Quality Attributes Analysis in a Crowdsourcing-based Emergency Management System

Ana Maria Amorim[1], Glaucya Boechat[1], Renato Novais[2,3], Vaninha Vieira[1,3] and Karina Villela[4]
[1]*Department of Computer Science, Federal University of Bahia, Salvador, Brazil*
[2]*Department of Computer Science, Federal Institute of Bahia, Salvador, Brazil*
[3]*Fraunhofer Project Center for Software and Systems Engineering at UFBA (FPC-UFBA), Salvador, Brazil*
[4]*Fraunhofer Institute for Experimental Software Engineering (IESE), Kaiserslautern, Germany*

Keywords: Crowdsourcing, Software Quality, Mobile Computing, Emergency Management.

Abstract: In an emergency situation where the physical integrity of people is at risk, a mobile solution should be easy to use and trustworthy. In order to offer a good user experience and to improve the quality of the app, we should evaluate characteristics of usability, satisfaction, and freedom from risk. This paper presents an experiment whose objective is to evaluate quality attributes in a crowdsourcing-based emergency management system. The quality attributes evaluated are: appropriateness recognisability, user interface aesthetics, usefulness, trust, and health and safety risk mitigation. The experiment was designed following the Goal/Question/Metric approach. We could evaluate the app with experts from the area of emergency. The results showed that the participants thought the app was well designed, easy to understand, easy to learn, and easy to use. This evaluation ensured the application improvement, and also the evaluation process adopted.

1 INTRODUCTION

In the context of crisis and emergency, not only material goods but also the physical integrity of people is at risk. In such situations there is a need for systems that have quick answers. Nowadays, more and more people are carrying a mobile device. The number of applications for these devices increases every day. Crowdsourcing systems take advantage of people having a mobile device to use them as data entry for their solutions. These applications have been used in different areas, and each one of them with their own challenges and characteristics. In the context of crisis and emergency management are emerging crowdsourcing solutions such as RESCUER. The RESCUER project aims to develop an intelligent solution based on computer, to support emergency and crisis management, with a special focus on incidents that occur in industrial areas or large events. An important aspect to improve the emergency and crisis management is better decision making and clearer coordination through better information flow.

The growth in the number of applications leads companies to invest in the development of products with high quality. According to ISO/IEC 25010:2011 (ISO/IEC, 2011), there are many aspects to be considered to measure the quality of a product, one of those is usability. If companies want to improve the quality of their products from the point of view of user's acceptance they have to take care of attributes like user experience and usability.In case of applications aimed at mobile devices, usability should consider aspects such as limited resources, connectivity issues, data entry models, and the varying display resolutions of mobile devices (Nayebi et al., 2012; Ouhbi et al., 2015). Users look for mobile apps that are easy to learn and user-friendly.

The Mobile Crowdsourcing Solution (MCS) is the mobile component of RESCUER to be installed in users' smartphones. MCS supports end-users in providing command center with information about an emergency situation. The app takes into account different smartphones and how people interact with them under stress. The goal is to benefit as much as possible from information provided without any explicit action of their users, but taking into consideration the users privacy.

One of the challenges for the development of software on distributed platforms and mobile devices is

Amorim, A., Boechat, G., Novais, R., Vieira, V. and Villela, K.
Quality Attributes Analysis in a Crowdsourcing-based Emergency Management System.
DOI: 10.5220/0006360405010509
In *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) - Volume 2*, pages 501-509
ISBN: 978-989-758-248-6

that both the users and the context of use will be known only after the applications are released for use (Sherief et al., 2014; Pagano and Bruegge, 2013). The main source of information of RESCUER are the citizens, who through their mobile phones report an emergency situation to the command center. The command center is formed by a group of people in charge of assessing the risks and making decisions.

In the case of RESCUER, another challenge is the context where the application will be used: an emergency situation. On such situation, the evaluation must also consider aspects such as interaction, people's stress, difficulty of being used by the people in charge of emergency situation, adverse environment due to the presence of too much smoke or noise. Those aspects were previously evaluated (Nass et al., 2014). Interaction is a key aspect in motivating, and keeping participation and collaboration of people, improving crowdsourcing systems (Pan and Blevis, 2011; Thuan et al., 2016).

Besides the reduction of damage to people's lives, one of the desired outcomes of RESCUER project is to improve the image of the organizations involved in an incident through the use of modern technologies for effective and efficient management of emergency and crisis situation. So the research question that guides this article is: how do we know if the app being developed for use by the crowd in emergency situations like fire and explosion is being developed to be easy to use and reliable.

This paper presents an experiment to evaluate a crowdsourcing-based system to be used in crisis and emergency situation. In the experiment, we were concerned with the quality of the product and the quality in use, and we reported a practical case of user experience evaluation.

We know that a well-accepted application has been previously evaluated by its potential users. With this in mind, we conducted our experiment using volunteers of different profiles considered in RESCUER solution: workforce, supporting force, and civilian. We simulated three different scenarios for each user profile. The scenarios were related to different areas of the incident according to the fire proximity as: hot, warm, and cold zone. We tried to contextualize each scenario to make the participant feel like in the real scenario. The experiment included 13 civilian, 3 supporting force, and 15 workforce. We addressed three questions related to product quality, and three questions related to quality in use.

The main contribution of the article is: it presents an experiment that other researchers can use in case they need to evaluate quality attributes such as usability, satisfaction, and freedom from risk, in a crowdsourcing-based emergency system. With this experiment, we ensured the application improvement, and also the evaluation process adopted.

2 RELATED WORK

The Quality in Use (QiU) is one of the challenges for the implementation of mobile systems (Osman and Osman, 2013). ISO/IEC 25010:2011 defines QiU as "degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk and satisfaction in specific contexts of use" (ISO/IEC, 2011).

ISO 25010:2011 breaks down the notion of quality-in-use into usability-in-use, flexibility-in-use, and safety-in-use, it also defines satisfaction-in-use considering four aspects: likeability, as satisfaction of pragmatic goals; pleasure, as satisfaction of hedonic goals; comfort, as physical satisfaction; and trust, as satisfaction with security. Flexibility-in-use is defined in the context of conformity-in-use, extendibility-in-use, and accessibility-in-use (Nayebi et al., 2012).

In (Aedo et al., 2012) the authors address emergency notification and evacuation adapting emergency alerts and evacuation routes to the context and the profile of each person. They take into account the importance of customized evacuation routes. They assess the usability and understandability of their solution through a user study performed by twelve students at the University in a fire simulation in an indoors environment. The user study consists of three phases: training, task execution, and usability questionnaire. The questionnaire is composed of four groups of questions: system and past experiences, system acceptance, task execution, and user satisfaction.

In (Kuula et al., 2013) the authors evaluated the performance and usability of a smartphone system for alerting, command and communication purpose of the national police. A group of ten policemen, and few civilians formed the test group. After the experiment, key persons in different levels of the police organization were interviewed face to face, and users were asked to answer an internet questionnaire.

In (Gomez et al., 2013) the authors analyzed more than 250 Android emergency applications from Google Play database related to emergency detection, notification or management, where they concluded that none of them serve citizens in emergency situations managed or solved by a public emergency service, and few applications handle the potential of citizens as witnesses and volunteers.

There are many studies on the use of crowdsour-

cing platforms in the evaluation of mobile systems (Sherief et al., 2014) (Birch and Heffernan, 2014). The problem of using this solution in the RESCUER app is the context in which it will be used, context of emergencies and crises in industrial or large events parks. In this case the assessments should take place in simulated environments.

Usability testing is important for the success of any application. According to Liu et al. (Liu et al., 2012) in a system project, usability tests should begin early and often, but the cost and effort required to recruit participants, engage observers and buy or rent equipment can make these tests impossible. However, an analysis of the cost-benefit of these tests has drawn attention to ROI (return on investment) when evaluating the usability is inserted at the beginning of the development of a product.

3 METHODOLOGY

In this work we followed the Goal/Question/Metric (GQM) approah (Basili, 1992). The object of study in this experiment is the MCS RESCUER's component. The general goal is assess the quality of the product and the quality in use of the component. The specific objective is to analyze some attributes of product quality, and some attributes of quality in use, for the purpose of evaluate. In product quality we are analyzing two characteristics of usability: appropriateness recognisability and user interface aesthetics. Related to quality in use we are analyzing freedom from risk and two characteristics of satisfaction: usefulness and trust. We are analyzing the product from the point of view of the user, in the context of large event. A system proposed to be used in a stressful environment might consider the social and cultural aspects of the place. Acceptance criteria were defined at the project requirements elicitation and during meetings with project members.

In the experiment, we address three research questions related to product quality, *PQ1* through *PQ3*, and three research questions related to quality in use, *QU1* through *QU3* (Table 1). *PQ1* and *PQ2* are related to appropriateness recognisability, a characteristic of usability. Appropriateness recognisability is the degree to which users can recognize whether a product or system is appropriate for their needs. *PQ3* is related to user interface aesthetics, another characteristic of usability. User interface aesthetics is the degree to which a user interface enables pleasing and satisfying interaction for the user. *QU1* is related to usefulness and *QU2* is related to trust, both are characteristic of satisfaction. Usefulness is the degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use. Trust is the degree to which a user or other stakeholder has confidence that a product or system will behave as intended. *QU3* is related to health and safety risk mitigation, a characteristic of freedom from risk. Freedom from risk is the degree to which a product or system mitigates the potential risk to people in the intended contexts of use. We used questionnaire to collect the answers of the users.

4 EXPERIMENTAL EVALUATION

In this experiment we tried to reproduce in textual form, different environments so that participants could somehow be transported mentally to those scenarios as they were reading. In each scenario, participants performed the requested task freely. After complete the tasks, participants answered a questionnaire where they could report their experience with the application.

4.1 Context

Three different scenarios of the incident situation were created for each user profile: civilian, supporting force, and workforce. The scenarios were related to different areas of the incident: hot, warm, and cold zone. In each scenario, we tried to make the participant feel like in the real scenario.

4.2 Participants

We performed the experiment with 31 participants at the *Congresso Internacional de Desastres em Massa* (CIDEM 2016), an event occurring in June 10-12, 2016. As one of our goals was to evaluate the usability of an application to be used by the crowd, this application should be intuitive, its use should not require any previous training (Gafni, 2009). So, we considered two groups: one with demonstration and other without demonstration. In order to get a close number of participants for each group, the first participant was introduced to the app with a brief demonstration, the second not, the third was, and so on. During the selection, we tried to assure that the participants had previously faced problems related to emergency situations.

Table 1: Questions Related to Quality Criteria According to ISO25010.

	Software Characteristic	Subcharacteristic	Research Question	Measure
Product Quality	Usability	Appropriateness recognisability	*PQ1* - Are there missing information in the incident report? *PQ1.1* - Is there information that you think to be relevant to handle a fire incident, but you could not provide using the app? *PQ1.2* - Is there information that you could provide, but you would like to have had a better support from the app?	At most 20% of participants responding Yes
			PQ2 - Are there no relevant information in the incident report? *PQ2.1* - Is there a part of the app (e.g. take picture, free text report) that you think not to be relevant to report an incident? *PQ2.2* - Is there a piece of information (e.g. color of the smoke) in the report form that you know not to be relevant to handle the incident?	At most 20% of participants responding Yes.
		User interface aesthetics	*PQ3* - How pleasant and satisfying is the system for the user?	Average score should be higher than 75%.
Quality in use	Satisfaction	Usefulness	*QU1* - How useful is the system for the user? *QU1.1* - Would you use this app to help operational forces if an emergency situation like this occurs during a big event?	80% of participants responding Yes.
		Trust	*QU2* - How trustworthy is the system to the user? *QU2.1* - Would you use this app to request help (for yourself) if an emergency situation like this occurs during a big event?	80% of participants responding Yes.
	Freedom from risk	Health and safety risk mitigation	*QU3* - Do you feel running further risks due to the use of the,app? *QU3.1* - Will I feel safer with the RESCUER app on my,smartphone during a large-scale event? *QU3.2* - Did you have the feeling of running further risks when using the app?	*QU3.1*: 80% of participants responding Yes. *QU3.2*: At most 20% of the participants responding YES.

4.3 Task Design

We defined representative tasks of the MCS. The definition of the tasks were based on demands identified during the processes of an emergency of fire. Therefore, they represent real world emergency scenarios. Additionally, these tasks were selected based on their different levels of difficulty. The tasks are: sending notification report (quick report), sending structured report (complete report), and sending help report.

4.4 Experimental Procedures

A pilot study was performed prior to the experiment with the intention of identifying certain problems in its procedures, or even in the application, which are difficult to predict during its execution. Four participants were selected to perform the pilot study. Two of them with demonstration and two of them without demonstration. It is important to highlight that these four participants did not take part of the final experiment. The pilot study allowed us to improve the tasks description, goals and degree of difficulty. In addition, it was also essential to predict the necessary average time for each participant execute the tasks.

We prepared the environment to the experiment in the area outside the auditorium, where the event was going on. We used Smartphones with OS Android Jelly Bean (4.1) or newer installed, connected to Internet. The MCS application were previously installed to be ready for use by the volunteers during the experiment execution. Each participant took the necessary time to conclude the experiment. The sessions were supervised to avoid parallel conversations among the participants.

First of all, we invited people to participate in the experiment. Then, we explained the experiment general purpose. After that, we classified the participant according with his/her profile.

If the participant was going to perform the task without demonstration, we presented the MCS app and its main objective, the scenarios of the incident situation (written in the flip-chart), and the tasks they had to execute.

In case the participant was going to perform the task with demonstration, we presented the MCS app showing how they can report an incident, and the different options of the MSC, the scenarios of the incident situation (written in the flip-chart), and the tasks they had to execute. Two small tasks of a simple example were performed. The participants could observe how to use the tools by means of these small tasks. They were encouraged to address any doubts

or concerns during the training session

After that, each participant could start to perform the task to report incidents, and had no limit of time to finish it.

After finish the tasks in all the three scenarios, we asked participants to spend between five to ten minutes to answer a questionnaire where we could get quantitative and qualitative feedback about the solution.

5 RESULTS

We collected a total of 31 questionnaires. Being 18 male and 13 female participants. The age were between 22 and 53 years old. Civilian participants answered 13 questionnaires, supporting force 3, and workforce 15. Table 2 shows the distribution of the participants. Figure 1 shows the distribution of workforces according to their area of expertise.

Table 2: Profile of the Participants.

Participant profile	With demonstration	Without demonstration	Total
Civilian	6	7	13
Supporting Force	2	1	3
Workforce	7	8	15

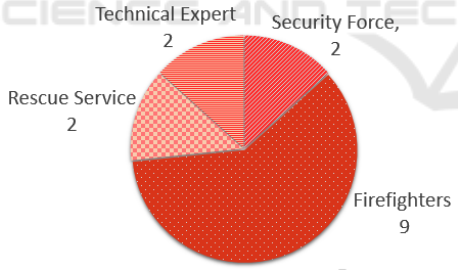

Figure 1: Area of expertise of Workforce profile.

5.1 Product Quality - Appropriateness Recognisability

With this characteristic of usability we measured the degree to which the participants could recognize whether the mobile application MCS was appropriate for their needs, in the case of this experiment, inform a fire incident report. Analyzing questions *PQ1*, *PQ2*, and their derivation we could measure appropriateness recognisability characteristic, which acceptance criteria must be at most 20%.

In question *PQ1.1 - Is there information that you think to be relevant to handle a fire incident, but you could not provide using the app?* 35% of participants answered YES (there is information that they think to be relevant to handle a fire incident, but they could not provide using the app), and others 65% participants answered NO. The answers of the participants to the question *PQ1.1* show that the mobile application MCS is still not in accordance with this characteristic of usability because to be in accordance with appropriateness recognisability of product quality is necessary that at most 20% of the participants answer YES.

We verified that participants without demonstration had more difficulties with the mobile application than the participants with demonstration. 46.67% of participants without demonstration answered that they could not provide information relevant about a fire incident using the mobile application, while participants with demonstration were 25%.

The answers of the participants to the question *PQ1.2 - Is there information that you could provide, but you would like to have had a better support from the app?* show the same result collected in the previous question, that the mobile application MCS is still not in accordance with appropriateness recognisability. In this case, 42% of participants answered YES (there is information that they could provide, but they could not provide using the app, they would like to have had a better support from the app), 55% of participants answered NO, and 3% of participants did not answer. And to be in accordance with appropriateness recognisability of product quality is necessary that at most 20% of the participants answer YES.

The answers of the participants to the question *PQ2.1 - Is there a part of the app (e.g. take picture, free text report) that you think not to be relevant to report an incident?* show that the mobile application MCS is in accordance with appropriateness recognisability. In this case, only 10% of participants answered YES (there is a part of the app that they think not to be relevant to report an incident), and 90% of participants answered NO. And to be in accordance with appropriateness recognisability of product quality is necessary that at most 20% of the participants answer YES.

The answers of the participants to the question *PQ2.2 - Is there a piece of information (e.g. color of the smoke) in the report form that you know not to be relevant to handle the incident?* show that the mobile application MCS is in accordance with appropriateness recognisability. In this question 16% of participants answered YES (there is a piece of information in the report form that they know not to be relevant to report an incident), 81% of participants answered NO, and 3% of participants answered that they did

not know. And to be in accordance with appropriateness recognisability of product quality is necessary that at most 20% of the participants answer YES.

5.2 Product Quality - User Interface Aesthetics

This characteristic of usability measures the degree to which a user interface enables pleasing and satisfying interaction for the user. We could measure this characteristic analyzing question *PQ3*, based on AttrakDiff (Hassenzahg, 2008). The question is composed by nine items to be valued according to the mobile application user's experience. If the average of the answers is close to the value 1, this means that the quality tends to the word on the left side, on the other hand, if the value tends to 7, the quality tends to the word on the right side.

Based on the answers of the participants, we noticed that participants with demonstration considered the mobile application more clearly structured and more beautiful than participants without demonstration (Figure 2).

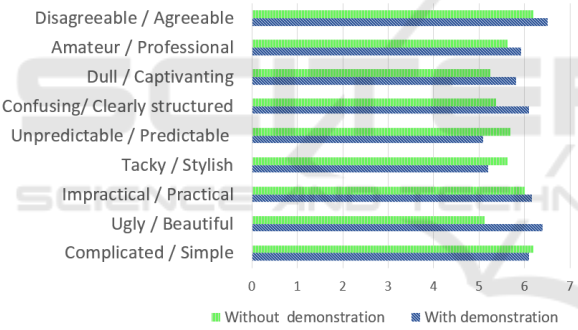

Figure 2: Answers of the participants referring to user interface aesthetics of product quality - Question PQ3: How pleasant and satisfying is the system for the user?

5.3 Quality in Use - Usefulness

With this characteristic of satisfaction we measured the degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use. We could measure this characteristic analyzing question *QU1* and its derivation.

Based on the answers of the participants, we can conclude that mobile application is in conformity with this characteristic of usefulness. In this case, 100% of participants answered YES, they agree that they would use the mobile application to help operational forces if an emergency situation occurs during a big event. To be in accordance with usefulness of quality in use is necessary at least 80% of the participants answer YES.

5.4 Quality in Use - Trust

With this characteristic of satisfaction we measured the degree to which a user or other stakeholder has confidence that a product or system will behave as intended. We could measure this characteristic analyzing question *QU2* and its derivation.

Based on the answers of the participants, we can conclude that mobile application is in conformity with this characteristic of usefulness. In this case, 93,55% of participants answered YES, they agree that they would use the mobile application to request help if an emergency situation occurs during a big event. To be in accordance with usefulness of quality in use is necessary at least 80% of the participants answer YES.

5.5 Quality in Use - Health and Safety Risk Mitigation

With this characteristic of freedom from risk we measured the degree to which a product or system mitigates the potential risk to people in the intended contexts of use. We could measure this characteristic analyzing question *QU3* and its derivation. Based on the answers of the participants, we can conclude that mobile application is in conformity with this characteristic of freedom from risk. In question *QU3.1*, 25 of 31 participants answered YES (corresponds to 80,64% of participants), they agree that they feel safer with the RESCUER app on their smartphone during a large-scale event (Figure 3). To be in accordance with health and safety risk mitigation of quality in use is necessary at least 80% of the participants answer YES.

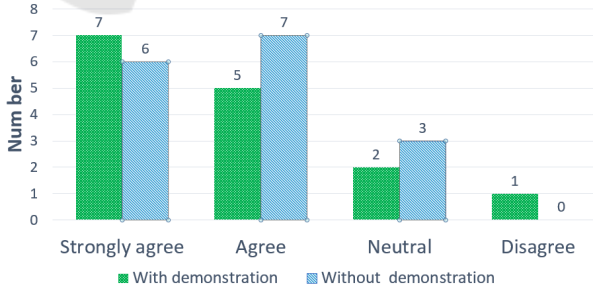

Figure 3: Answers of the participants referring to health and safety risk mitigation of quality in use - Question QU3.1: Will I feel safer with the RESCUER app on my smartphone during a large-scale event?

With the answers of question *QU3.2*, we can conclude that mobile application is not in conformity with this characteristic of freedom from risk. In this question 58.07% of participants answered YES, that

they had the feeling of running further risks when using the mobile application in at least one of the three zones (hot, warm or cold). To be in accordance with health and safety risk mitigation of quality in use is necessary at most 20% of the participants answer YES.

During the experiment, we received positive feedback about product quality and quality in use. Some participants of workforce reported that the application provides essential and important information for the tactical level , the MCS is very important to save lives, will speed up the arrival of workforce, and is intuitive and responsive to the use of civilians. A civilian and a supporting force emphasized that using the application will improve the process of deal with the incident.

We have also collected negative opinions as Internet connection availability. In fact, during an emergency, users feel stressed and worried about their own situation and they prefer report information using audio or video than having to choose multiple options. Some civilians mentioned some weaknesses in the application, such as the lack of an option to report the existence of fire hydrant, escape routes, and status of the incident. Some participants found it difficult to inform the color of smoke, point the location of the fire, the status of the incident, and the number of affected people. They suggested the addition of the panic bottom, and the possibility to have infrared vision. They also wanted to report details about the incident, like local temperature, security and radiation level, besides the possibility to send audio and video. It is worth emphasizing that the evaluated version was not the final version of the application, these features were already designed and will be part of the final version.

Summarizing the results obtained from the user study, we can conclude that the participants have found the application clearly structured and easy to understand, to learn and to use. In particular, they have appreciated as the mobile application is clearly structured, employs the usage of concise terminology and messages with a clear understanding of their meanings.

6 THREATS TO VALIDITY

There are some threats to the validity on this experiment (Wohlin et al., 2000). In this section we will discuss them.

Conclusion Validity. These threats affect the ability to make the right conclusion about relation between the experiment and its outcome. We identified one threat: design of the tasks - tasks may have been misinterpreted by participants. To mitigate this threat, we defined representative tasks based on real demands identified in a fire emergency.

Construct Validity. These threats are related to the result of the experiment according to its concept or theory (if the test measures what it claims). We identified one threat: operational procedures of the experiment - the participants may have not understood the experiment guideline, the change of scenario, for example, from the hot zone to the warm zone. To mitigate this threat, in addition to describing in detail each zone of fire, we used one sheet of paper for each incident zone on the flip-chart.

Internal Validity. These threats are related to issues that affect independent variables without the researcher's knowledge. We identified two threats: (i) instrumentation - the scenarios showed on the flip-chart were not the better way to simulate a real incident. In an emergency scenario this limitation cannot be easily removed; (ii) selection - people were volunteers in our experiment. According to Wohlin, volunteers are more motivated for a new task.

External Validity. These threats represent limitations to generalize the results of our experiment to industrial practice. We identified one threat: interaction of setting and treatment - the experimental environment is not representative of a realistic scenario, since we are dealing with emergency situation.

7 FINAL REMARKS

Resuming our research question, *how do we know if the app being developed for use by the crowd in emergency situations like fire and explosion is being developed to be easy to use and reliable*, we performed an experiment to evaluate usability, satisfaction, and freedom from risk in a crowdsourcing-based emergency system. This evaluation included not only finding any usage problems, but also collecting opinions and suggestions from participants. In general, positive feedbacks were received in particular about the easiness to understand, to learn and to use MCS RESCUER's component. Moreover, due to the relevant role of mobile phones on our daily life, users found very innovative the idea to use them in emergency situations.

During the evaluation process we observed some points that can be improved for an upcoming evaluation. A major challenge in evaluating emergency solutions is to get participants to feel in the emergency scenario. By the restriction of physical space we tried to locate the participant in the various scenarios of the

emergency (hot, warm and cold zone) through text. We noticed that this feature alone was not enough. Many people did not read the full text, and many did not realize the peculiarities characteristic of each situation. We suggest the use of audio-visual resources as a possibility to make the scenario more real. Another aspect to be improved is the app demonstration. We used a script, which should be followed by the research team, to demonstrate the use of the application. We believe that the use of a video is a better solution to standardize the demonstration and make it easier.

Based on the results of this study, we will take into account problems occurred during the user study and collected suggestions from user experiences, in particular concerning localization with Wi-Fi triangulation, and feel to run further risks due to the use of the app in hot zone. As future work, we are planning to use RESCUER on a simulation in a industrial park scenario. On this experiment, we intend to analyze the impact of using RESCUER in a different scenario with group of people.

ACKNOWLEDGEMENTS

This work was supported by the RESCUER project, funded by the CNPq/MCTI (Grant: 490084/2013-3) and by European Commission (Grant: 614154).

REFERENCES

Aedo, I., Yu, S., Daz, P., Acua, P., and Onorati, T. (2012). Personalized Alert Notifications and Evacuation Routes in Indoor Environments. *Sensors*.

Basili, V. R. (1992). Software Modeling and Measurement: The Goal Question Metric Paradigm. *Computer Science Technical Report Series, CS-TR-2956 (UMIACS-TR-92-96), University of Maryland, College Park, MD*.

Birch, K. and Heffernan, K. (2014). Crowdsourcing for Clinical Research An Evaluation of Maturity. *Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand*, pages 3–11.

Gafni, R. (2009). Usability issues in mobile-wireless information systems. *Issues in Informing Science and Information Technology*, 6:755–769.

Gomez, D., Bernardos, A. M., Portillo, J. I., Tarrio, P., and Casar, J. R. (2013). A Review on Mobile Applications for Citizen Emergency Management. *In Proceedings of the 11th International Conference on Practical Applications of Agents and Multi-agent Systems, Salamanca, Spain, 2224 May 2013*, pages 190–201.

Hassenzahg, M. (2008). The Interplay of Beauty, Goodness, and Usability in Interactiv Products. *Hum.-Comput. Interact.*, 19(4):319–349.

ISO/IEC (2011). 25010:2011. In *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*.

Kuula, J., Kettunen, P., Auvinen, V., Viitanen, S., Kauppinen, O., and Korhonen, T. (2013). Smartphones as an Alerting, Command and Control System for the Preparedness Groups and Civilians: Results of Preliminary Tests with the Finnish Police. *ISCRAM 2013 Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management*.

Liu, D., Bias, R. G., Lease, M., and Kuipers, R. (2012). Crowdsourcing for usability testing. *Proceedings of the ASIST Annual Meeting, v. 49, n. 1*.

Nass, C., Breiner, K., and Villela, K. (2014). Mobile Crowdsourcing Solution for Emergency Situations: Human Reaction Model and Strategy for Interaction Design. *In Proceedings of the 1st International Workshop on User Interfaces for Crowdsourcing and Human Computation (CrowdUI)*.

Nayebi, F., Desharnais, J.-M., and Abran, A. (2012). The state of the art of mobile application usability evaluation. *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4.

Osman, N. B. and Osman, I. M. (2013). Attributes for the quality in use of mobile government systems. In *Proceedings - 2013 International Conference on Computer, Electrical and Electronics Engineering: 'Research Makes a Difference', ICCEEE 2013*, pages 274–279.

Ouhbi, S., Fernández-Alemán, J. L., Pozo, J. R., El Bajta, M., Toval, A., and Idri, A. (2015). Compliance of blood donation apps with mobile os usability guidelines. *Journal of medical systems*, 39(6):63.

Pagano, D. and Bruegge, B. (2013). User involvement in software evolution practice: a case study. *Proceedings of the 35th International Conference on Software Engineering (ICSE), San Francisco, CA, 2013. IEEE.*, pages 953–962.

Pan, Y. and Blevis, E. (2011). A Survey of Crowdsourcing as a Means of Collaboration and the Implications of Crowdsourcing for Interaction Design. *In 2011 International Conference on Collaboration Technologies and Systems (CTS)*.

Sherief, N., Jiang, N., Hosseini, M., Phalp, K., and Ali, R. (2014). Crowdsourcing Software Evaluation. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, page 19.

Thuan, N. H., Antunes, P., and Johnstone, D. (2016). Factors influencing the decision to crowdsource: A systematic literature review. *Information Systems Frontiers*, 18(1):47–68.

Wohlin, C., Runeson, P., Hst, M., Ohlsson, M., Regnell, B., and Wessln, A. (2000). *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Norwell, MA, USA.

REFERENCES

Aedo, I., Yu, S., Daz, P., Acua, P., and Onorati, T. (2012). Personalized Alert Notifications and Evacuation Routes in Indoor Environments. *Sensors*.

Basili, V. R. (1992). Software Modeling and Measurement: The Goal Question Metric Paradigm. *Computer Science Technical Report Series, CS-TR-2956 (UMIACS-TR-92-96), University of Maryland, College Park, MD*.

Birch, K. and Heffernan, K. (2014). Crowdsourcing for Clinical Research An Evaluation of Maturity. *Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand*, pages 3–11.

Gafni, R. (2009). Usability issues in mobile-wireless information systems. *Issues in Informing Science and Information Technology*, 6:755–769.

Gomez, D., Bernardos, A. M., Portillo, J. I., Tarrio, P., and Casar, J. R. (2013). A Review on Mobile Applications for Citizen Emergency Management. *In Proceedings of the 11th International Conference on Practical Applications of Agents and Multi-agent Systems, Salamanca, Spain, 2224 May 2013*, pages 190–201.

Hassenzahg, M. (2008). The Interplay of Beauty, Goodness, and Usability in Interactiv Products. *Hum.-Comput. Interact.*, 19(4):319–349.

ISO/IEC (2011). 25010:2011. In *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*.

Kuula, J., Kettunen, P., Auvinen, V., Viitanen, S., Kauppinen, O., and Korhonen, T. (2013). Smartphones as an Alerting, Command and Control System for the Preparedness Groups and Civilians: Results of Preliminary Tests with the Finnish Police. *ISCRAM 2013 Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management*.

Liu, D., Bias, R. G., Lease, M., and Kuipers, R. (2012). Crowdsourcing for usability testing. *Proceedings of the ASIST Annual Meeting, v. 49, n. 1*.

Nass, C., Breiner, K., and Villela, K. (2014). Mobile Crowdsourcing Solution for Emergency Situations: Human Reaction Model and Strategy for Interaction Design. *In Proceedings of the 1st International Workshop on User Interfaces for Crowdsourcing and Human Computation (CrowdUI)*.

Nayebi, F., Desharnais, J.-M., and Abran, A. (2012). The state of the art of mobile application usability evaluation. *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4.

Osman, N. B. and Osman, I. M. (2013). Attributes for the quality in use of mobile government systems. In *Proceedings - 2013 International Conference on Computer, Electrical and Electronics Engineering: 'Research Makes a Difference', ICCEEE 2013*, pages 274–279.

Ouhbi, S., Fernández-Alemán, J. L., Pozo, J. R., El Bajta, M., Toval, A., and Idri, A. (2015). Compliance of blood donation apps with mobile os usability guidelines. *Journal of medical systems*, 39(6):63.

Pagano, D. and Bruegge, B. (2013). User involvement in software evolution practice: a case study. *Proceedings of the 35th International Conference on Software Engineering (ICSE), San Francisco, CA, 2013. IEEE.*, pages 953–962.

Pan, Y. and Blevis, E. (2011). A Survey of Crowdsourcing as a Means of Collaboration and the Implications of Crowdsourcing for Interaction Design. *In 2011 International Conference on Collaboration Technologies and Systems (CTS)*.

Sherief, N., Jiang, N., Hosseini, M., Phalp, K., and Ali, R. (2014). Crowdsourcing Software Evaluation. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, page 19.

Thuan, N. H., Antunes, P., and Johnstone, D. (2016). Factors influencing the decision to crowdsource: A systematic literature review. *Information Systems Frontiers*, 18(1):47–68.

Wohlin, C., Runeson, P., Hst, M., Ohlsson, M., Regnell, B., and Wessln, A. (2000). *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Norwell, MA, USA.