Towards using Serious Games for realistic evaluation of disaster management IT tools

Kenny Meesters
Tilburg University
k.meesters@tilburguniversity.edu

Abstract

Technological developments, along with political and financial incentives have encouraged an increasing number of developers and researchers to develop tools and systems to aid disaster responders. To ensure adoption of these innovations evaluations are necessary. Evaluations aid in assessing the impact, the fit with the users' requirements and the identification of improvements. However options for conducting realistic evaluations of tools tailored for use in a disaster context are limited. Real-life evaluations are unfeasible because of the disaster context, while dedicated simulations are costly.

This paper proposes the use of serious games as an evaluation method; striking a balance between realism required for valid and useful feedback and the light-weight resource use to ensure multiple iterative evaluations are feasible. In addition to the theoretical outline this evaluation method has been applied to a system for crisis-related data collection, demonstrating the usability of serious games for IT-tool evaluation in a disaster context.

Introduction

The importance of information during crisis responses and disaster management has been extensively studied, and described in detail. Information like damage assessments and situational overviews supports humanitarian organizations in their efforts to provide efficient and effective aid and relief to affected communities (Comfort et al. 2004; Fiedrich et al. 2000). While the importance of information has been established for quite some years, ongoing innovation and developments continue to present new opportunities for disaster responders and crisis managers. Due to an ever more connected world, the development of mobile technology and the technological advancements in the field of computer science and software engineering new options for information gathering, analysis and dissemination continue to emerge. Examples of such innovations, specifically in information technologies, are illustrated in the Disaster Relief 2.0 document (Crowley et al. 2010) and the World Disasters Report 2013 (Red Cross Societies 2013). These opportunities have invited scholars, developers, innovators and entrepreneurs to develop (digital) tools. This interest has been fuelled by the available funding for such project including the Harvard Humanitarian Innovation Fund, the innovation grant program of the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA), as well as numerous funding options from (supra)national governments such as the European Commission FP7 program. In addition to these funds, many NGOs have also started investing in the development of such tools themselves (Currion et al. 2007).

Problem Statement

The described developments have resulted in an increased number of tools developed in recent years, however only a small selection of these tools have reached wide-spread deployment and use. The reason for the limited adoption of tools comes from a multitude of reasons, including the need for an improvement understanding of the context in which these tools are being used, integration in the existing organizational structures or –more general- the level to which the designed system address the needs and match the expectations of the (end-) users (Venkatesh et al. 2000). In order to improve this adoptability developers would need to forge strong connections with the end-users. Such connection would improve the adoptability through an improved understanding of the context, and the improvement understanding of the needs and requirements of professionals for tools that aid them in their work. The importance of this connection and end-user involvement has been described extensively in the existing literature. Studies have shown that user and stakeholder involvement are key factors in successful information

systems development. Newman (Newman et al. 1990) and Atkinson (Atkinson 1999) amongst others show that user involvement is recommended as part of successful development. There are many ways to achieve such involvement (Sommerville et al. 1998), e.g. through a cyclic, user-centred design process.

This user-centred design process can take part in any stage of the development cycle. For example in requirement engineering phase, in the actual development or the testing stages. Regardless of the specific stage, the process always relies on user involvement to evaluate the development efforts and provide (continuous) feedback as illustrated in Figure 1. This forces developers to not only analyse the users and predict the use of the system, but also verify the validity of their assumptions in real-world tests with actual users. This approach optimizes the design of information tools around the way users' behaviours rather than dictating the users' behaviour in order to use the developed tool (Landgren 2007).

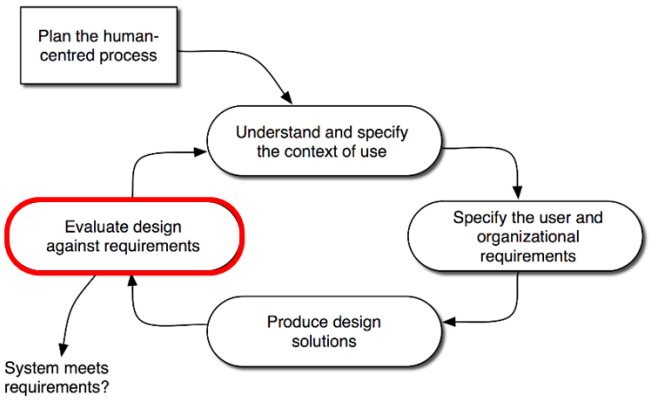

Figure 1 User-centered design process (Jokela et al. 2003)

The methods described in the aforementioned studies rely on the availability of the end-user to actively participate in the development. Engaging with humanitarian professionals in the development of tools, especially those directly involved in crisis response and disaster management however poses some challenges. The volatile nature of the work in disaster and crisis response makes it difficult for developers to plan continuous interaction with crisis responders in the development cycles. Especially when developing for large-scale disasters, in which responders are deployed to other countries for extended periods, in situations where there is limited options even for remote interaction (Turoff et al. 2004).

Another element that contributes to the difficulty in establishing a connection with the end user is that most tools are developed outside of the organizations that the targeted decision makers are part of. In other words, most tools are created by developers who are not part of the crisis response organizations. Additionally, many tools are developed for a generic problem, and target multiple organizations rather than a solution tailored specially to one organization. In such case developers would need to engage with multiple users from different organizations to ensure a representative group of evaluators.

Aim of the research

Wide-spread adoption of applications designed for crisis response and disaster management depends on the level to which the developed tools can address the needs of the end users, i.e. the crisis responders. In addition, developers need to ensure 'good fit' with the behavior of the users and the existing processes in order to increase the likelihood of adoption. In order to evaluate these aspects of in the development of systems, and more importantly to generate feedback for further improvement, a rigorous evaluation method is needed. While evaluations can yield valuable feedback, the quality and validity of the outcomes depend on the use of representative evaluators and a proper evaluation setting.

As mentioned evaluating systems designed for crisis response and disaster management poses some domain specific challenges, related to the volatile nature of the responders' work and their availability to contribute to evaluations and the limited options to test tools in an appropriate setting. The aim of our research is to design an evaluation approach enabling developers to evaluate their work in a valid way providing them with valuable and actionable feedback while reducing the dependency on the availability of the professionals as end users, and a full disaster context as the evaluation environment.

Research methodology

The study presented in this paper is part of ongoing research into evaluating and improving the impact of IT-tools in the disaster response and crisis management domain. In this paper we will explore the potential of *serious games* as an evaluation method for IT-tools developed for information management purposes during and following a disaster or crisis. We propose that serious games can provide a valid, yet light-weight alternative to other evaluation methods, balancing the required realism for valid evaluation with a light-weight, redeploy-able environment. As part of this explorative research we conduct a case-study in which a game instance is used to evaluate an information management tool designed for disaster response is used and evaluated. In the case-study we firstly and foremost examine *if* the concepts of serious games and IT evaluation can be aligned and applied in a real game instance. Secondly the case-study help to grasp an initial idea of the quality and usefulness of the evaluation and its outcomes. The specific evaluation results for the IT tool are however of lesser importance for the study presented here. Participating in the game instance are professional responders and developers of the evaluated tool. The former offers reflection on the game environment and experience, the latter on the evaluation value.

In this paper we first outline the related work of this research. In the next section we start by reviewing the relevant literature around IT development, and focus on the importance of evaluation and testing in the development cycle specifically. Following this review we examine the challenges related to the evaluation in context of disaster response and crisis management and the current methods used as well as their shortcomings. We continue by introducing the serious game concept and how this can be used to address these issues using a specific design approach. Next we describe how these concepts have been applied in our case-study. Finally we discuss the outcomes of the case-study and present some concluding remarks, limitations, and steps for future research.

Related work

In existing literature on designing solutions in general and information technology solution in particular, the value of investing in working with the users has been mentioned as being equally or even more important as the inspiration for a solution (Alexander 1979; Simon 1996). The literature demonstrates that the designer must be very critical as to the utility and satisfaction of the future users and other stakeholders. Therefore, in order to have the desired impact, the solution should satisfy a set of design criteria. In order to test whether the designed solution indeed does satisfy the criteria an evaluation process needs to be put in place. Particular research on evaluation methodology has already provided guidelines and insights for designers to setup and include evaluation as part of the overall design process (Patton 2000; Rossi et al. 2004). In order to understand the role and relation of the evaluation process with the overall design process, we briefly introduce the various steps in the design process first, followed by an overview of evaluation methods. We conclude by describing the shortcomings of these evaluation methods in relation to the introduced domain of IT-tools for emergency and disaster management.

Design Process

Expanding the design-cycle introduced in Figure 1, combined with the aforementioned literature on the design process we can define a generic design process consisting of six stages (Cross 2008). The first stage of the process concerns the initiation, idea-generation or other triggers that lead to an objective and/or a set of high-level goals. This stage is followed by the process of specifying and articulating the requirements and assumptions. These requirements do not only apply to the designed solution itself but also implicitly or explicitly define the demands or expectations of the future users. These requirements should therefore be checked and evaluated with the future users. The third stage is the specification of the solution to be. It describes the various elements of the solution that are required to satisfy the earlier mentioned requirements. The result of these first three stages is a first design that describes the solution in detail. The fourth stage is the realization of a prototype based on the design. This encompasses the entire design and is used for validation and evaluation. These prototype can vary to the extent in which they are functional and useable depending on the desired level of feedback. For example mock-ups are used to facilitate discussions on the functional and user requirements but lack the complete representation of all the interfaces and functions from the design (Stapleton 1997).

The fifth step is the implementation of the prototype, ideally in a real-life scenario. This requires that the context in which the system is intended to be used is created. This context should incorporate all the underlying assumptions about the users, including any training that is assumed in the earlier defined requirements. The final step of the design process is to conduct the evaluation and check whether the designed solution has the desired effect and impact. This includes the verification of the goals defined in the first step of the design process, including the goals of any involved stakeholders. The result of the evaluation often indicates that the solution does not fully satisfy all the goals and the associated requirements and requires refinement, initiating a subsequent design/improvement cycle.

IT-Tool Evaluation

The final two stages of the above mentioned design cycle explicitly relate to the evaluation of the software. However it is important to notice that the evaluation takes place during the whole design process, as illustrated in the referenced user-centered design process. The importance of continues evaluation is particularly important if the development stages require significant resources. To prevent large investments based on assumptions and uncertainty, early and frequent verification are necessary.

Based on the aim of the research, ensuring an accurate and realistic evaluation centered on the users and the context in which they operate, several evaluation methods seem particularly applicable. In general qualitative methods for evaluation are more suited than quantitative ones (Verschuren et al. 1997), the rationale being that the various elements involved in the design process are so complex that a large amount of variables would be needed to capture all relevant feedback. Such qualitative evaluation can be conducted with the use of observers, content analysis or in-depth interviews. This approach has been successfully employed by various researchers in their design process. Most commonly in research areas that involve a combined problem analysis as solution design process. Such combination is often observed in action research, an approach that addresses a problem through simulations research and problem solving efforts. Perry et al. (2004) have demonstrated the use of action research in (simulated) case studies. In this action-research based approach a collaborative effort similar to the one described in the aim of the research is established. Even more, the use of specific case studies ensure both the validity and the adoption of the outcomes. In addition to action research, several other studies have shown the importance of using case studies as part of the evaluation of software (Zannier et al. 2006)

Requirements For Evaluation Design

Evaluation in general is used to compare separate parts of a design with specific criteria in order to determine to what extent the evaluated subject satisfies these criteria (Verschuren et al. 2005). Designing and conducting an evaluation consists of three stages: (1) create a plan for the evaluation, (2) execute the evaluation and (3) asses the results and determine the effect of the solution on the set goals. As illustrated, qualitative data in general provides the most actionable data for designers to find specific points to improve their solution. This requires an implementation of the evaluation method which allows the collection of qualitative data. I.e. allow the presence of observers, interviews with the test-subjects or the capture of content for later analysis. Finally evaluation methods should enable the evaluators to contrast their findings with the set goals. The evaluation should therefore be designed in a way that it enables the drawing of conclusions based on the design choices and the outcome of the evaluation. For example by limiting the scope and confounding factors that could obscure potential causal links.

Evaluation in a disaster context

As illustrated, case-studies are one of the preferred methods for conducting evaluations centered on the user in appropriate context as they provide a realistic and comprehensive picture of how users adopt and interact with a provisioned system. However conducting case-studies in the disaster management and crisis response field is challenging. The real-world field conditions in which the designed system has to be evaluated pose certain difficulties for conducting an evaluation. For example the high-level of stress already excreted on the intended users (professionals), the limited availability of (financial and otherwise) resources or the general disorder, make conducting an evaluation in a real case difficult. In addition there are limited options for developers and researchers to psychically conduct the evaluation or obtain qualitative data and feedback (Van de Walle et al. 2009), because of geographical distance but also the

hazards posed when travelling to a disaster area without proper training. Finally there are limited opportunities, and when an 'opportunity' does present itself, it is unexpected and unpredictable. While the limited availability of cases is obviously a good thing, it limits the options to do rigorous evaluations.

Simulations and Exercises

An alternative to case-studies are simulations or exercises, ranging from abstract analytical models to full, realistic field exercises (Alexander 2000). However each of these methods has their drawbacks. Full scale exercises approximate the reality of a real deployment but require significant resources, for example simulants, materials and effects to mimic the conditions and comprehensively recreate the environment. Furthermore such exercises can often introduce confounding factors, especially when the assumption about the users (such as the training regarding the prototype) has not been implemented correctly. Due to the large scope of a full simulation many additional factors can be introduced that affect the outcome of the evaluation while not being related to the introduced solution, i.e. it will be harder to determine a relationship between the introduced tool and the observed impact and received feedback.

On the other side of the simulation spectrum are table-top exercises, models and computer simulations which are more efficient in their resource use. These simulations can be setup and run over and over again, providing a reusable evaluation environment. However these simulations lack the required realistic context for a valid and rigorous evaluation. Computer simulations for example disregard the human element in the evaluations and can only evaluate those elements introduced by the designer of the simulation. Table top exercises include the human element but lack the realism of a full scale exercise. This realistic setting does not only related to the limited immersion of the exercise but also to the limited options and freedom participants may have, resulting in significant different cognitive processes and behavior of the participants than those in a real-life case. When choosing one of these options a balance has to been found between the efforts required for and the validity (realism) of the simulation, especially considering the cyclic (recurring) nature of the evaluations.

Serious Games

Serious games can be applied to a broad spectrum of application areas, e.g. government, education, corporate, healthcare and crisis response (Michael et al. 2005). However the concept of a serious game itself is defined many different ways, depending on the involved actors and the specific context. The core element that is recurring in both the scientific literature and the applications of serious games is purpose of the game. Serious games serve a different purpose then providing mere entertainment. This however does not exclude entertainment. According to some players and game developers even games that are created purely for entertainment can be taken very seriously. Therefore it is rather the alternative purpose, an ulterior motive, which creates the distinction between serious and pure entertainment games. According to (Abt 1987) "reduced to its formal essence a game is as activity among two or more independent decision-makers seeking to achieve their objectives in some limiting context." Clark continues to add "the problem with this definition is that not all games are a contest among adversaries – in some games the players cooperate to achieve a common goal against an obstructing force or natural situation that is itself not really a player because it does not have objectives".

This definitions provide a clear link to the application of (serious) games in the context of crisis response and management. The independent decision-makers mentioned in the definition can also be identified in the crisis response, for example as the representatives of different organizations responding to a crisis (the limiting context). Even more the player (the crisis responders and managers) need to cooperate to achieve a common goal (the rescue and relief of the affected population) while being obstructed by a natural situation (for example the difficulty and uncertainty of a crisis situation).

When considering the motives or purpose for designing and playing serious games serious games often have a strong learning objective for the players (Aldrich 2003). Other objectives for serious games include the exploration certain ideas, for example of policy and strategy changes (Geurts et al. 2007) in which not (only) the players are part of the objective but also the outcome of the game has a specific value. Serious games can also be deployed as an environment for experimentation in which gaining an understanding of the players is the motive (Marsh 2011). These specific definitions provides a clear indication for the use of serious games as an evaluation environment.

Serious games for evaluation

The challenge in the disaster management context is to provide the developers with an evaluation environment, to better understand the use of developed tools. Although according to the literature there is a clear indication that the intended evaluation method for evaluating IT-tools designed for crisis response and disaster management *could* be implemented or considered a serious game, we also consider *why* a serious game is an appropriate method. As illustrated in the IT-evaluation section there is a need for a realistic but resource-efficient evaluation method. This realism will have to be evident in the behavior of the players and the environment in which the evaluation takes place. Serious games simulate certain elements which allows the designer of the game to create an effective environment while leaving the experience for the players intact. It allows the designers to balance realism with efficiency, for examples by replacing real people with dummies or damaged buildings by augmented/mixed reality. Serious games can put inexperienced players in a certain role, by providing specific briefings, in-game guidance and most importantly motivation (Wouters et al. 2013). Serious games therefore can enable evaluations with non-professional players, in a lightweight scenario while maintaining the validity of those evaluations.

Serious games design for IT evaluation

Serious games can be designed and employed to provide a light-weight sandbox environment using non-professional players for evaluating IT-tools while ensuring valid and reliable results. Thereby satisfying the requirements for a proper evaluation as illustrated in the related work, i.e. a realistic and real-life context and the feasibility to re-evaluate during iterative development cycles.

Game Design Approach

Serious games can be considered a form of exercise or training providing players with the opportunity to either demonstrate, apply or evaluate their knowledge in a setting resembling (parts of) the reality faced 'on the job' (Kirkle et al. 2005) When developing a serious game as an exercise, generally two elements are important: *planning* and *delivery* (Gagné et al. 1975). Planning involves the definition of objectives (purpose) and the design of activities, while delivery is the presentation of these activities to the players. Following this structure, we have created a design workflow that starts with the definition of the purpose of the game during planning, the interaction, options and environment during the delivery design and ends with the evaluation of the game after its execution. The individual steps of the workflow, as depicted in Figure 2 can be related to the roles of game Designers; responsible for *planning* and *delivery*, game Managers; responsible for the execution, "Players" and to the game "Environment" (Link et al. 2014).

In a very first step, the game Designers define the *purpose of the game*: in this case the evaluation of a specific IT-tool. This purpose dictates the further design steps and guides the further design. This purpose should relate to the goals set in design of the solution introduced. In other words, the purpose of the game is to present the players with the goal that the designed solution aim to accomplish. For example if the tool is design to increase situational awareness, the purpose of the game is to present the players with a game that requires them to build this awareness.

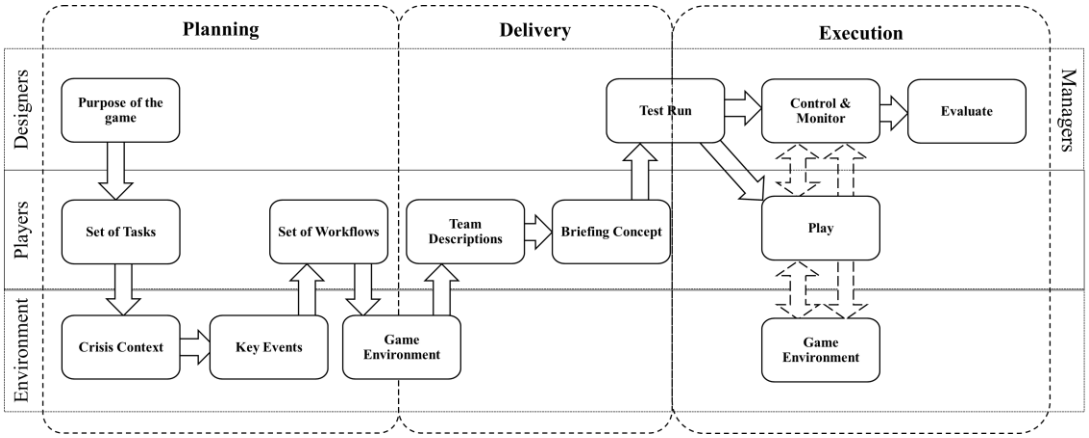

Figure 2 Game Design Workflow

When the purpose is set by the Designers, they may proceed to determine which tasks are appropriate for the Players to handle, thereby also limiting the scope of the game. These tasks will correspond closely with the requirements and assumptions defined in stage two of the design process. When they decided the purpose of the game and the tasks the Players shall handle, they may select the appropriate *crisis context* for the game Environment, e.g. a flood in Western Europe. Next, the tasks are meaningfully combined into in-game *workflows*. The Designers use these workflows for guidance while they further flesh out the game, creating a situation that can be "solved", e.g. by building situational awareness. The workflows are only intended to simplify the planning phase and lead to a consistent game design. Thus, Designers should thus also expect Players to deviate from workflows and think of ways to guide the Players in these cases, e.g. by introducing a non-Player character acting as an expert. These workflows should not be confused with the approach players take as the serious game does not dictate them to undertake certain actions, rather workflows are built on the assumptions of the users as defined in the design process. Next the description of the *game environment* can be completed, e.g. by key locations, information to gather, but also the locations of affected people and materials. With the game environment and the workflows, *team descriptions* can be created: starting location, mission statement, initial resources. Next, the assumptions of the players/users are again used, to create an appropriate *briefing concept*. For instance, your local Red Cross group can be expected to react differently to a flood context with new IT tools than a group of students who are curious about technology but have never given first aid to anyone.

Following the planning phase, the delivery phase consists of the Players *playing* the game and the Designers simultaneously *monitoring and controlling* the game. By monitoring the moves of the players, which can be supported by real-time tracking systems, the Designers are able to dynamically adjust the difficulty of the game (control). When the game script has ended, the Designers should *evaluate* the game itself quite quickly, in order to preserve as many impressions as possible. This can involve feedback sessions with the players but also within the Managers' team or via observers. Another follow-up evaluation is possible too, of course, to capture long-term impressions.

IT-tool Evaluation Within a Serious Game

Designing a serious game for IT-Tool evaluation requires that the game is designed around both the evaluation requirements, as well as the context that the designers of the tool that is evaluated considered in their design process. The incorporation of the design of the solution in the game setup ensures that the game will yield usable results in the evaluation. This alignment is visible in all the steps of the game design process, as illustrated above. For example the purpose of the game aligns with the objectives set for the designed solution (design stage 1). The set of tasks that the players are expected to perform is directly drawn from the functional requirements, in particular the specific processes facilitated by the developed IT-tool (design stage 2). Next the context and key-events are constructed based on the context used in the design for the IT-tool (design stage 3). Finally the team descriptions and briefing concept are used to align the user assumptions of the player with the user assumptions of the designed solution (design stage 4).

The various steps of evaluation for IT-tools as introduced earlier, can be incorporated in the design of the game as well. The game design process illustrated above allows the close examination of the players and their actions. This relates closely to preference for qualitative data-collection; the serious game can provide various options for facilitators, and pre-/post-game interviews with the players. Finally the serious games also ensure that the results are both valid and useable. Through an immersive delivery and game environment, the desired cognitive behavior (stress, uncertainty, etc.) can be invoked. More importantly, the scope of the game can be tailored to fit the processes for which a solution was designed, thereby eliminating confounding factors and facilitating linking the results obtained in the evaluation to the design choices. Finally, because the resulting game is design to fit the evaluation needs and nothing more, the resources use is efficient and the resulting game reusable for continues (iterative) evaluations.

Evaluating GDACSMobile in serious game Disaster in my Backyard

The design workflow and evaluation process presented in this paper have been applied in a case-study. In this case-study a game was designed around a specific software solution that was designed for use in a disaster context that had not been evaluated to such extend before. Note that the purpose of this case study is to demonstrate the feasibly of using serious games for IT-tool evaluation as mentioned in the research methodology. The specific tool-related outcomes are therefore not described or discussed here.

Introduction to GDACSMobile

GDACSMobile target groups, i.e. disaster management professionals ("authorized users") and affected population ("public users"), share their observations from the disaster-affected area. They do this by sending reports to the provider's server via the client application or Twitter. How the client application asks the users to describe an observation is determined by the report structure that is configured on the server. The provider uses the server application to receive and moderate user reports. By receiving reports the provider gains an overview of the situation, and may choose to reconfigure the structure. For instance, the provider may add a specific question about the health services that the affected population can access. When a client next accesses the server, it receives the reconfigured structure. The client then also receives new reports that are passed through moderation. Receiving both the current report structure and current situation reports closes the assessment cycle. With the newly gained information, users can better react to the situation they are in, e.g. by using a reported local transport capacity for delivering aid.

Introduction to Disaster in my Backyard Game

During the ISCRAM (Information System for Crisis Response and Management) summer school, the organizers wanted participants to experience what information management during disasters entails. To introduce participants to this experience the development of an exercise, based on a custom alternate reality game was proposed (Meesters et al. 2013). This provided the opportunity to redesign the game to serve as an evaluation environment for GDACSMobile. Participants of this game instance were 17 (PhD) students as players, 7 disaster management professionals as observers and 3 developers of GDACSMobile.

The game was redesigned using the game design workflow illustrated in the figure above. The *purpose* of the game was to allow non-professional players to experience the use of GDACSMobile in realistic context. More specifically the purpose of the game was to provide players with a challenge that GDACSMobile was designed to address: combining various observations and pieces of information. During the game each team would come across different pieces of information that, when put together, provided them with a common operational overview that enabled all teams to improve the quality of their decisions, for example by knowing where specific resources were located. The *set of tasks* of the game were defined based on functional requirements of GDACSMobile: the ability to find and collect information that is relevant for various disciplines. In order to emulate these tasks four teams were created: one logistic team, one assessment team and two urban search and rescue teams. The *crisis context* and *key-events* were constructed based on the scenario that GDACSMobile was developed for; teams with different specialties and information needs operating in an unknown area with unknown damage. Specifically the context involved a flooding in which a number of citizens needed medical attention and evacuation for which the players had to find the information necessary to setup a camp. Key-events included for example an actor (volunteer) in the role of UN official demanding status updates and various specific pieces information.

The *workflows* were designed to keep the earlier defined *tasks* intact, but simulating knowledge specific operations. Participants for example do not need to have training in doing in-field assessment in order to collect information. Rather pieces of relevant information were represented by markers placed throughout the game environment, allowing players to focus on the capture and sharing of information which GDACSMobile was designed for. Workflows do not dictated the choices of teams; teams are free to make their own decisions and their own strategy, for example exchange information with others teams or apply triage. Consideration was also given to the *team descriptions* and the *briefing concept*, as the players of this serious game are unfamiliar with the context (game) and the solution (GDACSMobile). The briefing included both training in the tool as well as the workflows that are part of the game, priming the players with right information and knowledge needed to be able to play the game. In addition the briefing concept and team description were used to bring players in the right mindset for the game, for example by using uniforms and mimicking an in-field briefing stressing the importance of the upcoming operations.

Evaluation Elements

In addition to the design of the game around GDACSMobile, the various elements of the evaluation design were included. The evaluation objective was to get an understanding of the usability of the developed tool in a crisis context, contrasted with other means of data collection, information sharing and communication methods. This objective is a clear formative objective, as the evaluation is not focused on

measuring if and to what level the design application performed worse or better than conventional means, but rather what the specific differences were, what elements inhibited the use or how the usability was perceived. The data collection for the evaluation was done using a debriefing concept on one hand, and the use of observers on the other hand. The team of observers consisted of 7 disaster management professionals and 3 developers of the introduced tool. In the debriefing players could reflect on the use of the tool and on the game itself, generating qualitative data. The debriefing concept and the instructions for the observers were constructed based on the various requirements and objectives of GDACSMobile. Because the game was designed specifically around the introduced tool, the feedback collected during the debriefing from players and observers could be linked directly to specific elements of the GDACSMobile.

Results and discussion

Overall the use of a serious game yielded useful and valuable feedback for the developers of GDACSMobile. Players of the game noted specific points for improvements, based on their assigned tasks and the overall purpose of the game. During the debriefing for example players indicated that an improved process support would be needed in order to make the tool more useable in situations where users would have limited time and training options. In addition user provided specific user-interface feedback and reported several technical and functional issues. This feedback was confirmed by the professional players, who provided similar feedback. Illustrating that both the non-professional and professional players observed the same issues and thus that game enables non-professionals to generate the same feedback as professionals. More importantly all groups indicated that, despite the introduction of workflows simulating operational procedures, the circumstances, challenges and context faced by the players result in the same behavior during a real-world deployment. Professional players and observers for example mentioned the experienced uncertainty, collaboration difficulties, and stress as realistic.

Furthermore the resulting game (design) can now be reapplied with minimal resources, allowing the developers of GDACSMobile to re-evaluate their solution on a more regular basis. In addition to the reduced amount of resources required, the game can be played with non-professional players, enlarging the pool of evaluators. Nevertheless there are some specific elements of the game design and evaluation process that require modification in order to further improve the quality of the evaluation results.

Game Design Process

Several elements, listed in the game design process are needed to create a consistent game, to allow players to exhibit the desired behavior, including the interaction with the designed solutions and the valid representation of the real-world context. All elements in the design process are important in order to fulfill that requirement and –more importantly- need to be considered and designed in relation to the design process of the solution itself. In order to achieve this integration it is important that the designer of the solution has a clear notion of the relevant contexts and assumptions from the start. Considering the implementation of evaluation in the form of a serious game from the beginning will enable an evaluator to more easily design a suited game. Furthermore a serious game in the early stages of development, can be used to bring together different stakeholders and help determine their requirements through interaction.

Game Experience

The game is designed to induce high levels of stress in several phases, which inhibits the Players' capacity for problem solving. The set of key events thus included options to dynamically adjust the game's difficulty, which turned out to be very useful. More often than not, we had to make the game easier because players took complicated rather than simple approaches to tackle challenges they faced. However, we found it difficult to adjust to the way the players used (or rather: did not use) the tools that were at their disposal. According to the players' feedback, they were overwhelmed by the environment, so they did only use tools that they understood perfectly. While a result in itself, in order to obtain more evaluation data the next iteration of the game should include more training, i.e. a training in the evaluated solution.

Furthermore, the designers have to be highly aware of the players' backgrounds/ specifically the gap between existing knowledge and expertise and the presented objective of the game (*purpose of the game*) has to be large enough to provide challenges to the teams and small enough as not to discourage them. This gap should be considered in every stage of the game design process.

Evaluation Validation

A particular point of interest in the design and use of serious games as an evaluation method, is not only the coherence between the game and the designed solution but also the match between the simulated game environment and the realistic environment for which the tool was designed. Since the serious game involved non-professional players it is difficult to assess purely on their experience if the game is a valid evaluation environment. Therefore it is necessary to include experienced professionals not only in the design of the solution, but also in the design and test of the evaluation method i.e. the serious game (Pawson et al. 1997). However once established that the evaluation method does approximate the required elements of the real-world context sufficiently, the dependence on those professionals become less, and designers are able to conduct valid evaluations on their own

Concluding remarks

With the increasing potential of information technologies, along with the realization that information is crucial part of modern disaster management and response, the incentives for developers, researchers and innovators to develop new solution are bigger than ever. Over the years to come there will be an increase of tools developed. Tools that will have to be tested and evaluated, not only determine whether the function correctly according to their requirements but also in the real-life context to ensure adoption by professional emergency services and disaster responders. The options for such evaluations are however limited, real-life testing is unfeasible and exercises are costly. In particular considering iterative evaluation, the limited availability of professionals and the limited resources, an evaluation approach that allows non-professionals to test new tools in a light-weight environment that provides the requires realism is needed. Serious games can provide such an environment, as illustrated in this paper it is possible to design a serious game that allows repeated evaluations satisfying these requirements. The introduced design process allow developers to design a game alongside their design process and choices.

The potential of such game however goes beyond evaluation purposes alone. Serious games, tailored around a specific solution, or better a specific problem that a solutions aims to address, can also be used for research into that problem. Furthermore such games can be used a training environment for new tools as well as be part of dissemination strategy, for example by demonstrating the impact. An interesting observation made by working with professionals is that evaluation does not only hold importance for the developers of systems to assess the quality and adoptability of their developed systems. Evaluation provides an equal function to users with specific information needs and requirements. Evaluations enable professionals organizations to make a well-considered selection in software to be used (Jadhav et al. 2009). While the evaluation setting presented in this paper is designed to be used by developers of information systems, the availability of the appropriate evaluation environment would also allow the intended users (professionals) to test, assess and evaluate various information systems.

Acknowledgements

We would like to thank the Chair for Information systems and Supply Chain Management, from the University of Münster for their contribution. In particular we thank Daniel Link and Carsten Bubbich for their support during the game and deployment of GDACSMobile. We are grateful towards the staff of the Campus Vesta, and the Belgian First Aid & Support Team for providing volunteers and the facilities.

References

Abt, C. C. 1987. *Serious games*, (University Press of America.

Aldrich, C. 2003. *Simulations and the future of learning: An innovative (and perhaps revolutionary) approach to e-learning*, (John Wiley & Sons.

Alexander, C. 1979. *The timeless way of building*, (Oxford University Press.

Alexander, D. 2000. "Scenario methodology for teaching principles of emergency management," *Disaster Prevention and Management* (9:2), pp 89-97.

Atkinson, R. 1999. "Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria," *International journal of project management* (17:6), pp 337-342.

Comfort, L. K., Ko, K., and Zagorecki, A. 2004. "Coordination in Rapidly Evolving Disaster Response Systems The Role of Information," *American Behavioral Scientist* (48:3), pp 295-313.
Cross, N. 2008. "Engineering design methods: strategies for product design,").
Crowley, J., and Chan, J. 2010. "Disaster Relief 2.0: The future of information sharing in humanitarian emergencies," in *Disaster Relief 2.0: The future of information sharing in humanitarian emergencies*, HHI; United Nations Foundation; OCHA; The Vodafone Foundation.
Currion, P., Silva, C. d., and Van de Walle, B. 2007. "Open source software for disaster management," *Communications of the ACM* (50:3), pp 61-65.
Fiedrich, F., Gehbauer, F., and Rickers, U. 2000. "Optimized resource allocation for emergency response after earthquake disasters," *Safety Science* (35:1), pp 41-57.
Gagné, R. M., and Driscoll, M. P. 1975. *Essentials of learning for instruction*, (Dryden Press Hinsdale, IL.
Geurts, J. L., Duke, R. D., and Vermeulen, P. A. 2007. "Policy gaming for strategy and change," *Long Range Planning* (40:6), pp 535-558.
Jadhav, A. S., and Sonar, R. M. 2009. "Evaluating and selecting software packages: A review," *Information and software technology* (51:3), pp 555-563.
Jokela, T., Iivari, N., Matero, J., and Karukka, M. Year. "The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11," Proceedings of the Latin American conference on Human-computer interaction, ACM2003, pp. 53-60.
Kirkle, S. E., Tomblin, S., and Kirkley, J. Year. "Instructional design authoring support for the development of serious games and mixed reality training," The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), NTSA2005.
Landgren, J. 2007. *Designing information technology for emergency response*, (Department of Applied Information Technology, IT University of Göteborg, Göteborg University.
Link, D., Meesters, K., Hellingrath, B., and Van de Walle, B. 2014. "Reference Task-based Design of Crisis Management Games,").
Marsh, T. 2011. "Serious games continuum: Between games for purpose and experiential environments for purpose," *Entertainment Computing* (2:2), pp 61-68.
Meesters, K., and van de Walle, B. Year. "Disaster in my backyard: a serious game introduction to disaster information management," Proceedings of the 10th International ISCRAM Conference–Baden-Baden, Germany. T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T. Müller, eds2013.
Michael, D. R., and Chen, S. L. 2005. *Serious games:* (Muska & Lipman/Premier-Trade.
Newman, M., and Noble, F. 1990. "User involvement as an interaction process: a case study," *Information systems research* (1:1), pp 89-113.
Patton, M. Q. 2000. *Utilization-focused evaluation*, (Springer.
Pawson, R., and Tilley, N. 1997. *Realistic evaluation*, (Sage.
Perry, D. E., Sim, S. E., and Easterbrook, S. M. Year. "Case studies for software engineers," Software Engineering, 2004. ICSE 2004. Proceedings. 26th International Conference on, IEEE2004, pp. 736-738.
Red Cross Societies 2013. "World disaster report," *Geneva: International Federation of Red Cross and Red Crescent Societies (http://www. ifrc. org)*).
Rossi, P. H., and Lipsey, M. W. 2004. *Evaluation: A systematic approach*, (Sage.
Simon, H. A. 1996. *The sciences of the artificial*, (MIT press.
Sommerville, I., and Kotonya, G. 1998. *Requirements engineering*, (John Wiley & Sons, Inc.
Stapleton, J. 1997. *DSDM, dynamic systems development method* (Cambridge University Press.
Turoff, M., Chumer, M., Van de Walle, B., and Yao, X. 2004. "The design of a dynamic emergency response management information system (DERMIS)," *JITTA* (5:4), pp 1-35.
Van de Walle, B., Turoff, M., and Hiltz, S. R. 2009. *Information systems for emergency management*, (ME Sharpe.
Venkatesh, V., and Davis, F. D. 2000. "A theoretical extension of the technology acceptance model: four longitudinal field studies," *Management science* (46:2), pp 186-204.
Verschuren, P., and Hartog, R. 2005. "Evaluation in design research," *Quality & Quantity* (39:6),733-762.
Verschuren, P., Somers, N., and Daems, A. 1997. "The need for qualitative methods in agricultural economic research," *Tijdschrift voor Sociaalwetenschappelijk Onderzoek van de Landbouw* (12).
Wouters, P., van Nimwegen, C., van Oostendorp, H., and van der Spek, E. D. 2013. "A meta-analysis of the cognitive and motivational effects of serious games," *Journal of Educational Psychology* (105:2), p 249.
Zannier, C., Melnik, G., and Maurer, F. Year. "On the success of empirical studies in the international conference on software engineering," Proceedings of the 28th international conference on Software engineering, ACM2006, pp. 341-350.