resources
WPQ_7	How much attention was required for manually respond to the task (eg. keyboard/mouse usage)?	Manual Response
WPQ_8	How much attention was required for producing the speech response (eg. engaging in a conversation, talk, answering questions)?	Speech Response

Table 4: Simplified Subjective Workload Assessment Technique (SWAT) questionnaire.

Label	Possibilities	Value	Area
$SWATQ_1$	Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.	1	Time Load
	Occasionally have spare time. Interruptions or overlap among activities occur infrequently.	2	
	Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.	3	
$SWATQ_2$	Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.	1	Mental Effort Load
	Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.	2	
	Extensive mental effort and concentration are necessary. Very complex activity required total attention	3	
$SWATQ_3$	Little Confusion, risk, frustration, or anxiety exists and can be easily accommodated	1	Psychological Stress Load
	Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance	2	
	High to very intense stress due to confusion, frustration, or anxiety. High extreme determination and self-control required	3	

$$WP_{HMW} : [0..8] \in \mathbb{R} \qquad WP_{HMW} = \sum_{i=1}^{8} (WPQ_i) \cdot \frac{1}{100}$$

SWAT - Simplified Subjective Workload Assessment Technique. For the SWAT questions (table 4) subjects provided a number between 1 and 3 for selecting the appropriate option among 3 possibilities. $SWATQ_i : [1..3] \in \mathbb{N}$ After they indicated a preference, for each pair combination among the 3 dimensions, (Time Load, Mental Effort and Psychological stress). This cross tabulation generates 3 preferences,

Table 5: System Usability Scale (SUS) questionnaire.

Label	Question	Label	Question
$SUSQ_1$	I think that I would like to use this interface frequently	$SUSQ_6$	I thought there was too much inconsistency in this interface
$SUSQ_2$	I found the interface unnecessarily complex	$SUSQ_7$	I would imagine that most people would learn to use this interface quickly
$SUSQ_3$	I thought the interface was easy to use	$SUSQ_8$	I found the interface very unmanageable (irritating or tiresome) to use
$SUSQ_4$	I think that I would need the support of a technical person to be able to use this interface	$SUSQ_9$	I felt very confident using the interface
$SUSQ_5$	I found the various functions in this interface were well integrated	$SUSQ_{10}$	I needed to learn a lot of things before I could get going with this interface

so each dimension has max 2 occurrences:

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = 3 \qquad SWATW_i : [0..2] \in \aleph$$

The number of occurrences, for each dimension, are used to weight the original scores, producing a final human mental workload assessment as in the follow:

$$SWAT_{HMW} : [1..3] \in \mathbb{R} \quad SWAT_{HMW} = \Big[\sum_{i=1}^{3}(SWATQ_i \cdot SWATW_i)\Big] \cdot \frac{1}{3}$$

SUS - System Usability Scale. The original answers of the SUS questionnaires (Bangor et al., 2008) use a Likert scale, bounded in the range 1 to 5. Volunteers were asked to answer the questions with a scale ranged 0 to 100: $SUSQ_i : [1..100] \in \mathbb{N}$ Individual scores are not meaningful on their own. For odd questions (SUS_i with $i = \{1|3|5|7|9\}$), the score contribution is the scale position ($SUSQ_i$) minus 1. For even questions (SUS_i with $i = \{2|4|6|8|10\}$), the contribution is 5 minus the scale position. Formally:

$$SUS_{TOT} : [0..100] \in \aleph \ , \qquad i_1 = \{1,3,5,7,9\} \quad i_2 = \{2,4,6,8,10\}$$
$$SUS_{TOT} = \big[\textstyle\sum_{i_1}(SUSQ_i) + \sum_{i_2}(100 - SUSQ_i)\big] \cdot \frac{1}{10}$$

4 RESULTS AND DISCUSSION

Experimental results are shown in table 6 and in figure 2 while the correlations between the outcomes of each computational model are presented in table 7.

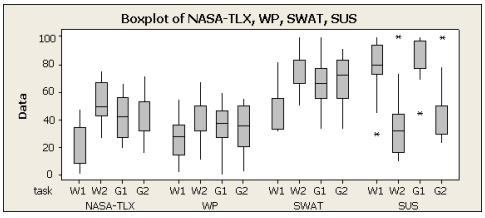


Figure 2: Boxplot of data grouped by model.

The Paired T-Test procedure was used to compare the mean difference between the results of the HMW-based models (NASA-TLX, WP, SWAT) and the ones of the usability model (SUS) on the tasks (W1 against W2, and G1 against G2). In particular, we tested the null hypothesis H_0 with a confidence interval $CI = 95\%$. The results are listed in tables 8 and 9.

Table 6: Distributions of data grouped by model.

Task	NASA		SWAT		WP		SUS	
	Avg	*Std*	*Avg*	*Std*	*Avg*	*Std*	*Avg*	*Std*
W1	21.3	14.3	46.1	15.2	26.0	14.8	78.3	18.7
W2	52.4	13.4	74.3	13.2	38.7	15.0	34.3	22.5
G1	41.9	15.1	66.6	18.7	34.9	16.2	84.9	13.7
G2	43.0	14.7	66.5	19.4	33.7	15.7	45.2	19.3

Table 7: Correlations of models.

Task	NASA/ SWAT	SUS/ NASA	WP/ NASA	SUS/ SWAT	WP/ SWAT	WP/ SUS
W1	0.75	-0.17	-0.10	-0.08	-0.07	-0.27
W2	0.52	-0.29	0.50	0.04	0.35	0.27
G1	0.82	-0.22	0.75	0.01	0.48	-0.14
G2	0.88	-0.03	0.80	-0.15	0.76	-0.12

The experiments conducted are preliminary, aimed at showing the potential of Human Mental Workload in Interaction and Web Design. From table 7 a clear correlation between the three HWM-based tools used (NASA-TLX, WP, SWAT) against the usability assessment technique (SUS) (Columns 3, 5, 7) does not appear. Further investigation needs to be carried out to discover correlations (if any). On the other hand, the three HMW-based models highly correlate, underlying a common view about the generated workloads on tasks (Columns 2, 4, 6). This suggests further experiments could consider only one of them, reducing the volunteers' answer-set. The only exception is for task W1 for the correlations WP/NASA and WP/SWAT: here the proportions of attentional resources required by the task do not clearly emerge. A possible explanation may rely in the uncertainty faced by volunteers, in answering the WP questions (table 3) for task W1. Tables 8 and 9 show the paired T-test of the four models for the task conducted on Wikipedia (W1 against W2) and on Google (G1 against G2). The first line (W1, W2) clearly shows that the null hypothesis is rejected in applying the four models (NASA-TLX, WP, SWAT, SUS): the two interfaces W1 and W2 have generated statistically significant workloads and usability scores. In turn this fact might be interpreted, by a web-designer,

Table 8: Paired T-test for NASA-TLX and WP.

Task	NASA			WP		
	T	P	H_0	T	P	H_0
W1, W2	−5.98	< 0.001	Rej.	−3.39	0.003	Rej.
G1, G2	−0.37	0.718	Acc.	0.81	0.428	Acc.

Table 9: Paired T-test for SWAT and SUS.

Task	SWAT			SUS		
	T	P	H_0	T	P	H_0
W1, W2	−5.56	< 0.001	Rej.	6.59	< 0.001	Rej.
G1, G2	0.02	0.987	Acc.	7.57	0.000	Rej.

negatively: the structural changes developed in the interface W2 negatively affect perceived usability and have an higher impact on the required mental workload of end-users. The second line of both the tables (G1, G2) shows that the null hypothesis is accepted for the three HMW-based tools, underling no statistical difference, in workload levels, between the two interfaces. On the contrary, for the paired T-test of the two SUS outcomes, the hypothesis is rejected: the usability of the two interfaces (G1 and G2) are perceived being statistically different. Considering the above interpretations, some potential research question arise:

- Is SUS, (or other usability assessment tool), sufficient for designing usable interfaces?
- Can the application of Human Mental Workload be a alternative or supporting procedure in the field of Interaction and Web Design?

5 CONCLUSIONS

The main aim of this contribution is to show the potential of the construct of Human Mental Workload in Interaction and Web Design. It has been shown how assessments of workload can be achieved on web-based tasks and how they can be applied for evaluating the impact of structural changes on web-sites. Three subjective psychological techniques for assessing mental workload were described: NASA Task Load Index, Workload Profile and a simplified Subjective Workload Assessment Technique. An experiment on two popular web-sites' interfaces was conducted: Wikipedia and Google. Four similar information search tasks were performed by 19 volunteers, two on the original interfaces and two on modified versions. Results show how, in the tasks performed on Wikipedia, increments in required mental workload correspond to decrements in usability perception and vice-versa, underlying an inverted correlation. This ideal case does not occur on the tasks performed on the two different Google interfaces where, for a decrement of perceived usability, the mental workload remains stationary. This suggests that usability should not be used in isolation for evaluating interactive interfaces but instead an analysis of the behaviour of end-users on typical web-based task, on a given interface, should be accounted. Our proposal is for the application of mental workload as a evaluation measure. This preliminary evidence should be supported by further investigations and applications. Future work will be focused on further experiments on other web-sites towards a general-applicable paradigm for aggregating workload and usability scores.

ACKNOWLEDGEMENTS

We are grateful to Noce L. and Rusconi F. who contributed to the design and execution of experiments.

REFERENCES

Albers, M. (2011). Tapping as a Measure of Cognitive Load and Website Usability. *Proceedings of the 29th ACM international conference on Design of communication*, pages 25–32.

Bangor, A., Kortum, P., and Miller, J. (2008). An empirical evaluation of the System Usability Scale (SUS). *International Journal of Human-Computer Interaction*, 24(6):574594.

Brook, J. (2008). SUS. A quick and dirty usability scale. *International Journal of Human-Computer Interaction*, 24(6):574594.

Cain, B. (2007). A Review of the Mental Workload Literature. *Technical Report, Defence Research and Development*.

Donnell, R. O. and Eggemeier, F. (1998). Modeling mental workload. *Cognitive Technology*, 3:9–31.

Gopher, D. and Donchin, E. (1986). Mental Workload Dynamics in Adaptive Interface Design. *Handbook of Perception and Human Performance*, 2(41):1–49.

Gwizdka, J. (2009a). Assessing Cognitive Load on Web Search Tasks. *Ergonomics Open Journal*, 2:114–123.

Gwizdka, J. (2009b). Distribution of Cognitive Load in Web Search. *Journal of the American Society for Information Science & Technology*, 61(11):2167–2187.

Hancock, P. and Meshkati, N. (1988). *Human Mental Workload*. Elsevier.

Hart, S. G. (2006). Nasa-Task Load Index (Nasa-Tlx); 20 Years Later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 50(9):904–908.

Kantowitz, B. (1988). Development of Nasa-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload*, 52:139–183.

Kramer, A. and Sirevaag, E. (1987). A Psychophysiological Assessment of Operator Workload During Simulated Flight Missions. *Human Factors*, 29(2):145–160.

Luximon, A. and Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics*, 44(3):229–243.

Tracy, J. P. and Albers, M. J. (2006). Measuring Cognitive Load to Test the Usability of Web Sites. *Usability and Information Design*, pages 256–260.

Tsang, P. and Velazquez, V. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3):358–381.

Wickens, C. (1987). Information processing, decision making, and cognition. *Cognitive engineering in the design of humancomputer interaction and expert systems*.

Xie, B. and Salvendy, G. (2000). Prediction of Mental Workload in Single and Multiple Task Environments. *International Journal of Cognitive Ergonomics*, 4(3):213–242.

Young, M. and Stanton, N. (2001). Mental Workload: Theory, Measurement, and Application. *International Encyclopedya of Ergonomics and Human Factors*, 1:507–509.

Zhu, H. and Hou, M. (2009). Restrain mental workload with roles in hci. In *Proceedings of Science and Technology for Humanity*, pages 387 – 392.