Practical Evaluation of the Emergency Usability Lab for Testing the Usability of Medical Devices in Emergency Situations

Peter Rasche(✉), Moritz Richter, Katharina Schäfer, Sabine Theis, Verena Nitsch, and Alexander Mertens

Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Bergdriesch 27, 52056 Aachen, Germany
{p.rasche,m.richter,k.schaefer,s.theis,v.nitsch,
a.mertens}@iaw.rwth-aachen.de

Abstract. In healthcare, clinical activities are increasingly being relocated to patients' home environment. Patients can thus continue their therapy in a self-determined manner at home. This trend challenges usability evaluations of medical devices, as patients represent a more heterogeneous user group in terms of experience and training than medical professionals and the use of medical devices by patients can also be influenced by their medical condition. The Emergency Usability Lab was developed as an approach for testing the usability of medical devices of home healthcare focusing on emergency situations. An emergency situation is determined as a situation the patient has to perform certain activities with a medical device while being limited in performance due to certain medical symptoms as dizziness, visual impairment or being short of breath.

This paper presents the results of the practical evaluation of the Emergency Usability Lab. Evaluation was carried out applying a between-group design. Within the study, two approved medical devices (blood glucose meter & emergency drug package) were subjected to a usability test. Each product was tested in two groups. In one group the Emergency Usability Lab was applied and the other served as control group. Both groups included 20 participants each. By using the Emergency Usability Lab, a blood glucose meter was tested under simulated conditions of hypoglycemia (lack of concentration and impaired vision) and an emergency drug package was tested under simulated conditions of pain-related shortness of breath (shortness of breath, dizziness and impaired vision). The results of both usability studies were compared with regard to the number of usability issues found and their risk in order to verify the applicability of the Emergency Usability Lab.

Keywords: Emergency · Usability · Simulation · Older adults · Digital health

Q. Gao and J. Zhou (Eds.): HCII 2020, LNCS 12207, pp. 222–230, 2020.
https://doi.org/10.1007/978-3-030-50252-2_17

1 Introduction

1.1 Motivation

Since the introduction of the Human Factors Guidelines of the Food and Drug Association in the USA in 2015 and the re-regulation of the Medical Device Directives in the European Union in 2017, usability of medical devices needs to be evaluated during official approval process of medical devices. As these directives are new and healthcare is changing fast due to digitization new methods are necessary to perform the usability evaluation. Additionally, clinical activities are increasingly being relocated to patients' home environment. Patients can thus continue their therapy in a self-determined manner at home. This trend challenges usability evaluations of medical devices, as patients represent a more heterogeneous user group in terms of experience and training than medical professionals and the use of medical devices by patients can also be influenced by their medical condition. The Emergency Usability Lab (EUL) was developed as an approach for testing the usability of medical devices of home healthcare focusing on emergency situations. An emergency situation is determined as a situation the patient has to perform certain activities with a medical device while being limited in performance due to certain medical symptoms as dizziness, visual impairment or being short of breath.

1.2 Aim of This Study

The aim of this study was to investigate in a between-group design what influence the use of the EUL has on the quality of results of two usability tests (EUL applied and control group). The first evaluated product is a blood glucose meter which can be connected to the iPhone (up to model 4S) and the second is a Bend-and-Peel tablet blister according to patent US8191711, which among other things is used for the distribution of fentanyl tablets. The aim is to determine whether the application of the EUL has an influence on the number of usability issues identified during a usability evaluation as well as the severity rating of these issues. Furthermore, it should be tested whether participants' subjective mental effort and workload differs between the two groups.

2 Method

A between-group design was chosen for this study. Accordingly, the usability test of both products was carried out in one group with and in a second (control group) without the application of the EUL. In accordance with the motivation described in the introduction, the two medical devices should be tested for their respective suitability for emergency situations. In the case of the blood glucose meter, the emergency situation was defined as an episode of severe hypoglycemia [1]. In such a situation, patients may show cognitive as well as visual impairment in extreme cases (lack of concentration and blurred vision) [1]. However, in order to react correctly, it is essential to determine the blood sugar level, and thus a blood glucose measurement needs to be performed by the patients. In the context of the Bend-and-Peel tablet blisters, emergency situation was defined as a situation in which patients suffer from extreme pain, which has led to shortness of breath, lack of concentration and visual impairment. Such pain is common in the context of cancer therapy or palliative care [2].

2.1 Questionnaire

For the evaluation of the usability of the products, a 16-sided questionnaire was developed, which was used in both groups in the same manner. The questionnaire initially collected demographic data of the study participants. Subsequently, the willingness to use technology according to Neyer and the health competence of the participants was surveyed using the EU-HLS-16 [3, 4]. A paper-based eye test for colour vision and visual acuity was then carried out. The Rating Scale of Mental Effort (RSME) was used to assess the mental stress of the participants during the individual usability tasks. In addition, the raw NASA TLX was recorded for each product to capture possible additional stress factors during product use.

Measuring Technology Readiness
Technology readiness was included as it might influence the use of modern information and communication technology, as well as the engagement with these products [3]. It is calculated based on 12 standardized items which are rated on a 5-point Likert scale (1 = not correct, 5 = fully correct). For positively formulated objects, the scale is converted so that a high point value corresponds with high technology readiness. Subsequently, the final score is calculated by mean value over all 12 items. The score therefore ranges between 1 point and 5 points [3].

Measuring Health Literacy
To gain further insight into the participants' knowledge and ability to manage potential diseases, a health literacy questionnaire was applied [4, 5]. This questionnaire assesses user's access, understanding, evaluation, and application of health-related information. The questionnaire contains 16 items rated on a 4-point Likert scale ranging from 'very difficult' to 'very easy'. Afterwards results are transformed into dichotomous answers and summed up. Thus, the score ranges between 0 point and 16 points [4].

Measuring Subjective mental Effort and Workload
All usability tasks were evaluated by the RSME to measure the subjective mental effort necessary for participants to accomplish the tasks [6, 7]. The participants rated mental effort on a 150-pointscale. All participants were familiarized with the RSME scale using five appropriate daily life examples.

Additionally, raw NASA TLX tool was used to analyse participants' workload. This tool divides a subjects' workload into six different factors. These factors are 'mental demand', 'physical demand', 'time pressure', 'expended effort', 'achieved performance level' and 'experienced frustration' [8].

Accessing Users' Experiences and Criticality of Identified Problems
Semi-structured interviews were performed to investigate whether participants identified any usability issues or misleading information or functions related to the two evaluated products.

2.2 Experimental Apparatus

Tested Products

Blood glucose meter

The blood glucose meter iBG-Star manufactured by AgaMatrix Inc., Salem, USA, was tested. Previous studies had shown the high accuracy of this device in measuring blood glucose [9]. Furthermore, the authors already used this device in a previous study including diabetes patients [10]. Thus, this product was chosen to enable a comparison between results of this study and the previous one.

Bend-and-Peel tablet blister

The second product to be tested was a Bend-and-Peel tablet blister. This blister is used in combination with fentanyl tablets, which can be used for breakthrough pain in cancer therapy and palliative care. It should be noted that the blister is opened in situations where the patient, and therefore the user, suffers extreme pain. It must be correspondingly easy to open. However, since it needs to be prohibited that children could take this medication, the blister must still be child-resistant. The structure of the blister is defined by patent US8191711.

Emergency Usability Lab

The Emergency Usability Lab is a modular system that can simulate breathing difficulties, cognitive impairment or visual impairment of performance by using stimulating stressors. Depending on the clinical picture under investigation, the stressors can be combined to provide a representation of the symptoms of the disease in an emergency situation [11]. In this study stroboscopic glasses and an interrupted white noise signal were used to simulate the symptoms of hypoglycemia. In combination with the tablet packaging, the two mentioned products and additionally a breathing mask and a weight vest were used, which simulate a respiratory distress under pain.

Cogstate Performance Testing Battery

Participant's cognitive performance was accessed using the CogState testing battery (www.cogstate.com). This battery was implemented on an iPad 3 (Apple, USA). This system was chosen as it provides a higher ecological validity of the results with regard to the performance with a mobile device than if a stationary computer-based system had been used. Results of this testing battery are comparable to pen and paper tests [12]. The testing battery itself consists of four different tasks. The task order was the same during all stressor exposures. The tasks were completed in the order shown below.

- Detection task (DET) - a reaction time test, measuring psychomotor function.
- Identification task (IDN) - a choice reaction time test, measuring visual attention.
- One Back task (ONB) - a task to measure working memory and attention.
- Groton Maze learning grid (GML) - a task to measure executive function by the total number of errors made.

2.3 Data Collection and Recruitment

The data collection was carried out at the Institute for Industrial Engineering and Ergonomics of RWTH Aachen University. For the participation in this study, an expense allowance of 15€ was paid. The participants were recruited online via social media channels, which are frequently used by students of RWTH Aachen University. Participants with experience in using one of the two evaluated products are excluded from participation to gain a realistic picture of product use by non-experts.

2.4 Procedure

Prior to the experiment, participants were informed about the purpose of the study and their right to withdraw at any time during the study without any adverse consequences. All sessions took place before lunch to avoid differences in performance due to the daytime. Participants answered a short questionnaire aiming to ascertain demographic information. Afterwards, participants' performance was accessed by performing the Cogstate testing battery. Execution of associated tasks was measured by RSME as well as NASA TLX as reference for following usability tasks. Next, the participants had five minutes to familiarize themselves with the blood glucose meter. They were provided with the original manufacturer's instructions. Afterwards the participants performed a simulated blood glucose measurement. Instead of pricking themselves with a real needle, the patients used a lancet without needle accordingly and had a drop of blood substitute (glucose test fluid) dripped onto their finger. Depending on which of the two groups (with/without Emergency Usability Lab) was tested, the mentioned stressors (stroboscopic glasses and interrupted white noise) were additionally applied during the simulated blood glucose measurement. After the simulated measurement, the participants filled out the raw NASA TLX and rated mental effort by RSME. Subsequently, a semi-structured interview was conducted in which the participants were asked about usability issues and challenges regarding the blood glucose measurement. At the end of the interview, the participants were asked to evaluate the identified usability issues on a 4-point Likert scale with regard to the need for a revision by the manufacturer. After a five-minute break, the procedure was repeated using the tablet-blister. Again, the participants had five minutes to familiarize themselves with the instructions for unpacking the tablet. They were then asked to unpack one tablet. If the blister was damaged during this attempt, the participants were free to start a second attempt in the same run by unpacking another tablet. Again, the raw NASA TLX and RSME scores were collected afterwards. Also, for this product, a semi-structured interview was performed to identify usability issues and challenges and their severity in terms of need for revision by manufacturer. Afterwards the participants were paid the expense allowance and were subsequently seen off.

3 Results

In this paper, the difference in subjective mental effort and workload (RSME and NASA TLX) and the number of identified usability issues and their severity were investigated in a between-group designed study.

3.1 Participants

In total 40 participants took part in this study, divided into two groups (EUL applied and control group). Each group was gender balanced. The average age of the participants was 25.55 years (standard deviation = 4.150). All participants showed normal average values regarding the measurements of the Cogstate performance testing battery and were enrolled at a university. None of the participants had experienced one of the two emergency situations or evaluated products before. Technical readiness and health literacy were also quite similar within both groups (Table 1).

Table 1. Participants demographics

	EUL applied	Control group
Age		
Minimum (years)	19	19
Maximum (years)	40	35
Mean (SD; years)	24.95 (4.477)	26.15 (3.815)
Technical readiness		
Mean (SD; points)	2.27 (0.493)	1.92 (0.492)
Health competence		
Mean (SD; points)	12.00 (2.772)	11.85 (2.942)

3.2 Usability Evaluation of the Blood Glucose Meter

The descriptive comparison of the results shows that based on the semi-structured interviews under the stress condition, participants named more usability issues than within the control group. Compared to this severity rating of the identified issues seems quite similar. RSME scores are also higher within the stressed group than within control group. It also applies to the NASA TLX scores.

The difference between the number of usability issues reported in stress and control group for the blood glucose meter is significant ($t(38) = 3.040$, $p = 0.004$, $r = 0.442$). Severity rating of issues was not significantly different. Within the stressed group participants named more usability issues than within the control group. NASA TLX scores between the stressed and the control group are significantly different ($t(38) = 3.946$, $p < 0.001$, $r = 0.539$). Within stressed group participants reported higher NASA TLX scores than within the control group. The difference between RSME scores for the blood glucose measurement is significant with a medium effect strength ($t(38) = 3.933$, $p < 0.001$, $r = 0.538$) (Table 2).

Table 2. Usability evaluation results within EUL and control group for blood glucose meter

	EUL applied	Control group
Problems/Challenges mentioned		
Mean number (SD; points)	3.00 (1.486)	1.80 (0.951)
Mean severity (SD; points)	2.44 (0.832)	2.74 (0.890)
NASA TLX		
Mean (SD; points)	49.33 (14.827)	32.04 (12.817)
RSME		
Mean (SD; points)	59.50 (27.285)	30.75 (18.011)

3.3 Usability Evaluation of the Bend-and-Peel Tablet Blister

For the Bend-and-Peel blister participants within stressed and control group did not report a different number of problems regarding the product. Compared to this severity rating of the identified problems seems quite similar. Scores for RSME did not differ for the Bend-and-Peel tablet blister between stressed and control group. Same observations were made for NASA TLX evaluation.

Although reported NASA TLX scores are higher within the applied EUL group, no significant difference was revealed ($t(38) = 1.532$, $p = 0.134$, $r = 0.241$). The difference between RSME scores for the Bend-and-Peel tablet blister is not significant ($t(38) = 1.950$, $p = 0.059$, $r = 0.302$) (Table 3).

Table 3. Usability evaluation results within EUL and control group for bend-and-peel blister

	EUL applied	Control Group
Problems/Challenges mentioned		
Mean number (SD; points)	2.0 (0.973)	1.95 (1.234)
Mean severity (SD; points)	2.50 (0.892)	2.78 (0.820)
NASA TLX		
Mean (SD; points)	44.75 (17.619)	36.45 (16.585)
RSME		
Mean (SD; points)	43.60 (29.491)	26.20 (26.889)

4 Discussion

For the first time, results of this study show that the Emergency Usability Lab can be applied fundamentally within the framework of usability tests and represents a benefit. It was shown that the stressors applied in the laboratory do not restrict the participants to such an extent that they would not be able to solve the tasks assigned within the usability test. Results show that the application of stressors results in significantly higher number of usability problems in relation to more complex medical devices. In this study, a blood glucose meter was chosen as a complex medical device. As a representative of the group of "simpler" medical devices, a bend-and-peel blister packaging was chosen. While the usability task for the blood glucose meter took about one minute to complete, the usability task for the bend-and-peel blister packaging was completed in about 15-30 s. Possibly the differences mentioned are due to the respective exposure time to the stressors or the complexity of the product. Further studies focusing these aspects are necessary to investigate these circumstances in more detail. Moreover, the study showed that the application of the stressors leads to a higher subjective stress level for the same usability task measured by the RSME scale and the NASA TLX.

Due to the study design, various limitations have to be mentioned. The stressors used are products which are used for training purposes in competitive sports. Although it is known from previous studies by the authors that the stressors cause stress and a limitation of the performance of study participants, the use of these products to generate stress must nevertheless be questioned. Furthermore, only two medical devices were examined within the scope of this study. Therefore, it cannot be conclusively clarified whether the results regarding the identification of usability problems with complex medical devices can be attributed to the complexity of these devices, the application time of the stressors or to unknown reasons.

This fundamentally demonstrated that the Emergency Usability Lab provides a useful contribution to the evaluation of the usability of products for use in emergency situations. Future analysis of the collected data will yield further results regarding the practical use of the Emergency Usability Lab.

References

1. Carroll, M.F., Burge, M.R., Schade, D.S.: Severe hypoglycemia in adults. Rev. Endocr. Metabol. Disord. **4**(2), 149–157 (2003)
2. Bausewein, C., Simon, S.T.: Shortness of breath and cough in patients in palliative care. Deutsches Ärzteblatt Int. **110**(33–34), 563 (2013)
3. Neyer, F.J., Felber, J., Gebhardt, C.: Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft. Diagnostica **58**(2), 87–99 (2012). https://doi.org/10.1026/0012-1924/a000067
4. RoÈ thlin, F., Pelikan, JM., Ganahl, K.: Die Gesundheitskompetenz der 15-jaÈhrigen Jugendlichen in Oesterreich. In: Abschlussbericht der oÈsterreichischen Gesundheitskompetenz Jugendstudie im Auftrag des Hauptverbands der oÈsterreichischen SozialversicherungstraÈger (HVSV). http://lbihpr.lbg.ac.at.w8.netz-werk.com/sites/files/lbihpr/attachments/hljugend_bericht.pdf. Accessed: 2017–09-19. (Archived by WebCite® at http://www.webcitation.org/6taVAKBVo). Accessed 19 Sep 2017

5. Hls-Eu Consortium. Comparative Report of Health Literacy in Eight EU Member States. In: The European Health Literacy Survey HLS-EU (2012)
6. Zijlstra, F.: Efficiency in work behaviour. In: A Design Approach for Modern Tools. In: Ph.D. thesis, TU Delft, Delft University of Technology. 1993. Available: http://repository.tudelft.nl/assets/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f/ide_zijlstra_19931123.PDF
7. Otto, T., Zijlstra, F.R.H., Goebel, R.: Neural correlates of mental effort evaluation-involvement of structures related to self-awareness. Soc. Cogn. Affect Neurosci. **9**(3), 307–315 (2014). https://doi.org/10.1093/scan/nss136
8. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index). Results of empirical and theoretical research. Hum. Mental Workload **1**(3), 139–183 (1988)
9. Pfützner, A., Mitri, M., Musholt, P.B., Sachsenheimer, D., Borchert, M., Yap, A., et al.: Clinical assessment of the accuracy of blood glucose measurement devices. Curr. Med. Res. Opin. **28**(4), 525–531 (2012). https://doi.org/10.1185/03007995.2012.673479. PMID: 22435798
10. Rasche, P., Mertens, A., Miron-Shatz, T., Berzon, C., Schlick, C.M., Jahn, M., et al.: Seamless recording of glucometer measurements among older experienced diabetic patients ±A study of perception and usability. PLoS ONE **13**(5), e0197455 (2018). https://doi.org/10.1371/journal.pone.0197455
11. Rasche, P., Mertens, A., Schlick, Christopher M.: Emergency usability lab - concept to evaluate the usability of healthcare systems in emergencies. In: Duffy, Vincent G. (ed.) DHM 2017. LNCS, vol. 10287, pp. 354–364. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58466-9_32
12. Benoit, A., Malla, A.K., Iyer, S.N., Joober, R., Bherer, L., Lepage, M.: Cognitive deficits characterization using the CogState research battery in first-episode psychosis patients. Schizophr Res. Cogn. **2**(3), 140–145 (2015). https://doi.org/10.1016/j.scog.2015.03.006