



HAL
open science

Usability Evaluation of Information Technology in Disaster and Emergency Management

Tilo Mentler, Henrik Berndt, Daniel Wessel, Michael Herczeg

► **To cite this version:**

Tilo Mentler, Henrik Berndt, Daniel Wessel, Michael Herczeg. Usability Evaluation of Information Technology in Disaster and Emergency Management. 1st International Conference on Information Technology in Disaster Risk Reduction (ITDRR), Nov 2016, Sofia, Bulgaria. pp.46-60, 10.1007/978-3-319-68486-4_5 . hal-03213119

HAL Id: hal-03213119

<https://inria.hal.science/hal-03213119v1>

Submitted on 30 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Usability Evaluation of Information Technology in Disaster and Emergency Management

Tilo Mentler, Henrik Berndt, Daniel Wessel, and Michael Herczeg

Institute for Multimedia and Interactive Systems, University of Luebeck, Luebeck, Germany
mentler@imis.uni-luebeck.de

Abstract. Apart from technical reliability, usability is one of the major criteria for safe and efficient usage of interactive information technology in disaster and emergency management. However, in this setting, usability evaluation is difficult due to the heterogeneity and unpredictability of operation conditions, as well as the difficult, usually mobile, context. However, there are ways to conduct usability evaluations in disaster and emergency settings. Thus, in this paper, advantages and disadvantages of empirical and analytical usability evaluation methods for interactive systems in disaster and emergency management are discussed. The importance of formative evaluation measures within an iterative human-centered design process is emphasized. It is illustrated by two case studies dealing with paramedics' and emergency physicians' usage of mobile and wearable devices in mass casualty incidents.

Keywords: Usability, Evaluation, Usability Engineering, Human-Centered Design, Disaster Management, Emergency Management.

1 Introduction

Representative for many researchers and practitioners, Meissner et al. [1] postulate “a tremendous potential for increasing efficiency and effectiveness in coping with a disaster” by supporting disaster relief forces, aid workers and general public with information technology. However, availability and technical reliability of interactive systems alone will not lead to improved workflows or more pervasive information flows. As any socio-technical system, disaster and emergency management “relies both on human and technical function carriers” [2]. It is characterized by the “reciprocal interrelationship between humans and machines” [2]. Thoughtful human-machine task allocation and user interface design are of utmost importance.

For more than 30 years, usability has been the major criterion for assessing human-computer interaction. It is defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [3]. Several methods and process models for systematic usability have been developed. Apart from involving users early and designing iteratively, evaluating usability at different stages and with defined measures is one of the basic principles of usability engineering [4-6].

However, within the context of disaster and emergency management gaining “credible feedback on how well [...] design is working and to how to improve it” [6] is challenging for at least two reasons:

- Time- and safety-critical characteristics of disasters and emergencies put high requirements on users and technology. Even early prototypes might require advanced levels of hardware and software quality.
- Mobile contexts of use allow for unique and often unpredictable operational conditions (e.g. weather, lighting, safety hazards, connectivity). Many influences should be considered but can neither be created intentionally in laboratories nor in the field.

For example, Leitner et al. [7] explain their decision to test a mobile emergency information system under laboratory conditions instead on-site as follows: “Firstly, we didn’t want to risk the test being interrupted when an emergency call arrives, resulting in incomplete data. Secondly, we had to consider the ethical aspect that instead of relaxing between emergency cases, the physician would have to participate in our test. This could have influenced his or her performance in a real emergency case. Other reasons were technical ones. For a realistic test it would have been necessary to establish a working wireless infrastructure in the car, which would have tied up too many resources of the development team necessary for other tasks in the project.”

In any case, usability must be considered while designing interactive systems for users acting and making decisions under the influence of various stressors. This topic will be addressed by describing background and related work in section 2. Several usability evaluation approaches will be compared with respect to the context of disaster and emergency management (see section 3). In section 4, two case studies about the human-centered design and evaluation of mobile and wearable interactive systems for members of Emergency Medical Services (EMS) in mass casualty incidents (MCIs) will be presented. Conclusions will be drawn in section 5.

2 Background and Related Work

As mentioned before, usability evaluations should be part of systematic approaches to usability rather than a single quality measure at the end of a development project. Therefore, basic principles of usability engineering will be elaborated in section 2.1. Following this, usability evaluation methods will be distinguished with respect to time of conduction and reasoning approaches in section 2.2. Finally, related work to usability evaluation in disaster and emergency management will be outlined.

2.1 Usability Engineering

Usability engineering processes refer to “concepts and techniques for planning, achieving, and verifying objectives for system usability” [8]. They embrace but are certainly not limited to user interface and interaction design. For different stages of development, several methods have been developed or derived from humanities, social or engineering sciences [5]:

- Observing and interviewing users in the field as well as formal task analysis and user research methods in order to understand the context of use.
- Writing scenarios or sketching storyboards as well as working with different kinds of low-fidelity and high-fidelity prototypes (e.g. wireframes, mockups, paper prototypes) as an aid to envision design solutions.
- Testing the suitability of design solutions with respect to users, tasks and operating conditions in the laboratory and in the field (see section 2.2).

Sophisticated process models like Contextual Design [8], Scenario-Based Design [9], the Usability Engineering Lifecycle by Mayhew [10] or the Usability Engineering Lifecycle by Nielsen [11] describe steps and methods to follow in great detail.

2.2 Usability Evaluation

Independently of the process used, there is no guarantee for a certain degree of usability. Therefore intermediate results and the final product have to be tested in order to identify usage problems and room for improvement [12]. Suitable approaches depend, among others, on the

- fidelity of prototype (from paper prototype to fully functional interactive system);
- stage of development (from proof of concept to final test);
- available resources (e.g. budget, time, access to participants, equipment);
- expertise of evaluators (incl. skill in applying different techniques).

Evaluation Approaches. In general, ex ante, formative and summative evaluations can be differentiated. Additionally, we highlight the need for continuous evaluations.

Ex ante evaluations are usually done prior to an implementation to evaluate the concept (e.g. regarding feasibility or costs vs. benefits). They are focused on the current situation and potentials for improvement [13-14]. Given the high costs and risks of developments for disaster and emergency management, interventions should be planned well and address the issues with the greatest potential for improvement. Ex ante evaluations can help to understand and determine the usage context and specify the requirements for new technology. Given the need of evaluations to have clear goals or comparison standards to objectively determine the quality of a solution, results of this primary assessment can be useful for formative and summative evaluations. Well-formulated and stated standards are crucial in disaster and emergency management because standards might differ between countries and even within different states of a country.

Formative evaluations are conducted several times during the development in order to gain suggestions for improvement. They are a crucial feedback mechanism to ensure the development stays on track and is optimized for the future users, not the developers. It is usually an interactive process with a focus on qualitative data that is helpful for the developers to come up with, implement and justify design decisions [15-17].

A summative evaluation is conducted once at the final stages of development. It is used to evaluate the overall success of the development. Objectivity, impartiality and

independence of the evaluator are crucial as the results may be addressed by politicians, the public or funding agencies [15-17].

Combining these approaches is recommended in all of the previously mentioned usability engineering processes. Furthermore, some of them consider user feedback after installation respectively placing a product in the market (e.g. [11]).

Valuable insights regarding usability can be gained by getting feedback from users after several weeks, months or years of using an interactive system in practice (continuing evaluation). These observations can highlight issues not apparent in the usually short-term summative evaluation or reveal the effects of changing conditions (e.g. new standards or technologies are introduced), a changing user base (e.g. generational changes), or new threats (e.g. previously never encountered disasters).

Observing the actual usage in the field over long time periods should be planned during development and implemented. If done well, the results can be used for continuous improvement. Log files can be very useful as they unobtrusively gather actual usage data. Anonymously recorded with the user's knowledge and consent, they can help to identify features or function rarely used. Following this, developers and users could decide either to remove them or to improve their accessibility in the following version or via a short-term update. To go beyond the currently available features, users should be able to express usability problems without any great effort, e.g. via an implemented feedback function.

Interviews, workshops and (online) surveys can also be conducted but they have the disadvantage that users afterwards would have to remember interaction problems or other critical situations that occurred. During disasters or emergencies, they will not even be able to take notes about them. Therefore, it is unlikely that feedback would contain a great deal of actual improvement possibilities. However, log files could be used as memory aids to determine usability problems.

Reasoning Approaches. With respect to reasoning approaches, analytic methods based on models, guidelines or experts' reviews can be distinguished from empirical methods involving tests with actual users or appropriate representatives. While the first ones are often applied at early stages of usability engineering processes, the second ones are usually utilized at advanced stages. Because there is no "silver bullet" to usability evaluations, analytic and empiric methods are often combined in iterative design processes. Guiding empirical tests with results of analytic approaches was named *mediated evaluation* in the domain of education [18]. This term has been adopted by some researchers in the field of human-computer interaction and usability engineering [8].

Evaluation Designs. Data from evaluations have to be seen in context, so the research design of the evaluation matters as well. An *experiment* (i.e. randomly assigning people to either an experimental or a control group while controlling for confounding variables (cf. [19]) allows for determining cause-and-effect relationships. However, this requires a large amount of participants (roughly 30 per condition) and a high degree of control over the setting (see section 3.2).

Quasi-experiments (variation, e.g. two conditions but no control of confounding variables) are more common. A typical example is the implementation of the solution in one workplace and comparing it with the performance in another one without that solution. However, the evidence is weaker as they do not allow casual inferences (there might be other reasons for the performance difference). Correlation studies are easier to conduct, but correlation does not imply causality so the values of conclusions are limited.

At the very least, some context should – and usually can – be provided, either with data from a previous/alternative solution or the current gold standard, or compared to earlier tests (pre-post-tests). While this data might not allow for casual inference, it provides at least some reference points for interpretation. Practical constraints usually determine what is possible, and even if the best design is not possible, the best possible design should be chosen.

2.3 Usability Evaluation in Disaster and Emergency Management

As mentioned before, technical reliability of systems in disasters and emergencies alone would not lead to better outcomes because “when the user interaction with a safety-critical system goes wrong, the result can be catastrophic” [20]. The importance of usability evaluations has not only been recognized by single research groups (e.g. [21-23]) but also by principal organizations dealing with disaster and emergency management. The Sendai Framework for Disaster Risk Reduction 2015-2030 is “the first major agreement” [24] with respect to disaster risk management and has been endorsed by the General Assembly of the United Nations. Although it does not name the term usability (evaluation), several statements in this matter can be found in the document:

- “...taking into account the needs of different categories of users...” [25, p. 14]
- “To support the development of local, national, regional and global user-friendly systems and services for the exchange of information on good practices, cost-effective and easy-to-use disaster risk reduction technologies and lessons learned on policies, plans and measures for disaster risk reduction.” [25, p.16]
- “...develop such systems through a participatory process; tailor them to the needs of users, including social and cultural requirements...” [25, p. 21]

The United Nations Office for Disaster Risk Reduction (UNISDR) adds that “it is imperative to improve the usability of such services [for information dissemination in disasters] by strengthening technological infrastructure in all locations and providing information in a clear and concise way” [26, p. 88].

These statements show that usability of interactive systems is an important aspect of disaster and emergency management. However, terms like “user-friendly systems” and “easy-to-use” have long been challenged by usability and human-computer interaction experts, because they are either “unnecessarily anthropomorphic” [10, p. 23] and can hardly be operationalized [27].

3 Assessing Usability Evaluation Methods and Settings

In the following sections, several analytic and empirical usability evaluation methods and settings will be assessed in their suitability for disaster and emergency management evaluations. They have been selected because they have proven to work in several domains with safety-critical characteristics.

3.1 Methods

Heuristic evaluations, model-based approaches, and cognitive walkthroughs will be outlined as usability evaluation methods in the following sections.

Heuristic Evaluation. In the context of usability evaluations, heuristics address “some basic characteristics of usable interfaces” [10, p. 115] like “Speak the users’ language: The dialogue should be expressed clearly in words, phrases, and concept familiar to the user, rather than in system-oriented way” [10, p. 20].

A systematic and individual analysis of an interactive system according to such principles by one or more usability experts is called heuristic evaluation [8, 10]. Well-known and often applied sets of heuristics are Nielsen’s 10 Usability Heuristics [10] and Shneiderman’s 8 Golden Rules of Interface Design [28].

Because of their context independence, such heuristics cannot be used to evaluate domain-specific and safety-oriented aspects of interactive systems for disaster and emergency management. Furthermore, some recommendations might conflict and products might have different requirements that make it necessary to violate generic heuristics. However, violating a guideline or heuristic should be done deliberately and for good reasons. In general, if applications fail such basic principles of user interface and interaction design, they will hardly be usable in a safe and efficient way, especially under extraordinary circumstances. Therefore, heuristic evaluations by usability experts are an important and helpful method for formative evaluations. They can easily be complemented but not completely replaced by domain-specific inspections by disaster and emergency management experts. If possible, decisions based on guidelines should be tested via empirical data, e.g. in formative or summative evaluations.

Model-based Approaches. In order to predict and analyze courses of human-computer interaction, task-oriented modeling approaches were developed [5,12]. They have been applied to safety-critical domains like healthcare [29].

Methods like GOMS (Goals, Operators, Methods, Selection Rules) or KLM (Key-stroke-Level Method) are among the most commonly used model-based approaches [12, 28]. In this regard

- Goals are “simply the user’s goals, as defined in layman’s language” [30, p. 81];
- Operators are “the actions that the software allows the user to take” [30, p. 81];
- Methods are “well-learned sequences of sub-goals and operators that can accomplish a goal” [30, p. 81];

- Selection Rules are “the personal rules that users follow in deciding what method to use in a particular circumstance” [30, p. 81f].

By predicting the time to complete (TTC) tasks for single operators, overall performance can be estimated. In the case of disaster and emergency management, these estimations might be less precise than in other domains because GOMS specifically “applies to situations in which users will be expected to perform tasks that they have already mastered” [28, p. 84].

Therefore, model-based usability evaluation should be regarded as a valuable complement to other approaches and limited to the most time-critical parts of user interfaces for professional operators.

Cognitive Walkthrough. Cognitive Walkthrough and its variations like Cognitive Jogthrough are other analytic approaches to step-wise analysis of human-computer interaction by system designers and usability experts. It is based on the CE+ theory of exploratory learning [5, 31].

Apart from understanding important characteristics of future users, regular and extraordinary usage scenarios as well as defined sequences for efficient task completion are required [5]. Based on them, an analyst performs each interaction step individually and answers questions whether users would

- try to solve the problem in the right way;
- notice that an appropriate function or feature is available;
- associate available actions with desired effects;
- recognize the progress they made towards their goal [5].

Cognitive Walkthroughs have been applied to safety-critical domains, e.g. air traffic control [32]. With regard to computer-supported cooperative work, some modifications were found to be necessary in order to deal with individual and group tasks [33].

Therefore, cognitive walkthroughs might be especially helpful to improve applications for single non-professional users or individual operators of emergency and rescue services. Cooperation and team aspects characterizing disaster and emergency management to a certain degree could hardly be assessed with this approach.

3.2 Settings

For settings, the typical division of laboratory and field settings will be followed. Each of them will be assessed for its feasibility.

Usability Tests in the Laboratory. In the context of usability engineering, laboratory studies are characterized by users testing an interactive system in an artificial environment intended for evaluations (i.e. not in the users' workplace). The examiner's control of the setting is used to ensure stable testing conditions and to allow the use of standardized benchmarks [5, 8]. Laboratory environments usually allow for detailed

unobtrusive data gathering, e.g. when audio and video recording equipment is hidden from view.

The main advantage of laboratory studies is high internal validity. Different users can use the product under the same circumstances, reducing the influence of confounding factors in the environment (e.g. disturbances, variance in environmental conditions, etc.). Given the high degree of control regarding confounding variables, laboratory studies are well-suited for experiments. This way, the improved version can be tested against the current version to determine whether the improved version is actually better.

However, laboratory studies suffer from limited external like ecological validity. The effects determined in the controlled situation in the laboratory might not carry over to actual use in the field. This is the case, if important aspects of the actual use (e.g. stress, environmental factors) were not replicated in the laboratory. For example, a solution supporting voice input and optical head-mounted display output, which is usable under well-defined lighting conditions and low noise level, might be useless while standing next to a highway in the rain under high traffic. This is frequently a problem with experiments — the usual focus on depth (few variables and a high degree of control) is at odds with the breadth of different conditions in disaster and emergency settings.

In summary, usability tests in the laboratory are an important empirical approach to measuring usability of disaster and emergency management systems, but have to consider the critical aspects of future usage scenarios. For example, both noise levels and lighting conditions can be modified with the aid of audio equipment or dimmers. For mobile and wearable devices, users should be requested to stand up or walk around. If necessary, obstacle courses have to be built.

Usability Tests in the Field. As opposed to laboratory studies, field studies are characterized by testing an interactive system in the users' familiar work environment. Preparing, conducting and analyzing results of usability tests in the field is challenging for several reasons:

- Access to critical areas, especially in safety-critical domains, might only be possible to a limited extent or not at all.
- Working conditions can hardly be controlled by examiners, e.g. disruptions by colleagues or environmental conditions.
- Data recording is more difficult than in laboratories. Retrospective interviews might be necessary in order to make sense of certain usage situations.

However, field studies can deliver valuable insights that are hard to get otherwise. The external, e.g. ecological validity is high, although at the cost of internal validity. While field experiments are possible, they typically require large amounts of resources. In general, field studies are useful in summative evaluations. Results of field studies can also be used to improve laboratory studies, e.g. by replicating important environmental factors.

4 Case Studies

In the following sections, two case studies conducted in close collaboration with Emergency Medical Services are outlined. Special respect is given to the applied formative and summative usability evaluation measures.

4.1 Supporting Collaboration in MCIs with tablet PCs

Mass Casualty Incidents (MCIs) are characterized by “more patients at one time than locally available resources can manage using routine procedures” [34]. Dozens or hundreds of paramedics and emergency physicians have to collaborate in order to treat patients in the best possible way and apply resources efficiently. Incident commanders are in charge of task prioritization and personnel management.

Currently, paper-based artifacts (e.g. forms, tables, charts, maps) and various means of communication (e.g. radio, mobile phone, messengers, face-to-face conversations) are used for record keeping and information management. Situation awareness of incident commanders in MCIs is a demanding challenge [35].

Mobile computer-based tools might help to improve the situation because data can be exchanged and updated within a narrow time frame independent of the incident commanders' locations. Scalable visualizations (e.g. filter functions, summary vs. detail view) would allow for solving different information needs. Problems because of poor handwriting could be prevented. However, designing such solutions is both a technical and a usability challenge. Incident commanders and other rescue forces are under high physical and mental stress due to the extraordinary circumstances.

Within a two-year user-centered system design project, we developed a prototype of a tablet-based data gathering and information management system in close collaboration with German Emergency Medical Services (cf. [36-38] for more details).

Apart from regular interviews, focus groups and workshops with members of Emergency Medical Services, several specific formative and summative usability methods were applied. They are described subsequently.

First of all, prototypes and single interaction elements were reviewed by usability experts a number of times. The well-established dialogue principles of ISO standard 9241-110 served as heuristics (see section 3.1). For example, error tolerance as one of these dialogue principles was considered by proposing a button-oriented design (see Fig. 1) “offering a choice of valid input values, e.g. for triage categories, diagnoses, drug names, doses or feed rates” [39], “favoring filtering over searching functions” [39] and “searching phonetically when needed” [39].

Furthermore, alternative designs were presented to members of EMS. They were invited to choose their preferred version or to make proposals for better solutions. For example, four different drafts of a bar chart visualizing the number of patients in a specific triage category were discussed by 36 paramedics and physicians (see Fig. 2).

While 5 participants came up with own solutions, the 3 versions preferred most got 9, 6 and 5 votes, respectively. Disregarding minor differences, e.g. exact label positions, one design approach got 14 votes. The drafts were subject of a controversial debate

which “enabled us to better match our conceptual model with their mental model beyond the use cases associated with the actual chart” [35].



Fig. 1. Button-oriented design offering a choice of valid input values

The advanced prototype was shown to visitors of the leading German professional emergency medicine and rescue fair “akut 2012” at an own booth for two days (see Fig. 3). Although this is not a standard form of formative evaluation, we gained both important feedback from various experts as well as confirmations for our basic approaches [40].

Summative evaluation was performed during an exercise with an EMS and 40 virtual patients represented by cards. “Staging, triage, treatment, transport and assembly areas as well as an emergency control room were in place and equipped with tablet PCs” [40]. Participants were observed by 10 persons (members of the research team and volunteers) and asked to answer the validated ISONORM questionnaire after finishing the test. Following this, debriefing with all parties involved was conducted. Due to some network connectivity problems and minor usability flaws, results support the basic design approach. Based on these findings, we continued our research about usability of mobile and wearable devices in safety-critical domains using the example of EMS.

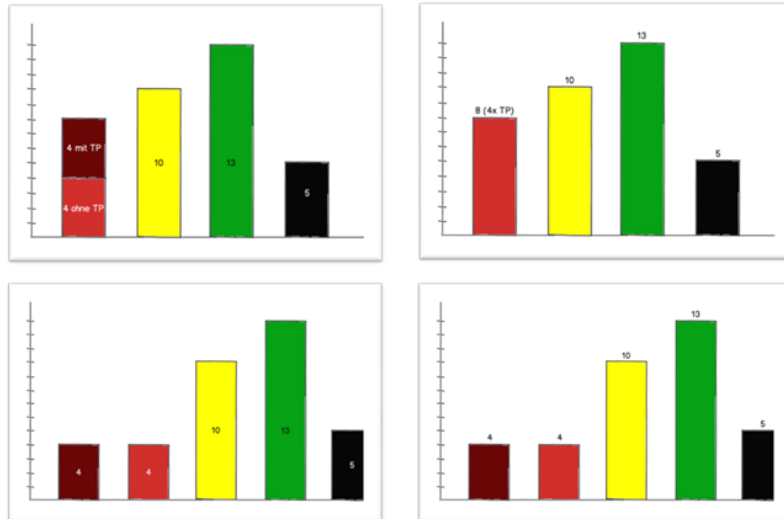


Fig. 2. Alternative design of a bar chart for triage categories [34]



Fig. 3. Discussing the prototype at the emergency and rescue fair "akut 2012"(left)

4.2 Supporting Triage and Hazard Identification with Smart glasses

In a half-year case study we dealt with the question whether smart glasses (in our case Google Glass) could be usable for supporting members of EMS. We focused on the triage process as it turned out to be one of the most important and challenging tasks in MCIs. Streger states that "most incidents are won or lost in the first 10 minutes after arrival" [41] and "that the first task that needs to be accomplished is triage of casualties" [41]. Triage shall ensure the best possible outcome for the casualties by assigning a treatment priority for each of them. Since it has to be done as fast as possible, it is

usually conducted by a member of one of the first arriving ambulances. Because MCIs are rare events for individual EMS members, most of them have only little experience in this matter.

The decision for starting research on smart glasses resulted from potential advantages compared to mobile devices (e.g. tablets), e.g. hands-free operation or the display in the user's sight of view. Carezzo et al. [42] and Cicero et al. [43] confirm this assumption. Google Glass supports voice recognition and touch gestures on the right side piece as primary interaction forms. Both of them were considered [44].

A user-centered design process was chosen. Experts from EMS and disaster relief units participated throughout all stages of development. First of all, semi-structured interviews were conducted with ten domain experts. Basic requirements as well as challenges for system design were derived. For example, the usage of algorithms was recommended. There was broad consensus that members of EMS shall not only have a command of the algorithms but also use them mandatory in the triage process [45]. Nevertheless, algorithms for triage in MCIs might not be remembered completely by rescue workers because of their mental load in extraordinary and rare circumstances. A computer-based solution not only could replace paper-based versions but gather data for documentation purposes and exchange it with incident commanders [46].

Before coding, mockups were designed and evaluated with usability and domain experts. Subsequently, two versions of the application with different interface and interaction design were implemented. One of them had to be discarded quickly because of technological flaws and not being consistent with design principles of Google. The other one was implemented iteratively (see Fig. 4).



Fig. 4. Instructions and questions are shown to the user interacting via voice commands.

The application was evaluated by 14 domain experts [46]. Because one of them saw the display distorted, only 13 EMS test persons could be considered. Participants were given four case descriptions and asked to perform triage. They were observed and notes were taken about usability problems. 11 of 13 users were able to solve all tasks without problems after a short learning phase. 2 participants were uncertain about system's state and further actions even at the last test case [46]. 8 participants called the triage support system "*useful*" [46] and "*helpful*" [46].

In addition, the usage scenario of two EMS members working together while one of them is using smart glasses was simulated for demonstration purposes (see Fig. 5). Applying a pressure dressing or performing other routine tasks seems to be possible while interacting with the wearable device. However, further research and usability evaluations are necessary.



Fig. 5. Triage of a casualty with a heavy bleeding. The user wearing smartglasses helps applying a pressure dressing [46].

5 Conclusions

Usability evaluations of disaster and emergency management systems are both crucial and challenging. Focus should be laid on ex ante, formative and continuing evaluation approaches utilizing heuristic evaluations, cognitive walkthroughs, log files and integrated feedback functions.

Usability tests in the laboratory are characterized by a high degree of control regarding confounding variables. When performing them for disaster and emergency settings, critical aspects of future usage scenarios have to be considered. Modifying noise levels, lighting conditions or interrupting users are quite easy measures to derive a certain degree of realism.

On the contrary, usability tests in the field require large amounts of resources. While they can provide insights hard to get otherwise, examiners have to judge the prototype's level of quality, the available resources and their expertise critically before preparing such a test. For example, several persons might be necessary in order to observe different areas of operation and responsible users belonging to just one public authority. Conducting data, analyzing results and drawing appropriate consequences are major challenges under these circumstances. Cost-benefit ratio should be kept in mind.

References

1. Meissner, A., Luckenbach, T., Risse, T., Kirste, T., Kirchner, H.: Design Challenges for an Integrated Disaster Management Communication and Information System. In: First IEEE Workshop on Disaster Recovery Networks (DIREN) (2002).
2. Ropohl, G.: Philosophy of Socio-Technical Systems. In: Agazzi E, Lenk H (eds) Society for Philosophy and Technology (1999). http://scholar.lib.vt.edu/ejournals/SPT/v4_n3html/ROPOHL.html
3. ISO (1998): Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.
4. Good, M., Spine, T.M., Whitesid, J., George, P.: User-derived impact analysis as a tool for usability engineering. In: Mantei M, Orbeton P (eds) CHI '86 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, N.Y., pp 241–246 (1986).
5. Benyon, D. Designing interactive systems. A comprehensive guide to HCI and interaction design. Addison Wesley, Harlow (2010).
6. Butler, K.A.: Usability engineering turns 10. Interactions 1, pp. 58–75 (1996).
7. Leitner, G., Ahlström, D., Hitz, M.: Usability of Mobile Computing in Emergency Response Systems – Lessons Learned and Future Directions. In: Holzinger A (ed) HCI and usability for medicine and health care. Third Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2007, Springer, Berlin, pp 241-254 (2007).
8. Beyer, H., Holtzblatt, K.: Contextual design. Defining customer-centered systems. Morgan Kaufmann, San Francisco (1998).
9. Rosson, M.B., Carroll, J.M.: Usability engineering. Scenario-based development of human-computer interaction. Academic Press, San Francisco (2002).
10. Mayhew, D.J.: The usability engineering lifecycle. A practitioner's handbook for user interface design. Morgan Kaufmann Publishers, San Francisco (1999).
11. Nielsen, J.: Usability engineering. Kaufmann, Amsterdam (1993).
12. Herczeg, M.: Software-Ergonomie. Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme. Oldenbourg Wissenschaftsverlag, München, (2009).
13. CORE Consultancy Research Evaluation (n.d.). http://www.evaluation.co.at/evaluation_konzept.htm
14. European Commission: Ex Ante Evaluation. A practical guide for preparing proposals for expenditure programmes. (2001). http://ec.europa.eu/smart-regulation/evaluation/docs/ex_ante_guide_2001_en.pdf
15. Wottawa, H., Thierau, H.: Lehrbuch Evaluation. Verlag Hans Huber, Bern (1998).
16. Sarodnick, F., Brau, H.: Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. Verlag Hans Huber, Bern (2006).
17. Sauro, J., Lewis, J.R.: Quantifying the user experience. Practical statistics for user research. Elsevier/Morgan Kaufmann, Amsterdam (2012).
18. Huber, O.: Das psychologische Experiment: Eine Einführung. Verlag Hans Huber, Bern (1987).
19. Scriven, M.: The methodology of evaluation. Purdue University, Lafayette (1966).
20. Redmill, F., Rajan, J.: Human factors in safety-critical systems. Butterworth-Heinemann, Oxford (1997).
21. Fischer, H., Klompaker, F.: Enriching Disaster Control Management based on Human-Centered Design. In: Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM. ISCRAM, Vancouver (2012).

22. Yang, D., Zhang, D., Frank, K., Robertson, P., Jennings, E., Roddy, M., Lichtenstern, M.: Providing real-time assistance in disaster relief by leveraging crowdsourcing power. *Personal and Ubiquitous Computing* 8, 2025–2034 (2014).
23. Li, J., Wilson, L., Stapleton, S., Cregan, P.: Design of an advanced telemedicine system for emergency care. In: Robertson T (ed) *OZCHI '06 Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, pp. 413–416 (2006).
24. United Nations Office for Disaster Risk Reduction. <http://www.unisdr.org/we/coordinate/sendai-framework>
25. United Nations Office for Disaster Risk Reduction: The Sendai Framework for Disaster Risk Reduction 2015-2030. http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf (2015).
26. United Nations Office for Disaster Risk Reduction: Progress and Challenges in Disaster Risk Reduction. http://www.unisdr.org/files/40967_40967progressandchallenges_indisaste.pdf (2014).
27. Cooper, A., Reimann, R., Cronin, D.: *About face 3. The essentials of interaction design*. Wiley, Indianapolis (2007).
28. Shneiderman, B.: *Designing the user interface. Strategies for effective human-computer interaction*. Addison-Wesley, Indianapolis (1998).
29. Magrabi, F.: Using cognitive models to evaluate safety-critical interfaces in healthcare. In: *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, New York, pp 3567-3572 (2008)
30. Bonnie, J.: Why GOMS?. *Interactions*, vol.4, pp. 80-89 (1995).
31. Rieman, J., Franzke, M., Redmiles, D.: Usability evaluation with the cognitive walkthrough. In: *Conference Companion on Human Factors in Computing Systems (CHI '95)*. ACM, New York, pp 387-388 (1995).
32. Novick, D.G.: Using the cognitive walkthrough for operating procedures. *Interactions*, vol.3, pp. 31-37 (1999).
33. Ereback, A.L., Höök, K.: Using cognitive walkthrough for evaluating a CSCW application. In: *Conference Companion on Human Factors in Computing Systems (CHI '94)*, ACM, New York, pp 91-92 (1994).
34. World Health Organization: *Mass casualty management systems - strategies and guidelines for building health sector capacity*. (2007). http://www.who.int/hac/techguidance/MCM_guidelines_inside_final.pdf
35. Mentler, T., Herczeg, M.: Interactive cognitive artifacts for enhancing situation awareness of incident commanders in mass casualty incidents. *Journal of Interaction Science*, vol.3 (2015).
36. Kindsmüller, M.C., Mentler, T., Herczeg, M., Rumland, T.: Care & Prepare - Usability Engineering for Mass Casualty Incidents. In: *ACM EICS4Med 2011: Proceedings of the 1st International Workshop on Engineering Interactive Computing Systems for Medicine and Health Care*, ACM, New York, pp 30-35 (2011).
37. Mentler, T., Herczeg, M., Jent, S., Stoislow, M., Kindsmüller, M.C.: Routine Mobile Applications for Emergency Medical Services in Mass Casualty Incidents. In: *Biomed Tech - Proceedings BMT 2012*, vol. 57 (Suppl. 1), pp 784-787 (2012).
38. Mentler, T., Herczeg, M.: Routine- und Ausnahmebetrieb im mobilen Kontext des Rettungsdienstes. In: Boll S, Maaß S, Malaka, R (eds.) *Mensch & Computer 2013*. De Gruyter, München, pp 109-118 (2013).

39. Mentler, T., Herczeg, M.: Applying ISO 9241-110 Dialogue Principles to Tablet Applications in Emergency Medical Services. In: Comes T, Fiedrich F, Fortier S, Geldermann J, Müller T (eds) 10th International ISCRAM Conference. ISCRAM, Baden-Baden. pp 502-506 (2013).
40. Mentler, T., Herczeg, M.: Human Factors and Ergonomics in Mobile Computing for Emergency Medical Services. In: Ahram T, Karwowski W, Marek T (eds) Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics. AHFE, Krakow, pp 4149-4160 (2014).
41. Streger, M.R.: Prehospital triage. *Emergency medical services* 27, 21-27 (1998).
42. Carengo, L., Barra, F.L., Ingrassia, P.L., Colombo, D., Costa, A., Della Corte, F.: Disaster medicine through Google Glass. *European Journal of Emergency Medicine* 3, 222-225 (2015).
43. Cicero, M.X., Walsh, B., Solad, Y., Whitfill, T., Paesano, G., Kim, K., Cone, D.C.: Do you see what I see? Insights from using google glass for disaster telemedicine triage. *Prehospital and disaster medicine* 1, 4-8 (2015).
44. Berndt, H., Mentler, T., Herczeg, M.: Smartglasses for the Triage of Casualties and the Identification of Hazardous Materials. *i-com* 2, 145-153 (2016).
45. Bundesamt für Bevölkerungsschutz und Katastrophenhilfe: In: 6 Sichtungs-Konsensus-Konferenz (2015). http://www.bbk.bund.de/SharedDocs/Downloads/BBK/DE/Downloads/GesBevS/6_Konsensus-Konferenz_Protokoll.pdf?__blob=publicationFile.
46. Berndt, H., Mentler, T., Herczeg, M.: Optical Head-Mounted Displays in Mass Casualty Incidents: Keeping an Eye on Patients and Hazardous Materials. *International Journal of Information Systems for Crisis Response and Management* 3, 1-15 (2015).