Using fNIRS in Usability Testing: Understanding the Effect of Web Form Layout on Mental Workload

Kristiyan Lukanov*
Fadata AD
Sofia, Bulgaria
kristian.lukanov@fadata.eu

Horia A. Maior*
Horizon CDT
University of Nottingham, UK
horia.maior@nottingham.ac.uk

Max L. Wilson
*Mixed Reality Lab
University of Nottingham, UK
max.wilson@nottingham.ac.uk

ABSTRACT
Amongst the many tasks in our lives, we encounter web forms on a regular basis, whether they are mundane like registering for a website, or complex and important like tax returns. There are many aspects of Usability, but one concern for user interfaces is to reduce mental workload and error rates. Whilst most assessment of mental workload is subjective and retrospective reporting by users, we examine the potential of functional Near Infrared Spectroscopy (fNIRS) as a tool for objectively and concurrently measuring mental workload during usability testing. We use this technology to evaluate the design of three different form layouts for a car insurance claim process, and show that a form divided into subforms increases mental workload, contrary to our expectations. We conclude that fNIRS is highly suitable for objectively examining mental workload during usability testing, and will therefore be able to provide more detailed insight than summative retrospective assessments. Further, for the fNIRS community, we show that the technology can easily move beyond typical psychology tasks, and be used for more natural study tasks.

ACM Classification Keywords
H.5.2. [Information Interfaces and Presentation] User Interfaces. - Graphical user interfaces

Author Keywords
Mental workload; web forms; usability; functional near-infrared spectroscopy; fNIRS;

INTRODUCTION
One process that is time consuming, and often one that we cannot avoid, is filling web forms with information. These might be mundane, like registering with a new website, procedural, like online banking, or complex and important, like completing a tax return. Regardless, we encounter them frequently, if not many times a day. Complex forms, like tax returns and car insurance claims, can be very difficult for users, and potentially create a significant amount of stress. From a Usability perspective, the forms should be designed so they are intuitive and easy to use, aiding users through the filling process, and helping them to avoid making errors. Although usability has many facets [9], one concern is for the mental workload (MWL) of the users [10], which is typically measured using retrospective subjective forms [19]. In this paper, however, we explore the use of concurrent objective measure of MWL: functional Near Infrared Spectroscopy (fNIRS). Although research has demonstrated that this technology is suitable for user study conditions, the evidence has been based upon constrained psychology tasks like N-Back tests [4, 20, 28]. Here, we demonstrate that fNIRS can be easily integrated into typical usability testing conditions, as we compare alternative designs for an online insurance claim form.

Mental workload is *"the relationship between primary task performance and the resources demanded by the primary task"* [44]. In their recent model of MWL, Sharples and Megaw highlight that both high and low MWL, can reduce performance [44]. As task complexity increases, performance reduces. Conversely, a repetitive task that does not utilize a person's mental or physical resources may result in boredom and apathy [2], which also means that a user becomes prone to errors [32]. It is therefore important in the field of HCI evaluation and Usability to understand users' capabilities and limitations, in order to assess the demands placed upon them.

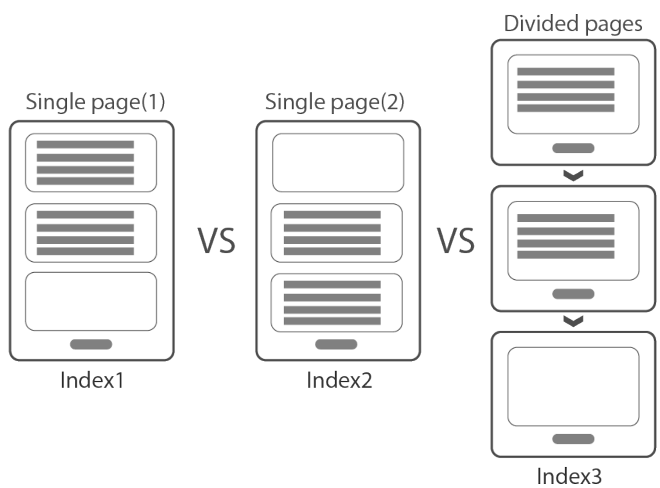

Figure 1. A sketch of the three web form layout variations investigated.

The aim of this usability study is to evaluate a web form filling interface for the insurance domain, and we use the example

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/5 $15.00
DOI: http://dx.doi.org/10.1145/2858036.2858236

of an online auto insurance claim process. We therefore evaluate three variations (shown in Figure 1) in the layout of the insurance claim form, based upon form design recommendation and discussed below, and examine their effects on user preference, emotional response, workload, and performance.

Usability and Web Form Filling

Web form filling is a task often encountered within our daily activities, however, according to our knowledge, there is scarcely any *empirical* research literature for this topic. Most of the research on web form filling and design is focused on optimizing experience and accessibility for an elderly population [14, 23, 35, 36]. Wästlund et al [1] compared two web page layouts, one in which all the text is on single page, and one where the text is separated over four pages, and concluded that users experienced less workload with the divided web form. Further research [22, 45] also suggest splitting long web forms into several pages in order to improve the process. It is suggested that the longer it takes for a task to be completed (short or long term) the more the perceived frustration users experience [8, 27]. Consequently, time to complete was also considered in our study.

Instead, the design of web form interfaces is often based on usability guidelines, as they are widely accepted in practice. The two most popular usability heuristics are those of Nielsen [30], and Shneidermann [39]. Because both of them advocate minimizing the load on working memory, we consider that reducing MWL will provide better user experience. This, therefore, motivates our concern with being able to accurately measure MWL, as most measures are summative retrospective and subjective assessments provided by users after completing an entire task. With a concurrent objective measure, as proposed for fNIRS, ideally we would be able to examine the MWL at different parts of the task, and to combine with techniques like Think Aloud Protocol [33] in usability testing.

Measuring Mental Workload

Subjective measures are based upon user opinions and capture the users' experience of effort. Due to their simplicity, cheap running costs and recognised validity, they tend to be the most used and accepted workload measures. One of the most used subjective techniques is the multidimensional NASA-TLX scale [19]. Using the 6 sub-scales, it provides high diagnosticity, identifying different aspects of workload. Other such measures include SWAT [34] and the Workload profile [43]. Longo et al. [24] compared the three measures mentioned above in a web browsing/searching task, and observed correlations in the results of the three measures claiming that they measure the same concept of workload.

Psychophysical measures are used to give objective data about MWL by not relying on subjective scales or performance measures. They can be obtained by recording e.g. variable heart rate [5], electrodermal response (EDR) and galvanic skin response (GSR) [15, 38], pupil dealation [7], brain imaging [6], and facial skin temperature [41]. These techniques detect the change in the arousal from the autonomic nervous system level which can be inferred to as MWL. However, different psychophysical measures capture different aspects workload [13], therefore consideration should be put in choosing the most appropriate measure for the given task.

Recent research has shown that fNIRS is a suitable brain measurement technique for HCI studies [25, 33, 40] as it provides useful information about the user while allowing for more normal interaction with a computer system. fNIRS uses blood oxygenation, rather than electrical levels that are affected by limb movement, for determining the activation of areas in the brain, where higher oxygenation indicates increased activity. This makes it non-invasive, portable, and suitable for periods of extended monitoring relative to other neuroimaging techniques. fNIRS measures the delivery of blood to active neuronal tissues and it is designed to be placed directly upon a participant's scalp, typically targeting the prefrontal cortex (PFC). It has been suggested by cognitive neuroscience studies that PFC is involved in higher order cognition [12] and emotion processing [18]. In 2009, using abstract psychology tasks, Hirshfield et al [21] concluded that fNIRS should be suitable for evaluating usability. Later, Peck et al [31] used fNIRS to compare and evaluate different data representations using memory tasks. Peck et al found a negative correlation between the fNIRS levels of Hbr data and NASA-TLX, which was further confirmed by Maior et al [26].

In this paper, we examine the prospect of using fNIRS to measure MWL within a typical usability study. Although the work above has looked at using fNIRS in HCI user studies, it has all focused on proving its value using psychology experiments. Our primary contribution is to demonstrate how MWL can be objectively measured during usability testing to evaluate different user interface layouts. Further, as a secondary contribution, we also demonstrate that fNIRS can translate from constrained psychological tasks to more natural ones and still provide insight into MWL.

EXPERIMENT DESIGN

We chose to focus our study of form filling on a car insurance claim scenario, as it was a) an example of an important web form that people encounter in everyday lives, b) involves careful thought rather than just entering data, and c) represents a current concern for the insurance industry. Video clips of car accidents were used as the stimulus for filling in insurance claims during the study, with a time gap, such that participants had to recall aspects of the accidents. These video clips provided all the necessary data for filling in the forms, and were between 30-60s long. Further, the video clips were of low-speed accidents and were carefully chosen to be lightweight avoiding any scenes of gore, injured bodies, or fatalities. We compared three alternative designs, described below, and measured performance, time to complete, emotional responses, and MWL.

Layout variations of web forms

Three HTML/CSS variations of a standard web form for an insurance claim were produced, as shown in Figure 1. They were created to resemble an actual online insurance claim form, but excluded aspects such as brand and colour, to avoid these having an impact on the results. All conditions included three main parts: 1) Personal information, 2) Accident information and 3) a summary of the accident. The personal information, such as name, date of birth, etc, was given to participants on paper, such that they neither had to input their own personal information, nor would they have to apply additional MWL to generate fake information. The accident information (part 2) consisted of information about the time and location, and a series of drop down lists about the number of passengers and cars involved in the accident and the location of the damage on their vehicle. The final part consisted of a text-area, where participants had to write an overall description of the accident, what led to it, and so on.

Our control condition, Index1, was simply a form that contained all three parts on one page. Our experimental conditions were drawn from standard web form design recommendations [45]. The first experimental condition, Index2, tested the hypothesis that beginning with a general summary of the accident (part 3 at the top) would reduce the level of MWL overall for the form; summarising it would make subsequent box filling easier. The second experimental condition, Index3, tested the general design recommendation that breaking down a form into subforms (one for each part) would make it easier to fill out in stages. Users navigated between the three parts of Index3 using a submit button with the label “Next”.

Objective Measures

Aside from recording time to complete each task, hemodynamic data was recorded during each condition using the fNIRS300 device along with the COBI studio recording software developed by Biopac Systems Inc. The device consists of a headband with 4 infrared LED emitters and 10 infrared detectors, operating on 730nm and 850nm wavelengths. The combination between them was used to calculate 16 channels which can measure the associated oxygenated (HbO), deoxygenated (Hbr) and total (Hbt) haemoglobin concentration in the PFC. The emitter-detector separation was 2.5cm and the sampling rate was 2Hz.

Data Processing was performed using fnirSoft [3]. A low-pass filter with cut off frequencies of 0.1 Hz, was used in order to remove high-frequency noise, physiological artefacts such as heartbeats and motion derived artefacts. The fNIRS signal was then processed with modified Beer-Lambert law [16] in order to calculate HbO, Hbr and Hbt. The correlation based signal improvement (CBSI) [17] method was applied to remove motion artefacts. Data was divided into conditions using time markers at the start and end of each form-filling period, and after data preprocessing, the mean and standard deviation for Hbo, Hbr and Hbt was calculated from all channels.

Self-Reporting Measures

Nielsen and Levy [29] advise researchers to use combination of subjective and objective data in usability studies, in order to identify bias and provide richer information about the process. Accordingly, we also examine related subjective measurements within our study. To assess operator perceived MWL we used the paper version of the multidimensional subjective workload scale NASA-TLX [19]. The individual scales are presented in the following order: Mental demand, Physical demand, Temporal demand, Performance, Effort, and Frustration. The NASA-TLX scores were obtained from participants after completing each web form condition. To capture participants emotional valence and arousal we used a 5 point Self Assessment Manikin (SAM) [11] after each video and form-filling condition.

Procedure

A total of 15 right handed participants (5 female) with mean age of 26 (SD = 4.71) were recruited from the University of Nottingham. All participants were undergraduate or graduate students, had normal or corrected vision, and reported no history of brain damage. To begin with, informed consent was obtained from the participants. Then, when ready, the fNIRS headband was placed and calibrated using a period of rest, in order to obtain a baseline. Participants then followed the procedure shown in Figure 2. The data from 3 participants was excluded due to recording problems.

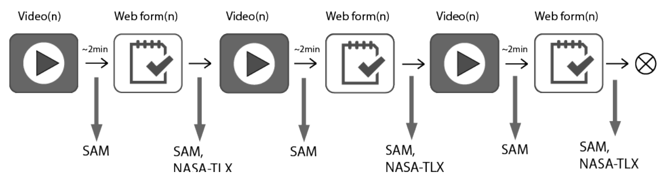

Figure 2. The study procedure followed in this experiment.

The three study conditions were counterbalanced using a Latin square, to avoid learning effects. Within each condition, participants first viewed a video clip, and then filled in a SAM test. SAM answers were informally monitored, as well as casual conversation between conditions, to check that participants were not experiencing distress from watching the videos. After a 2-minute gap, to allow for the accident to begin to decay from working memory, participants proceeded to fill in the current form condition (also counterbalanced). To conclude the condition, participants filled in a NASA-TLX form and another SAM test. After completing all three conditions this way, the study then concluded with a short debriefing interview.

RESULTS

Overall, based on post-study ratings, the bulk of participants preferred Index1 and Index3 with 8 and 6 votes respectively compared to Index2, which was preferred by 3 participants. Beyond using the SAM to monitor for ethical concerns whilst participants viewed footage of car crashes, we used this data to check that the videos could be considered comparable. There was no statistical difference as assessed with a Friedman test between the three videos for SAM emotional valence and arousal. There was also no significant difference in the mean time to complete the task in each web form condition ($F(2,28) = 0.498, p < 0.613$), nor the emotional valence associated with the three web forms ($X^2(2) = 5.15, p = 0.076$). Instead, below we examine our measures of MWL for a difference between conditions.

Objective Mental Workload - fNIRS

A one-way repeated measures ANOVA test found a significant difference in the mean levels of Hbo between the 3 web forms ($F(2,22) = 4.324, p < .026$, as shown in Figure 3. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $X^2(2) = 0.975, p = 0.879$. The measured mean Hbo was *higher* for Index3 $M = 0.644(SD = 1.37)$ compared to both Index2 $M = 0.384(SD = 1.28)$ and Index1 $M = 0.139(SD = 1.21)$. A Post hoc analysis with a Bonferroni adjustment revealed that Index3 had significantly higher Hbo activation than Index1: $p = 0.047$. No significant differences were found in Hbr and HbT.

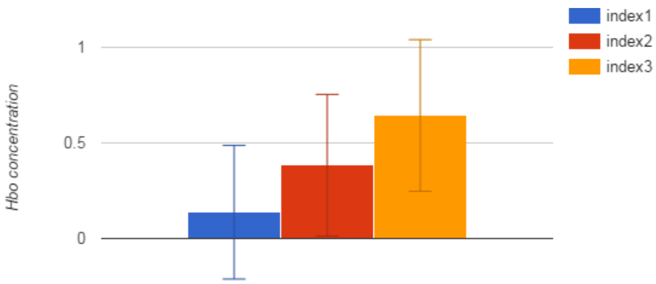

Figure 3. Mean Hbo activation between the three web form conditions as measured by fNIRS. Higher Hbo values indicates higher workload.

Subjective Mental Workload - NASA-TLX

A statistically significant difference was found between the three web forms in the perceived mental demand $F(2,28) = 4.677, p < .018$ score as assessed by one way repeated measures ANOVA. In line with the objective data from the fNIRS, participants perceived Index3 as the most mentally demanding with mean score of $M = 11.87(SD = 5.04)$, followed by Index2 with $M = 9.67(SD = 5.02)$ and Index3 $M = 8.73(SD = 4.41)$. Post hoc analysis with Bonferoni correction revealed significant interaction between Index1 and Index3 with $p = 0.018$.

DISCUSSION

The purpose of this study was to improve usability of web form filling of insurance claims and find which of three web forms layouts elicited the least workload and suggest implications for future design. Overall, we found that the designs of the forms did not create an emotional affect, and did not affect performance in terms of time. Further, we found that asking users to begin with summarising the accident (Index2) did not create a significant difference in MWL within our study. Dividing the form into separate subforms (Index3), however, increased MWL, according to both objective and subjective measures. Of the three forms, 8 out of 15 users expressed their preference towards our baseline Index1 condition. From herein, we focus on comparing single and divided forms.

One explanation for our findings is that a single form approach allows participants to go back and verify what information they have already entered: *"I remember in the second one (Index1) I'm not sure whether the option provided left or right, so I rechecked"* (P12). This way working memory resources are saved because participants have the ability to quickly recheck what they have already entered, relying on recognition, rather than recall [30]. Further, work in information browsers has shown that spatial configurations have significant benefits over temporal layouts [37]. Although often discounted by Human Factors researchers [44], an alternative model of Cognitive Load from the field of Education research [42] also notes that temporal distance increases Cognitive Load, as users cannot cross-reference material.

Despite the perspective of these models, our results were contradictory to our expectations because the use of subforms is very popular in practice, and we assumed that the divided page approach should reduce visual search and clutter, thus demand less attentional resources. Also, the more informational cues (web form fields) are present on an interface, the more time the user should spent on searching for information, thus the performance should drop. This claim can be partially supported with the feedback from P5: *"so the second one (Index3) I felt having to [navigate between subforms] broke it down a little bit, like you didn't have to think about everything in one go..."*.

Another advantage of the single page approach is that participants can choose which form to start first: *"the good thing about the number two(Index1) is everything is on the same page I can choose whatever I like"* (P9). Generally, some of the participants preferred to fill in the description field first, and then the rest of the form, so researchers and practitioners have to give users the power to choose from where they can start. Together, these insights indicate that future work should focus directly on the issue of creating optimal subforms, considering both the length of the form and the similarity of the focus of each subform. Its possible that the same study procedure will find that dividing a different form, of a different size, might reduce MWL. Future work might also investigate the concern of flexibility, as a heuristic, allowing participants to easily start at the part of the form they prefer.

CONCLUSION

In this paper we explored the use of fNIRS as an objective concurrent measure of MWL, alongside other measures, within a typical usability testing procedure. In particular, we examined performance, MWL, and preference of three alternatives to a car insurance claim form. While no differences were found in emotional or performance measures regardless of the completed form, both subjective and objective measures of MWL concluded that dividing the forms into separate pages significantly increased MWL. This was contrary to our expectations based upon the prevalence of divided forms online. As a primary contribution we further showed fNIRS to be suitable and useful within a typical usability evaluation settings. We also found our objective measure of MWL agreed with the subjective retrospective assessments. This means that future work can examine how MWL fluctuates during a task, as fNIRS generates a concurrent ongoing measure. Finally, we showed how fNIRS could translate from psychology tasks to natural form-filling tasks. This is a useful addendum to our findings, providing additional evidence for the utility of fNIRS as a MWL evaluation tool.

ACKNOWLEDGMENTS

This work was supported by the Horizon Centre for Doctoral Training (EP/G037574/1) and EPSRC Grant (EP/M000877/1). **Data access statement:** Consent was not gained from participants to put their data online; a dataset is not openly available.

REFERENCES

1. 2008. The effect of page layout on mental workload: A dual-task experiment. *Computers in Human Behavior* 24, 3 (2008), 1229 – 1245.
2. Daniel Afergan, Evan M Peck, Erin T Solovey, Andrew Jenkins, Samuel W Hincks, Eli T Brown, Remco Chang, and Robert JK Jacob. 2014. Dynamic difficulty using brain metrics of workload. In *CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3797–3806. DOI: http://dx.doi.org/10.1145/2556288.2557230
3. Hasan Ayaz. 2010. Functional Near Infrared Spectroscopy based Brain Computer Interface. *PhD Thesis, Drexel University, Philadelphia, PA* (2010).
4. Hasan Ayaz, Meltem Izzetoglu, Scott Bunce, Terry Heiman-Patterson, and Banu Onaral. Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy. In *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on*. DOI: http://dx.doi.org/10.1109/CNE.2007.369680
5. Richard W Backs and Kimberle A Seljos. 1994. Metabolic and cardiorespiratory measures of mental effort: the effects of level of difficulty in a working memory task. *International Journal of psychophysiology* 16, 1 (1994), 57–68. DOI: http://dx.doi.org/doi/10.1016/0167-8760(94)90042-6
6. Michela Balconi, Elisabetta Grippa, and Maria Elide Vanutelli. 2015. What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing. *Brain and cognition* 95 (2015), 67–76.
7. Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276. DOI: http://dx.doi.org/doi/10.1037/0033-2909.91.2.276
8. Katie Bessiere, Irina Ceaparu, Jonathan Lazar, John Robinson, and Ben Shneiderman. 2004. Social and psychological influences on computer user frustration. *Media access: Social and psychological dimensions of new technology use* (2004), 169–192.
9. Nigel Bevan. 2001. International standards for HCI and usability. *International journal of human-computer studies* 55, 4 (2001), 533–552. DOI: http://dx.doi.org/10.1006/ijhc.2001.0483
10. Nigel Bevan and Ian Curson. 1997. Methods for measuring usability. In *Human-Computer Interaction INTERACT'97*. Springer, 672–673.
11. Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59. DOI: http://dx.doi.org/10.1016/0005-7916(94)90063-9
12. Todd S Braver, Jonathan D Cohen, Leigh E Nystrom, John Jonides, Edward E Smith, and Douglas C Noll. 1997. A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 5, 1 (1997), 49–62.
13. Brad Cain. 2007. *A review of the mental workload literature*. Technical Report. DTIC Document. http://www.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf
14. Ann Chadwick-Dias, Michelle McNulty, and Tom Tullis. 2003. Web usability and age: how design changes can improve performance. In *ACM SIGCAPH Computers and the Physically Handicapped*. ACM, 30–37. DOI: http://dx.doi.org/10.1145/960201.957212
15. C Collet, E Salvia, and C Petit-Boulanger. 2014. Measuring workload with electrodermal activity during common braking actions. *Ergonomics* 57, 6 (2014), 886–896. DOI: http://dx.doi.org/10.1080/00140139.2014.899627
16. M Cope and David T Delpy. 1988. System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical and Biological Engineering and Computing* 26, 3 (1988), 289–294. DOI: http://dx.doi.org/10.1007/BF02447083
17. Xu Cui, Signe Bray, and Allan L Reiss. 2010. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 4 (2010), 3039–3046.
18. Antonio R Damasio, BJ Everitt, and D Bishop. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences* 351, 1346 (1996), 1413–1420. DOI: http://dx.doi.org/10.1098/rstb.1996.0125
19. Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9
20. Christian Herff, Dominic Heger, Ole Fortmann, Johannes Hennrich, Felix Putze, and Tanja Schultz. 2013. Mental workload during n-back task - quantified in the prefrontal cortex using fNIRS. *Frontiers in human neuroscience* 7 (2013). DOI: http://dx.doi.org/10.3389%2Ffnhum.2013.00935
21. Leanne M Hirshfield, Erin Treacy Solovey, Audrey Girouard, James Kebinger, Robert JK Jacob, Angelo Sassaroli, and Sergio Fantini. 2009. Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2185–2194. DOI: http://dx.doi.org/10.1145/1518701.1519035
22. Caroline Jarrett and Gerry Gaffney. 2009. *Forms that work: Designing Web forms for usability*. Morgan Kaufmann.
23. Lorna Lines, Oluchi Ikechi, K Hone, and Tony Elliman. 2006. Online form design for older adults: Introducing web-automated personalisation. In *Proceedings of HCI, the Web and the Older Population, workshop at HCI 2006*.
24. Luca Longo, Fabio Rusconi, Lucia Noce, and Stephen Barrett. 2012. The Importance of Human Mental Workload in Web Design. In *WEBIST*. 403–409.
25. Horia A Maior, Matthew Pike, Sarah Sharples, and Max L Wilson. 2015. Examining the reliability of using fNIRS in realistic hci settings for spatial and verbal tasks. *CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing* 15 (2015), 3807–3816. DOI: http://dx.doi.org/10.1145/2702123.2702315
26. Horia A Maior, Matthew Pike, Max L Wilson, and Sarah Sharples. 2014. Continuous detection of workload overload: An fnirs approach. *Contemporary Ergonomics and Human Factors 2014: Proceedings of the international conference on Ergonomics & Human Factors 2014, Southampton, UK, 7-10 April 2014* (2014), 450. DOI: http://dx.doi.org/10.1201/b16742-79
27. Valerie Mendoza and David G Novick. 2005. Usability over time. In *Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information*. ACM, 151–158. DOI: http://dx.doi.org/10.1145/1085313.1085348
28. Erika Molteni, Michele Butti, Anna M Bianchi, and Gianluigi Reni. 2008. Activation of the prefrontal cortex during a visual n-back working memory task with varying memory load: a near infrared spectroscopy study. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 4024–4027. DOI: http://dx.doi.org/10.1109/IEMBS.2008.4650092

29. Jakob Nielsen and Jonathan Levy. 1994. Measuring usability: preference vs. performance. *Commun. ACM* 37, 4 (1994), 66–75. DOI:http://dx.doi.org/10.1145/175276.175282

30. Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 249–256. DOI: http://dx.doi.org/10.1145/97243.97281

31. Evan M M Peck, Beste F Yuksel, Alvitta Ottley, Robert JK Jacob, and Remco Chang. 2013. Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 473–482. DOI: http://dx.doi.org/10.1145/2470654.2470723

32. Reinhard Pekrun, Thomas Goetz, Lia M Daniels, Robert H Stupnisky, and Raymond P Perry. 2010. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology* 102, 3 (2010), 531. DOI: http://dx.doi.org/doi/10.1037/a0019243

33. Matthew Pike, Horia A Maior, Martin Porcheron, Sarah Sharples, and Max L Wilson. 2014. Measuring the Effect of Think Aloud Protocols on Workload using fNIRS. In *CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. DOI: http://dx.doi.org/10.1145/2556288.2556974

34. Gary B Reid and Thomas E Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology* 52 (1988), 185–218.

35. Sergio Sayago and Josep Blat. 2007. Some Aspects of Designing Accessible Online Forms for the Young Elderly.. In *WEBIST (2)*. 13–17.

36. Sergio Sayago, José-María Guijarro, and Josep Blat. 2012. Selective attention in web forms: an exploratory case study with older people. *Behaviour & Information Technology* 31, 2 (2012), 171–184. DOI: http://dx.doi.org/10.1080/01449291003767920

37. M.C. Schraefel, Maria Karam, and Shengdong Zhao. 2003. Listen to the Music: Audio Preview Cues for Exploration of Online Music.. In *Interact 2003 - Bringing the Bits Together*. http://eprints.soton.ac.uk/258800/ Event Dates: Sept 1-5.

38. Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*. ACM, 2651–2656. DOI: http://dx.doi.org/10.1145/1240866.1241057

39. Ben Shneiderman. 1992. *Designing the user interface: strategies for effective human-computer interaction*. Vol. 3. Addison-Wesley Reading, MA. 60–63 pages.

40. Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert JK Jacob. 2009. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 157–166. DOI: http://dx.doi.org/10.1145/1622176.1622207

41. John Stemberger, Robert S Allison, and Thomas Schnell. 2010. Thermal imaging as a way to classify cognitive workload. In *Computer and Robot Vision (CRV), 2010 Canadian Conference On*. IEEE, 231–238. DOI: http://dx.doi.org/10.1109/CRV.2010.37

42. John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285. DOI:http://dx.doi.org/10.1016/0364-0213(88)90023-7

43. Pamela S Tsang and Velma L Velazquez. 1996. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* 39, 3 (1996), 358–381.

44. John R Wilson and Sarah Sharples. 2015. *Evaluation of human work*. CRC Press. 521–522 pages.

45. Luke Wroblewski. 2008. *Web form design: filling in the blanks*. Rosenfeld Media.