



Project Start Date: 1.9.2016
 Project Duration: 30 months

D5.2 GAP Application Periodic Evaluation Report

Deliverable details	
Deliverable number	D5.2
Author(s)	Michael Cooke Sarah Curristan Melanie Labor
Due date	31/05/2018
Delivered date	30/05/ 2018
Reviewed by	Anne Holohan
Dissemination level	PU
Contact person EC	Carla Rocha-Gomes

History of Changes			
Date	Version	Person in Charge	Description
27/05/2018	V1.0	Michael Cooke	First draft
30/05/2018	V2	Michael Cooke	Revised draft
30/05/2018	V3	Michael Cooke	Final version



Table of Contents

TABLE OF CONTENTS	2
INTRODUCTION	3
METHODOLOGY	5
HEURISTIC EVALUATION	7
<i>Usability Heuristics (Nielsen)</i>	7
<i>Process Of Evaluation Using Heuristics</i>	8
<i>Limitations Of Heuristic Evaluation</i>	9
HEURISTIC EVALUATION RESULTS.....	11
COOPERATIVE EVALUATION	14
<i>Concept Of Cooperative Evaluation</i>	14
<i>Overview Of The Method Process</i>	16
<i>Limitations of Cooperative Evaluation</i>	16
<i>Designing Sessions</i>	17
<i>Recruitment Of Participants</i>	18
<i>Data Collection And Analysis</i>	18
<i>Running The Sessions</i>	19
SUMMARY OF EVALUATION RESULTS	21
CONCLUSIONS	27
REFERENCES	28
APPENDIX 1 DETAILED NOTES FROM COOPERATIVE EVALUATION SESSIONS	30
APPENDIX 2 SAMPLE INFORMED CONSENT FORM	48
APPENDIX 3 COOPERATIVE EVALUATION TASK LIST	51



Introduction

As discussed in D5.1, the primary objective of the GAP project is to develop a soft skills curriculum and training game for the delivery of the curriculum for international personnel involved in international missions for peacekeeping, peace-building, and conflict prevention. The curriculum is being implemented through a digital role-playing game, intended to be part of a broader blended skills learning approach, which will provide training on essential soft skills for the mission context. The nature of these soft skills has already been identified through the examination of a combination of existing training practices, the available literature on soft skills and peacekeeper training, and, most importantly, through direct consultation with GAP end-users.

The curriculum has been scripted with close reference to the experiences of those who have participated in overseas deployment for peacekeeping missions and close attention has been paid to the realism of the scenarios on and off base and how the interactions with international colleagues as well as locals should be conducted. What the GAP game in its current instantiation represents is a sample of that script text implemented as an interactive role-playing game for the purpose of providing early feedback to the game development team, Haunted Planet. The version of the game tested is a medium fidelity prototype that provides most of the functionality and partial but representative training context comprising of different scenes on and off the base. It has been developed in the Unity platform and is an advanced wireframe game. As a wireframe it differs from the proposed final version in terms of sound, graphics and animated elements and some interactive functionalities but nonetheless it fully supports independent game-play by an individual player with interactive dialogue elements.

The purpose of task T5.3 and this report, D5.2, is to provide formative evaluation feedback on the interaction design of the game to guide the further development of the user interface. The emphasis is primarily on the issues around usability, including interaction with the interface, understanding the content, following the narrative, use of language and managing the progress through the game. Feedback on the nature of the content is, at this point, of secondary importance. This is because it is only a partial implementation of the training content which is not enough to judge the overall effectiveness of the tool as a training support, and also the priority is to establish the optimal user-experience and usability results before further content



implementation is carried out to reduce the risk of post-hoc dissatisfaction with the final version of the game. Having said that, some of the issues discussed of a usability and user experience nature will interact with content aspects, which will be pointed out where relevant.

The section that follows will discuss the methodology used for this evaluation phase which includes both a user-centred, co-design method known as cooperative evaluation as well as an expert review method known as heuristic evaluation based on best practice around user interface design.



Methodology

The usability of a system is defined as “the extent to which a product can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO 9241-11, 1998). The aim of usability inspection is to assess product design and to ensure the product behaves as expected and is capable of meeting user requirements (Greenberg & Buxton, 2008). However, a perspective on usability that is overly product-oriented can be too restrictive as it emphasizes the need for a product to be usable, but not useful. Assessing the usability of a product cannot be confined to its ease of use; it must also satisfy its intended purpose and the needs of the end user (Bevan, 1995). For our purposes therefore, while the usability is our primary concern at this point, it cannot be separated from the subject matter that it supports on the platform. Usability problems may be caused or influenced by the narrative of the game and how the task-requirements of the game are perceived in their game-play context. Clarity or confusion, for example, in what is being asked of the player can be a function of script rather than purely interaction design issues. More realistically we have to consider the issues of game-play and user-experience in conjunction with usability. That is why we approach this evaluation by having users directly experience the game and provide their feedback on both the overall experience and the specific points where issues are discovered and require modification. This is also supplemented by expert evaluation by a usability expert but the primary focus and weighting is given to the end-user response to the game as encountered in the evaluation sessions. The player had not seen or played the game prior to these sessions and had a mixed range of game-play experience in general.

Usability refers to the degree to which the user can operate the functions of the system. It is not a singular property of a system; it consists of several attributes. First, for a system to be considered usable it should be easy to learn, so that users who have no experience with the system should be able to become proficient quite quickly. Second, the system should be efficient to such an extent that it increases productivity for the desired task. Third, using the system should be memorable in such a way that the user can easily re-engage with it after a period of inactivity. Fourth, a well-designed system should have a low error rate and where errors do occur the means for implementing recovery should be easily accessible. Finally, the



usability design should ensure that users are satisfied in their use of the system (Nielsen, 1994). In order to demonstrate usability a system should uphold these qualities.

Nielsen (1993) argues that there is a difference between usability problems, which refer to how the system should be operated, and problems of utility, which refer to what the system can do. The provision of an operational definition for usability problems before beginning the evaluation is important for identifying these issues. In this research, a usability problem was regarded as any aspect of the system that is difficult for the user to engage with, that slows the user down or is incompatible with the user's preferred method of working (Hertzum & Jacobsen, 2001).

Usability evaluation methods are looked upon as a collection of resources that need to be specially tailored to the specific demands of a project. Usability testing methods do not specify what aspects of a system to test, which group is most representative of the target end-users or how best to carry out the task of inspection. These decisions are at the discretion of the evaluator and there is wide scope for variation (Woolrych et al, 2011). Examining the usability of a system can uncover all manner of conflicting findings and user needs that can only be resolved through further analysis. Undoubtedly, some contradictions will remain and compromises will certainly be made. Ultimately, the purpose of usability engineering is to further design towards the best solution for the needs of users and their tasks (Nielsen, 1994).

The GAP platform offers is aimed at two categories of user, one being trainers and another being peacekeeping personnel preparing for deployment. As these groups are engaging with the platform for different purposes their usability requirements are distinct. However, for our purposes here we are focusing on the experience of the player, that is the trainee, due to the fact that it is part of an overall blended learning approach as opposed to a standalone training game. The trainer may provide prior instruction and subsequent facilitation or debriefing, or may sit alongside the player in order to discuss the decisions and choices made.

In considering the perspective of the peacekeeping personnel end-users, the usability of the GAP platform was assessed using a combination of heuristic evaluation and cooperative evaluation. Heuristic evaluation is an approach that provides a checklist of qualities of interest for assessing the usability of a new system (Nielsen, 1994). Cooperative evaluation is a form



of participatory or co-design approach that asks end-users to provide feedback on a prototype or system concept. It is a form of iterative design that uses feedback from the target user-group to improve the prototype's design and ensure that it is appropriately matched to end-user needs (Monk, Wright, Haber & Davenport, 1993).

Heuristic Evaluation

The heuristic evaluation approach was developed by Molich and Nielsen (1990). It is an informal method of usability evaluation that relies upon an experienced evaluator carrying out an analysis of the system against several principles of usability. The purpose of a heuristic evaluation is to detect usability issues in the system's design in order to address these during the iterative design process (Nielsen, 1993). There are no rigorous guidelines for how this method should be implemented and it is recommended that the use of these heuristics be supplemented by other usability assessment methods (Nielsen and Molich, 1990). This approach has been adopted in this research and further usability testing has been conducted with a representative test-user group.

Usability Heuristics (Nielsen)

Nielsen's (1994) research identified ten general principles of good interactive design. The first is the visibility of system status; a system should always inform users of its current status and activity by giving the user appropriate and timely feedback. Second, the system should demonstrate a good match to real-world activity; it should be appropriately tailored to users' real-world conventions and use familiar language and concepts. Third, the system should provide the user with control and freedom to carry out activities; if an error is made the user should have the opportunity to leave an unwanted state and return to a previous condition, to put it briefly the system should support 'undo and redo'. Fourth, the system should demonstrate consistency and standards among the language, actions and features that it displays. Fifth, the design of the system should discourage errors by eliminating error-prone conditions and providing confirmation messages for actions; the display of error messages is also useful in this regard.



Sixth, the system's design should minimize the user's memory load; by making the system's actions and options both visible and accessible users can rely on recognition for their activity rather than recalling necessary information as they perform different tasks. Seventh, the system should be flexible and efficient; it should provide the user with customized accelerators, an opportunity to perform frequent actions and key functions more quickly. Eighth, the design should be both aesthetic and minimalist; functions should be accessible and task-relevant but the amount of information that the system displays should not overwhelm the user. Ninth, the system should be capable of recognizing, diagnosing and solving errors; the user should be informed of any errors that occur in plain-language messages which should also suggest potential solutions. Tenth, where necessary, the system should provide help and documentation relevant to the user's tasks.

These heuristics provide a general guideline of qualities to note and assess. They aim to cover a large variety of problems that occur in user interface design. Additional components can be added that are specific to the platform of interest. Experience with these principles is preferable to ensure that they are correctly interpreted and applied. However, Nielsen (1994) advocates that they can also be used by inexperienced evaluators if the evaluation is supplemented with other usability testing protocols.

Process Of Evaluation Using Heuristics

The aim of a heuristic evaluation is to compile an overview of the positive and negative aspects of a system that is still in development. This method aims to identify usability problems and address them as part of an iterative design process. It is an intuitive and practical method that relies both on the evaluator's experience in usability assessment as well as common sense. As different evaluators can observe different issues it is recommended that the evaluation process make use of the outputs of at least three different evaluators, and optimally five (Nielsen, 1993).

As part of this process the evaluator examines the system independently to avoid biasing the evaluations (Hertzum & Jacobsen, 2001). Typically, the interface is examined once in order for the evaluator to become accustomed to its layout and features at a general level. Evaluators are expected to question design decisions and aspects of the system. The system is then examined a second time in further detail in order to assess particular components against the



ten usability principles, as well as any additional principles that may be specific to the system in question. Any issues that are identified are recorded by the evaluator and given a severity rating. Evaluators categorised each issue as ‘minor’, ‘mild’ or ‘severe’ depending on how the extent to which they perceived the issue to affect users’ engagement with the platform.

The evaluation process should take approximately one to two hours depending on the complexity and duration of the game. The independent report of the expert evaluator, who is not involved in the game development, is for our purposes here combined with a report on the cooperative or user-centred evaluation.. The result of the evaluation process is a list of usability issues that correspond to Nielsen’s principles of good usability design; while these issues may not present an obvious and immediate resolution the use of the principles provides a frame of reference for understanding why such issues would pose a problem for end-users (Nielsen, 1993).

Limitations Of Heuristic Evaluation

The advantage of evaluating a platform through Nielsen’s heuristics is that it is a flexible and quick way to identify usability problems. It can be implemented in the early stages of the development process and this ensures that findings can be incorporated into the system’s design in a timely manner in order to produce a more effective system prototype (Nielsen & Molich, 1990). Although these qualities are appealing, this approach carries several limitations that must be recognized and remedied where possible. In addition, the limitations of heuristic evaluation also indicate that it should not be undertaken as the sole means of usability assessment.

Firstly, it must be acknowledged that it is impossible to identify all potential usability problems within a system. For this reason, Nielsen and Molich (1990) propose that heuristic evaluations should be carried out by several evaluators, independently, and this feedback be then accumulated. Jeffries and Desurvire (1992) argue that in order for heuristic evaluation to be effective it must be carried out by evaluators who are not only experienced in usability testing but are also familiar with the domain in which the system is to be applied. Alternatively it can be combined with the results from end-user testing. This approach was observed by this study, and the system evaluations were carried out an evaluator with experience in both HCI research



and interaction design along with operational end-users during the cooperative evaluation phase.

Another potential limitation of this method is that different evaluators may differ in the severity ratings they attach to detected issues. This can present an issue for the reliability of heuristic evaluations (Greenberg & Buxton, 2008). As Hertzum and Jacobsen (2001) discuss, two evaluators assessing the same platform and using the same method of evaluation can detect different usability issues. Research has demonstrated that this effect is not due to the evaluator's level of experience or the extent of the usability problem. 'Good' evaluators and 'bad' evaluators are equally likely to overlook significant usability problems (Nielsen & Molich, 1990). They may also differ in the severity ratings that they attach to these problems. Severity can be judged in different ways. For example, usability issues could be ranked in terms of priority or they may be classified according to predetermined levels of severity.

Additionally, the validity of usability problems is an important consideration. Problems picked up through heuristic evaluation by a usability expert may not pose a problem for the user in their actual use of the system. Similarly, problems identified by users in usability testing may have been overlooked during heuristic evaluation. Jeffries and Desurvire (1992) add that the results of heuristic evaluation are more likely to reveal an extensive list of minor problems, and perhaps at the expense of revealing the system's severe usability problems. Resolving a list of minor problems may not be the most effective application of resources in comparison to addressing more severe system problems (Nielsen & Molich, 1990).

It is important to identify as many usability problems as possible, but as these may be numerous it is important to address the problems of consequence, explicitly problems that pose a significant difficulty for a user relying on this product in a real-world setting (Nielsen & Molich, 1990). This issue was addressed by ensuring that the evaluations were carried out by an evaluator with experience in usability testing as well as actual intended end-users who are members of the target audience of the application. In addition, the heuristic evaluation method was supplemented with a cooperative evaluation of the system in order to give greater insight on usability issues from the end-users' perspective.



Heuristic Evaluation Results

These design/evaluation Heuristics are taken from Nielsen (1994).

<p>Visibility of system status</p>
<p><i>Criterion:</i> The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.</p>
<p><i>Result:</i> The game is based on a narrative structure with no scenery changes within scenes. The progression of the game relies on the dialogue between characters, who are largely static, and the player. From this point of view the status of the system is well maintained through the dialogue. The possible addition of way markers or visual progress indicators to indicate to the player where they are in the narrative timeframe would not necessarily be an improvement due to the narrative and open ended nature of the game allowing for multiple choices to be made at various points. A significant point for improvement relating to visibility of status refers to the identity of characters and the identification of the character-speaking. As there are several characters per scene, especially in the Day1 scenes, it is possible to forget the names of individuals as there are no identifying features such as name tags which are commonplace on uniforms. Also, the greying-out of the character who is not speaking is confusing as this is the only character or scene element that changes. Normally for figure/ground distinction the element that changes against a static background is the focus of attention. Greying out in this context leads the user to believe he or she is the active voice. The operational role and rank of the main characters is not always explained.</p>
<p>Match between system and the real world</p>
<p><i>Criterion:</i> The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.</p>
<p><i>Result:</i> While the game in its current version is basic in terms of its graphics and sound, the script elements and scenery depicted match with the target scenery (base and market place) and deployment conditions. Gameplay is by its nature is removed from the real-world but the GAP game uses the professional language of military and police relative to the deployment context (Middle East). It is neither too generic so that it could be anywhere, nor too specific to any particular locality. Some phraseology is used that represents the deployment context language</p>



of the locals. However there are some idioms or slang phrases that may not be familiar to personnel who have good English but lack familiarity with certain phrases used by native English speakers. Also, some of the dialogue in Day 1, pertaining to the motivations of the characters and the meaning of the mission may not be realistic to an introductory meeting between colleagues or commanders.

User control and freedom

Criterion: Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Result: The user is offered very few control features for managing the game. There is no capability to a) select a scene to start with, b) to navigate backwards in the scene to either attempt another alternative decision or simply to remind oneself about some information presented earlier if forgotten, c) to save one's progress at a given point and resume from that point at a later time, d) to skip scenes or scan forward to a later part in a scene, and e) to quit the game from within the app. The scene in the market place does allow for panning to explore further but the user is not guided towards the presence of this affordance and how that affordance is implemented. Navigation arrows left and right would remedy this.

Consistency and standards

Criterion: Users should not have to wonder whether different words, situations, or actions mean the same thing.

Result: The system is mainly consistent with general PC/Mac operations as well as tablet. Users familiar with these devices OSs will be familiar with the games operations, although there is no internal app functionality to quit or save the game and no drop-down menus (or any kind of menu apart from the cog icon representing "main" menu and "resume" function). While the function for panning the screen in the market place was invisible it does conform to the convention of side-scrolling with a mouse or fingers. One point of inconsistency exists between the day 1 scenes (dorm and canteen) and the market place. In the former the user does not select with whom to speak. A character approaches and the system controls the dialogue. In the market scene it is for the user to decide with whom to initiate dialogue and there is consequently some potential for confusion as to how that scene should progress sequentially between



characters (or whether or not it matters) given the difference in how the interaction is structured in different scenes. The conventions are not explained from the start. Also, the downward arrow used to exit the instant messenger, while intuitive, is not consistent with how a messaging app is exited. This is not necessarily a problem within the context of this game as the apparent affordance of the arrow is perceived.

Error prevention

Criterion: Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

Result: The user may inadvertently click on the wrong dialogue or action choice and wish to correct this. However, backward navigation or “undo” is not available.

Recognition rather than recall

Criterion: Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Result: The user is required to remember some significant information which can be forgotten as the game progresses such as names of characters and roles. Name tags on the uniforms would address this as well as more frequent reference in the dialogue text to what characters' roles are. Also, locations should be better indicated, such as when Maria returns to the dorm in the evening. The pace of the instant messaging on the phone in the night dorm scene when Maria receives a call from her sister is too fast and the affordance to scroll back up is not sufficiently visible. The user then is reliant on memory to follow the train of the conversation. The scroll capability within the phone (which essentially means scrolling an app within an app) should be made clear.

Flexibility and efficiency of use

Criterion: Accelerators — unseen by the novice user — may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.



Result: There are no short-cuts or accelerators present, either in player mode or admin mode. Accelerators may not always be relevant during game play as they might interfere with the learning experience of a structured narrative, but for game management and review purposes the ability to navigate quickly to specific points in the game for review or further instruction would be necessary. Also, saving game progress and choosing starting point at the outset (main menu) would be important to add.

Aesthetic and minimalist design

Criterion: Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

Result: Much of the dialogue in Day 1 is inconsequential and does not advance the narrative of the game or address any obvious learning outcomes. The graphics are very too basic in this version where the characters do not express changing body language, facial expressions or other indicators of mood tone.

Help users recognize, diagnose, and recover from errors

Criterion: Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Result: The opportunities for error are minimal as the main interaction with the system is in terms of dialogue with the characters. Errors are mainly related to game play decisions but not

Help and documentation

Criterion: Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Result: there is no help or documentation available in this current version.

Cooperative Evaluation

Concept Of Cooperative Evaluation



As discussed above, the cooperative evaluation approach was developed by Monk, Wright, Haber and Davenport (1993) as a quick, efficient and collaborative user-centred method for evaluating computer systems and various stages of their development. This approach looks to involve users in the design process by asking them to provide feedback on a system prototype. Users are given a task list that is intended to take them through the main functionalities of the system. They are instructed to think aloud, or verbalise their thoughts, perceptions and decision-making as they complete these tasks, and their interaction with the system is observed. The observer records any problems that the test-user encounters in using the system or any verbal or behavioural feedback they provide. This data is then analysed to give an indication of the system's usability issues.

The Landauer-Nielsen formula stipulates that a group consisting of five test-users should be sufficient to assess the usability of a platform (Cockton and Woolrych, 2001). This rule of thumb is based on a cost-benefit analysis performed by Nielsen that demonstrates that a group of five test-users is adequate for the successful detection of approximately 80% of usability problems within a system. Beyond five test-users, the cost of including additional test-users outweighs their contribution to the evaluation process. Nonetheless we believed that considering the variation in nationalities involved in the project and the realities of international deployment it was beneficial to have some representation from all participating countries, military and police branches, and a mix of genders so as not to assume the homogeneity of the user group.

This claim of diminishing returns has been met with some scrutiny as it assumes that all usability issues are equally likely to be detected, and this does not reflect the realities of testing. Usability testing is guided by a set of tasks and completed users who demonstrate a variety of individual differences. Thus, it is unlikely that this assumption can hold true in all cases. Specific system features, testing techniques, the complexity of the task list and the level of problems of interest all influence the amount of problems an assessment will return (Spool & Schroeder, 2001).

A study by Lindgaard and Chattrachart (2007) propose that there is no significant relationship between the number of test-users that undertake usability analysis and the amount of problems found. Rather, the amount of tasks a participant is given correlates with the amount of problems



that the user will identify. For this reason special consideration was given to the task list to ensure that it provided a comprehensive demonstration of the functionalities of the platform. Having said that, in the case of the GAP application, the number of separate functionalities is limited and the tasks are determined mainly by the narrative direction of the game. It was also expected that while the number of purely usability issues related to the platform would be relatively limited, a greater set of issues might become apparent when considering the different operational, training, and prior game-play experience of the players, interacting with the narrative of the game's script, particularly in terms of the dialogue between characters and speech/action choice selection.

With these issues in mind, the first phase of the cooperative evaluation of the platform aimed for a test-user sample of eight participants.

Overview Of The Method Process

Each user was paired with an observer and given a list of tasks to complete using the GAP platform. The users were encouraged to give comments on the platform and to ask questions about its functionalities. Users were informed that the observers may ask them questions during the evaluation regarding the tasks that were being performed and if the functionalities were suitable for end-user needs regarding major emergency management. Finally, users were reminded that **their performance on the task was not being evaluated** and that the purpose of the study was to assess the usability of the platform in its current state and find specific problems. Users' participation in the task was audio recorded for the purpose of capturing notes, the anonymised details of which are included in appendix 1. Completion of the task list game play took approximately 40 mins per user, depending on their own individual pace of game-play.

Limitations of Cooperative Evaluation

As with heuristic evaluation, there are some limitations to this method of usability testing. Bevan (1995) points out that the cooperative evaluation presents a situation where users are artificially guided through the system as opposed to exploring the system on their own. In this way, their use of the system may not fully reflect how they would naturally engage with the



platform. For example, encountering a minor problem may hinder success on a particular task but may not indicate a real issue for the user's use of the system.

Another issue is that task time cannot be compared among users and this is often a useful means of examining the efficiency of a platform. It is also worth bearing in mind that the player is being observed while playing which may reflect the reality for some of the real training situations that are instructor facilitated, but not all GAP gameplay will be observed. The difference in flow and freedom to experiment by the player would be expected to differ from solo game-play to observed game-play.

Designing Sessions

In order to produce reliable measures for testing the quality of use, the context for the evaluation should resemble the context of the envisaged real-world use of the system. It is important that evaluators understand this context and how the platform should be integrated. During evaluation key aspects of this context should be replicated where possible so that the test-user can interact with the system in a way that closely resembles the purpose it is intended for (Bevan, 1995). For the purpose of efficiency, and given the fact that this game can be played in any location, it was considered unnecessary for the evaluation sessions to be carried out in an actual military training facility, especially given the early stage of the prototype development and the fact that participants were from numerous different countries.

A study by Cockton and Woolrych (2001) has demonstrated that variability in the tasks given to test-users can influence what problems are identifiable. They advocate that the task list should be modified throughout usability testing in order to ensure that the system is being approached from different perspectives and is thoroughly tested. For this purpose the task list for participants was standardised and printed for all participants, as was the list of follow-up interview questions.

As Bevan (1995) argues, it is difficult to perform an accurate assessment of a system's usability when evaluation is taking place in an artificial environment or with scenarios that the participant would be unlikely to encounter. For this reason the evaluation task list was developed around the gameplay function designed to support a script based on a non-disclosed



Middle-Eastern environment and peacekeeping scenario reflecting recent and topical contexts and peacekeeping personnel test-users were instructed to act as if they were a peacekeeping personnel training for deployment which is the case for many of them. The task list was developed as a means to prompt the users to engage with all functionalities offered by the platform but equally to provide a mock-up of the context of a peacekeeping/building deployment, including interacting with other personnel on based as well as the environment outside of the base. In this way test-users use of the system aimed to reflect the envisaged use of the platform.

Sample content was created by the researchers to populate the GAP platform and provide the test-users with game scenarios they could work through to test the platform's functionalities. The central scenarios included a female CIMIC officer's first day of arrival at her deployment base including meeting members of her team and commander as well as encountering cultural differences with colleagues of different nationalities.

Recruitment Of Participants

As mentioned, the operation of the GAP platform is envisaged as being carried out by peacekeeping personnel players. The recruitment of participants for the cooperative evaluation of the system aimed to utilise peacekeeping personnel participants who are typically trained to perform these roles but without having previously being deployed, in order to test the suitability of the platform against its intended purpose. Eight participants were recruited in all from five countries (Ireland, UK, Finland, Portugal, and Poland) including 5 males and 3 females.

Data Collection And Analysis

Test sessions with the participants were audio recorded with the informed consent of each individual participant. Any unexpected behaviours demonstrated by participants in their navigation of the system were noted by the observers alongside a note of when this behaviour occurred during the narrative for comparison with the recordings. Comments made by the test-user were also recorded with respect to each individual task.



Full transcripts of the recorded data were not required as this is considered a rapid evaluation method. The recordings were used to facilitate note taking. Instead the usability and script problems were noted during the session and the recording were analysed subsequently to capture any detailed that may have been missed.

Running The Sessions

The cooperative evaluation sessions were run between the 20th and 22nd of March 2018. The duration of each test-session was approximately an hour and a half including the briefing, informed consent, gameplay and de-briefing. The user was welcome and accompanied to a quiet meeting room with plenty of natural light. Refreshments were provided.

The player was instructed as to the purpose of the GAP project and the purpose of the evaluation emphasising that it was not them or their knowledge that was being tested but the system's development with the intention of feeding back this information to the developers in order to improve it. Usability was explained as the main goal. The participants were handed an informed consent and instructional document to read and sign and were given the opportunity to ask any questions. After they were happy to sign, they were asked to initiate the game and follow any on-screen instructions.

One of the requirements of the method is that the end-user is asked to think aloud, provide a running commentary of what they see, what they are doing, and how they are reasoning about tasks and decisions. They allows us to not only identify when a problem occurs and what the nature of it is, but it also gives us some insight into the reason why it is a problem in the first place. When there is a divergence between the developers mental model and the operational end-users mental model and that is apparent in the system image, then this contradiction can be highlight through the thinking aloud protocol.

While it is unnatural and often difficult for a user to maintain a flow of thinking aloud, the co-evaluator gently prompts the player with questions such as “what are you seeing now?”, “what do you think is happening”, “how would this normally work”, “what do you think you should do now?”, etc. In circumstances where the user got stuck then the co-evaluator would intervene



and help them through the task reassuring that this is a point where the system needs improvement rather than a reflection on their abilities.

Ethics

The cooperative evaluation process relies on the active participation of volunteer end-users. Because of this, it is a method that carries ethical considerations that other usability measures do not. It is important to ensure that the test-users are aware that their participation is entirely voluntary and that they are free to withdraw from the research at any point. Test-users volunteered to take part in the evaluation and no incentives were offered. Each test-user was given a briefing sheet detailing the purpose of the research and an informed consent form to sign (see Appendix 2 for a blank sample) in order to indicate their willingness to participate and to have their session audio recorded. Participants were assured that they were free to withdraw from the study at any time. All participants were informed that their data would remain confidential and accessed only by the relevant researchers, and also that their data would be anonymised and that pseudonyms would be used in any report.



Summary of Evaluation Results

In this section we have summarised the overall results from the evaluation including the combined expert evaluation and cooperative evaluation results. As mentioned before, end-user reports from the cooperative evaluation take precedence over the interpretation of the expert in usability but generally the best-practice principles do not diverge from the end-user's experience.

It should be noted that the initial headline results from the evaluation sessions were made immediately available to the game developers on the 23rd of March to allow them to start working on improvements. This included specific recommendations for redesign where possible. There is more detail available from the cooperative evaluation sessions in Appendix 1.

Results are structured here corresponding to the requested structure of the game developers to allow them to most efficiently work with them.

Gameplay and navigation issues
Generally players found their way through alright. One player didn't understand that the purple text was clickable, eventually discovered it. Nobody immediately understood that you can pan across the market scene. This was discovered by accident every time. The "... " option in response to the fruit vendor was understood by about half of participants as meaning "don't say anything". Other cases players didn't understand it was an option at all, a technical glitch or oversight.
Recommendation
Provide visible navigation arrows at the market scene to suggest panning affordance.
Remove the "... " option and replace with "say nothing", or equivalent text.

Number of dialogue choices
It was generally agreed that the number of dialogue choices was adequate although there as some acknowledgement that the reality of decision-making may lie in the space between the available choices.



Recommendation

Either increase the number of choices to include a greater number of nuances (5-6) or else leave as is recognising that during instructional debriefing the student player will discuss their decision rational with an instructor.

Points of confusion in the gameplay

There was almost universal agreement that it is difficult to tell who is speaking and who is being spoken to. The greying out of characters doesn't clarify this because they greyed out character is the only part of the scene that changes when there is speech hinting that it is they who are speaking. Speech bubble or some other visual indicators were requested to more clearly indicate who is speaking. Also, players lose track of characters' names and who is who. Especially when characters reappear. Tove, for example, goes unnoticed despite the fact that she is introduced at the beginning. She is not on screen contingent with the moment that her name is used. The Indian "cutlery" scene was commented on as not being clear that the people in the background at the table were listening because there is no visual reaction from them. Name badges are conventional on uniforms. They would help if they were visible in the game. Tove was thought at one point to be Maria having taken her hat off.

Recommendations

Have visual link between the text of a speaker and the speaker themselves, ensure that they are on-screen at the moment.

Use a clearer indicator of who is speaking, such as speech bubbles, rather than greying out the silent character.

Use some visual identification of characters such as uniform name tags.

Progress through black-screen dialogue and progressing through game

Progressing through the options presented was immediately perceived in most cases, exception being the individual who did not understand the purple text as a clickable object. Mobile phone was understood, but not immediately obvious who it is on the other side. One point made about the mobile phone is that the messages coming in are too many and too quick, they scroll off the screen and then you can't see earlier ones if you have forgotten what they said. The arrow pointing downwards was understood as progressing to next scene as opposed to a part of the



phone. All players were just drawn to the arrow without hesitation despite the fact that it does not appear elsewhere. Characters with the yellow aura were perceived as clickable. Some did try to click other characters in background but it was mostly understood that only the yellow ones are interactive.

Recommendation

Use a device other than different colour text to indicate where text is clickable/interactive. For example, enclosing it within a box (button) or simply providing verbal instruction.

Provide more perceptible visual cue that the mobile phone text can be scrolled within the app, or alternatively slow the rate of message delivery.

Identification of characters, especially the player's character.

It was not always clear who she is (Maria), and she and Anna are often confused until a clearly male (Nowak) appears and allows for differentiation. While the game presents alternative options, players didn't all tend to want to go back and try other options. Some did but most were happy to let the game progress with the options taken. However, going back in the game was desirable for the purpose of reminding themselves of who Maria is and where she's from. The early text is frequently forgotten.

Recommendation

Allow for the affordance to return to a previous state/screen to remind the player of earlier information. Use name tags and clearly identify who the active speaker is.

Understanding who is speaking

Eventually users learn the convention but it seems to always be confusing at the start as previously discussed.

Recommendation

Use of speech bubbles could overcome this.

Confusion of characters moving in and out of screen

Often people being spoken to are not on screen and then appear after the utterance is made. This was confusing. It was asked why could we not see more than two characters at a time



when there are more than two characters implied in the conversation. The way in which the characters pop in and out briefly to make a quick remark was found to be slightly comical.

Recommendation

Allow for all active characters to be visible and not simply “pop up” on screen and withdraw to make an utterance.

Indicate visually that static figures in the background are actively listening (overhearing) as this makes a difference to how the dialogue would normally progress.

Understanding Italicised actions.

Although with reference to the distinction from the dialogue (italicised choices with the * were seen to be different but still choices nonetheless). Most players understood whenever there was a choice presented and went along whether it was a verbal act or a physical act.

Recommendation

In this case no problems were encountered. If in future scenes there may be ambiguity between verbal and physical action this may perhaps be highlighted either by setting the conventions from the outside or providing verbal/visual indication in the scene.

Do users understand scrolling the explore screen (panning the market place)?

No, not at all. Scrolling was discovered by accident and in some cases after mild frustration. Players would have understood the importance of scanning a scene tactically, and if this were mentioned it might have cued them for it, but as it was they thought what they could immediately see is all there was. Direction arrows were also requested to indicate screen panning as an affordance.

Recommendation

Provide either a textual cue to pan across the screen to scan the full scene or use direction arrows to suggest this.

Do users understand clicking on outlined characters in the explore scene?

Yes, that was mostly clear. Occasionally someone tried to click other people but mostly understood that only the highlighted ones were clickable. Clicking non-yellowed characters



was most likely when the affordance to pan was not perceived or when the system froze at that point.

Recommendation

The conventions on clicking on characters was clear in the market scene, but this convention was different from Day one, where you do not click on the characters. This could be made more consistent. Users would like to choose with whom to speak and in which order.

Game value

The market scene was considered much meatier as it involved operational decisions and dealt with the subtle process of intel gathering. The Day1 scenes were less interesting, too much small talk and not enough substantial conversation.

The Indian cutlery scene was considered useful and worthwhile as it dealt with a dilemma. The other topics were considered less important, in some cases referred to as philosophical discussions (why are we here, world peace, etc). Players recognised that these are relevant issues in the broader scheme but found it less credible as something discussed immediately upon arrival. Practical actions such as zeroing the weapon were considered realistic, also getting to know people, but discussing the meaning of the mission and personal motivations less so.

Visuals

The visuals can be improved but not completely panned by players, and the colour graphics considered highly favourable. More of that requested.

A visual introduction showing scenery would have been welcomed as opposed to text on a black screen. It's necessary to give a flavour of the context.

Recommendation

Add visual scenery at the start to set the context of the region

Sound

Soundscape would also be welcomed. The bleeping sound with the text was distracting and annoying in most cases although it did alert to the fact that dialogue was happening.

Recommendation



Remove the bleeping for scrolling text.

Add soundscape

Understanding Emotion Tone of scenes and dialogue

Currently there is nothing indicating the emotional tone of the dialogue apart from the text itself. Body language and facial expressions are absent and missed as a result. Non-English speaking players struggled with some of the idioms and didn't always understand the presence of a joke/sarcasm. The fruit vendor was almost always seen as someone to avoid because of his facial expression, but then turns out to be friendly. That's not necessarily a problem (lesson about judging books by covers) but mostly people have the same expression in every scene no matter what they are saying.

Recommendation

Add visual cues to changing mood or tone to make it clear that there is a reaction corresponding to a natural reaction to a joke, insult, confusion, etc. Body language and facial expressions are considered important by users.

Gender Aspects

This issues goes beyond a usability issue but it is of relevance for the overall experience of the gameplay. The game is perceived as being female-centred. While this is not seen as a problem in and of itself, especially when the purpose is to allow men in particular to experience the differences that a woman may encounter in these situations, It can also be of value to allow the player to be male so as to understand better the issues about, for example, approach a woman in the market, or shaking a woman's hand.

The content of day one was perceived as more feminine than male. As one female end-user reported (Liliana), woman are more concerned with how they feel in a situation and men are more concerned with what they have to do. Whether or not this is accurate it would be beneficial to orientate the content away from discussion about the meaning of the mission, personal views of being a peacekeeper, and more towards what needs to be done, such as intelligence-gathering and conflict resolution.



Conclusions

Overall, from perspective of both evaluations (expert and end-user) the game was seen to work well and with few and minor problems from an interaction design and narrative point of view. In most cases the players found the game to be enjoyable and would welcome the opportunity to continue playing other days and scenes, particularly those involving operational/tactical work and decision making. There was a genuine sense of wanting to know what happens next.

As a training tool it was also considered potentially valuable especially alongside a blended learning approach where it may be possible to reflect on decisions made with an instructor/facilitator. There was some commentary in the cooperative evaluation about whether some of the options were realistic or not, but it was agreed that they had a role to play given that there were multiple choices. Even “wrong” or unusual answers have learning value based on their consequences.

The game was considered to be a female orientated game, not only because the main character and companions are female but some of the dialogue reflect what are perceived by players as womens’ issues. This is not necessarily a negative point as it can be useful to a) train women about how to be successful in working in a male-dominated environment, especially in a very different society and b) to train men to be able to empathise some more with the particularly challenges that a woman has to face when, for example, dealing with males in a traditional and conservative society. However, some of the in-base scenes were consider by both male and female players to be what would traditionally be thought of as issues corresponding to womens’ sensibilities, such as world peace and making the world a better place, and in general the dialogue around motivation and meaning of the mission were not considered relevant and reduced the value of Day1. The more “tactical” issues involved in Day9 were considered much more relevant and enjoyable but at the same time they do allow for gender differences in terms of how the characters might be approached depending on whether the player is male or female.

Apart from the low-level usability issues, the proposed future visual were considered to be a potential major improvement over the line drawings currently present. Also the use of a soundscape was considered highly beneficial as an addition to the next version.



References

- Bevan, N. (1995). Measuring usability as quality of use. *Software Quality Journal*, Vol. 4 (2), pp. 115-130.
- Bucher, H.-F., & Schumacher, P. (2012). *Interaktionale Rezeptionsforschung*. (H.-J. Bucher & P. Schumacher, Eds.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-93166-1
- Firsching, J. (2012). Eye-Tracking Studie zur Facebook Chronik. Retrieved from <http://www.futurebiz.de/artikel/eye-tracking-studie-zur-facebook-chronik/>
- Greenberg, S. & Buxton, B. (2008). Usability evaluation considered harmful (some of the time). *Proceedings of the SIGCHI Conference in Human Factors Engineering* (pp.111-120). New York: ACM DOI: 978-1-60558-011-1/08/04.
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität 1 Grundannahmen und Vorarbeiten. *Mensch & Computer 2003. Interaktion in Bewegung*, 187–196. doi:10.1007/978-3-322-80058-9.
- Hertzum, M. & Jacobsen, N.E. (2001). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, Vol. 13 (4), pp. 421-443.
- ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals, Part 11*. ISO/IEC.
- Jeffries, R. & Desurvire, H. (1992). Usability testing vs. heuristic evaluation: Was there a contest? *SIGCHI Bulletin*, Vol. 24 (4), pp. 39-41.
- Laugwitz, B., Schrepp, M., & Held, I. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In *Mensch & Computer 2006 - Mensch und Computer im Strukturwandel* (pp. 125–134). Oldenburg Verlag.



Lindgaard, G., & Chattratchart, J. (2007, April). Usability testing: What have we overlooked?. In Proceedings of the SIGCHI Conference on Human factors in Computing Systems (pp. 1415-1424). New York: ACM.

Monk, A., Wright, P., Haber, J. & Davenport, L. (1993). Improving Your Human-Computer Interface: A Practical Technique. New Jersey: Prentice Hall.

Nielsen, J. (1994). Usability inspection methods. Conference Companion on Human Factors in Computing Systems (pp. 413-414). New York: ACM.

Nielsen, J. (1993). Usability Engineering. San Diego, CA: Academic Press.

Nielsen, J. & Molich, R. (1990). Heuristic evaluation of user interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 249-256). New York: ACM.

Prümper, J., & Anft, J. (1993). Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel. Software-Ergonomie.

Spool, J., & Schroeder, W. (2001, March). Testing web sites: Five users is nowhere near enough. In CHI'01 extended abstracts on Human factors in computing systems (pp. 285-286). New York: ACM.

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. Decision sciences, 39(2), 273-315.

Woolrych, A., Hornbaek, K., Frokjaer, E. & Cockton, G. (2011). Ingredients and meals rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. International Journal of Human-Computer Interaction, Vol. 27 (10), pp. 940-970



Woolrych, A. & Cockton, G. (2001). Why and when five test users aren't enough. In Proceedings of IHM-HCI 2001 Conference (Vol. 2, pp. 105-108). Toulouse, France: Cépadèus.

Appendix 1 Detailed notes from Cooperative Evaluation Sessions.

Player #1

Pseudonym: Barbara.
Age: 28.
Gender: Female
Nationality: Portuguese
Experience game player: Yes (e.g., GTA/Sims)
Previous deployment: No.
Culture awareness training: No
Platform – iMac OSX 10.13
Graphics
Nice Graphics. Catching attention
Graphics are simple but that is good. Can be improved but not essential.
Proposed colour graphics are good.
Sound
Sound could be softer. Too loud. (volume reduced).
Sound draw attention to the dialogue but can also be distraction.
Ambient soundscape would be beneficial
Idioms
Not understanding “no mint on pillow”
Understanding of narrative
It's clear that Maria is trying to understand a new unfamiliar place
Never been deployed but seems to be realistic meeting other people from different nations (canteen scene)
Interesting that there are these cultural difficulties. It's realistic.
“looks like Morocco”.
Amount of reading is not too much.
Pace of game is fine.
Understanding characters
Unclear about which character is which (e.g., misidentified Maria)



The person talking should be different from the others (highlighted, figure/ground distinction) Noted in Canteen and in Market.
Correctly identifies sister as person texting
Understanding emotional tone and humour
Clearly understood from text alone what the emotional context was, but only from text. (e.g., handshake scene)
Humour was understood when and where used.
Body language/facial expression could be helpful but not necessary to be high detail. Overall body expression would be sufficient.
Understanding navigation
Understood the purpose of arrow to end SMS conversation and scene
Understood need to click on person in market to progress game
Did not understand scene panning function on screen. Prompted that perhaps screen could be moved, then understood how to move screen side to side with mouse.
Believes there is a sufficient number of dialogue choices per scene.
Training value
Interesting as training tool. We do not know what to expect. We are not aware of obstacles we will find, this is a good way of understanding.
Gender aspect helps to fit the character, helps to identify with another female.
Game value
“I could play this for hours”. Interesting as a game.
More interesting in the market. Day 9 is more interesting as you are actually “working”.
Recognises value of blended learning approach, in conjunction with theory training.
Would like to play more. Looking forward to next version. Engaging.
Best feature
Script writing is good, as if by someone experienced in deployment
Worst feature
Not clear who is talking
Miscellaneous
There needs to be some content related to preparation for returning home.

Player #2

Pseudonym: Mirosław.
Age: 31.



Gender: Male
Nationality: Polish
Experience game player: Yes. Role player Games
Deployed. Never deployed.
Culture awareness training: Yes, classroom based.
Platform – iMac OSX 10.13
Graphics
Not good. “Quite lame”. Would not be enticed to download based on graphics quality.
Proposed new colour graphics would be a sufficient improvement over current.
Sound
Did not notice use of sound. Until pointed out afterwards. “thought it was something from the street”. Didn’t recognise it as feature.
Soundscape would be important for atmosphere.
Idioms
“Translator” should perhaps be “interpreter”.
Some slang words are good to use. Nothing confusing.
Understanding of narrative
What am I doing in a military installation an odd question. Motivation questions, meaning of mission questions too philosophical.
A good idea to let someone show you surroundings. Tour ends at canteen.
Normal activities are represented well in game (on base).
Understands that some options are “traps” in the gameplay. Trick questions.
“I don’t need information about what she did back home or why she is here”. Perhaps better to just ask if she has been here before.
Understood arrival of SMS message from home.
Surroundings are like many countries. Everyday life (Market place).
Understanding characters
Understood person not greyed out as person speaking
Looks like a Polish flag on uniform (Capt. Novak).
Mark “looks like a local”.
Understands the CIMIC officer role only after talking with Capt. Novak. Not clear from earlier.
Unclear who “new person” in canteen is (commander Novak returning but this is unclear)
Understands message from home based on hairstyle of picture.
No issue understanding who to identify with. Main character is yellowed.
Understanding emotional tone and humour



Discussion of people suffering “blah, blah, blah”. Very philosophical. Unrealistic for a canteen scene.
“even soldiers want to have something more” in response to “ <i>couldn’t ask for more</i> ” response. “Could be better but I’m doing fine” might be more realistic.
Understood purpose of “shaking hands” dilemma. Cultural sensibilities.
“I would build some barrier between me and Mark”.
“ <i>Of course, we are perfect</i> ” “I assume this is a joke.”
Not a good idea to send too much information via messenger for security reasons.
Understands concerns from home from messages.
Predicts it will be hard to talk to locals
Understands emotional context with woman in market, and that she has been encountered before. Perceives facial expression. Not a good idea to ask what happened to her face (but still a good dialogue option for training).
“I’m a little bit worried about clicking second character but I will try”. Due to facial expression, seems hostile.
Clearly understands the emotional tone of conversation through text only. Not matched by non-verbal cues or expressions.
Understanding navigation
Understands navigation in general (moving from scene to scene, selecting options).
Understands purpose of arrow on phone to end conversation and move to next scene.
Understands that characters surrounded in yellow outline can be interacted with.
After being offered fruit from vendor, third option is “...”. Don’t know what this means as an option. Assumes it means to not respond. This is unclear.
Discovers by accident that the screen moves but recognises that this is typical for video games.
Understands to use “cog” icon to go to main menu.
Can’t locate icon or menu choice to quit game. Tries F4. Familiar with Linux. Directed by interviewer to select “quit” from OS menu.
Sometimes wanted to navigate backwards in order to “undo” options but none available.
No “save” option. Save would be beneficial.
Saving/quitting, play management needs to be improved. Quit option from within application should be present.
Pace of game is similar to that of a book, but that is ok.
Training value
The purpose of dilemmas (choices) is clear. They are often trick questions which can be useful.
Has good learning value, cultural features are well covered.
Some answers are obviously more appropriate than others but would like to option to explore through navigating back the outcomes of different choices.
Overall training value of game is well appreciated.
Gender
Carpet seller has only one wife. Perceived to be “untypical of the culture”.



Used to playing male characters. Didn't really notice any differences in most cases, although some scenes such as market place playing a woman made a difference in tactical approach.
Most characters are female, noted.
Game value
Game script is quite good. Clearly written by an experienced person (deployment).
Game style is ok, reminiscent of Japanese style games.
Graphics need to be improved.
Appreciates the multiple-choice aspects of the narrative.
Wondering what would happen next, so still interested in continuing playing after having finished. Looking forward to more.
Best feature
Jokes were good. Good for learning because you don't harm anyone in the game, but you can learn from the experience from playing.
Worst feature
Graphics are not sufficiently engaging.
Miscellaneous
Some glitches noted at several points (flicker on screen).
Doesn't like tablets in general. Prefers to use keyboard & mouse.

Player #3

Pseudonym: James. Age: 33. Gender: M
Nationality: Northern Ireland
Experience game player: Yes, Role playing, SIMS.
Previous deployment: No
Culture awareness training: Yes
Platform – iMac OSX 10.13
Graphics
Graphics are ok but facial/bodily expressions are missing.
Colour graphics would be an improvement along the lines of those presented.
Sound
Scrolling text sound not appreciated. Unnecessary and distracted.
Soundscape would be important and valuable.
Idioms
Assumes the Arabic text means hello.



As a native English speaker no issues with slang or expressions.
Understanding of narrative
Scenes understood and context generally understood.
Player scans market scene noting different characters, as per training.
The “...” response to the fruit vendor is not understood as meaning “say nothing”. Assumed to mean “insert free text”.
Doesn’t perceive how to pan across the screen. Tried clicking various non-yellowed characters. Directed to try dragging with the mouse to reveal other characters
Understanding characters
Didn’t get that Tovi is another character. Assumed it was Maria with hat off.
Sender of message assumed to be a significant other. Possibly a boyfriend/girlfriend. Concern from home is clear.
Reference to Mom and Dad clarifies that it is sister. Not immediately clear.
Name badges would help keep track of who is who.
Understanding emotional tone and humour
Perceives informality of introduction between Maria and Tovi.
“trying not to judge” is perceived as a judgemental statement.
Canteen scene with the Indians clearly perceived as a test of cultural sensitivity and how to handle it.
Understands that the options to indirectly acquire intelligence through small talk is a beneficial option. Never directly ask about trouble but approach in an indirect way, slowly build up. Good that this is supported in the script.
Lacking facial expressions and body language.
Fruit seller seen as hostile, but then understood not to be. That’s not unrealistic.
Understanding navigation
Would be better to have a wider range of more nuanced choices in dialogue.
Navigation is simple, no option to get lost.
Multiple pathways is good. Going backwards is not available but not required. Follow the course of the game.
Difficulty finding the options for saving, main menu and quitting. Needed direction.
Pace is ok, reading is a little slow due to the pace of dialogue appearing on the screen, but reading text as a requirement is fine.
Pacing for the phone messages was too fast. Earlier messages disappeared off screen and forgotten. Didn’t perceive option for scrolling back.
Training value
More nuanced answers in dialogue could perhaps be accommodated. The “real” answer someone would realistically give is not necessarily included in the list of options. “Something in between those options”.



Low entertainment value but useful as a training game so serves its purpose.
Very useful to have as part of a blended learning structure. The ability to discuss your responses with an instructor would be beneficial.
Gender
Comes across as a female orientated game. A little confusing to play a female role, but that is a positive from training point of view.
On the other hand, it would also be a good option to sometimes play as yourself using your own name as the central character.
Was conscious of playing a female role, as a male. Give some good insights.
Game value
For its training purpose it's good, but wouldn't play it for recreation. Not a lot happening. Not very entertaining.
Did learn something from it, confirming or disconfirming pre-conceived understandings.
Best feature
Simplicity. Doesn't need to be over complicated. Focuses on the content.
Worst feature
Limited options of responses in dialogue, and at least some body language is required to enrich the dialogue.
Miscellaneous

Player #4

Pseudonym: Borislav Age: 43. Gender: Male
Nationality: Bulgarian
Experience game player: Yes, but not since 10 yrs.
Previous deployment: No
Culture awareness training: Yes
Platform – iMac OSX 10.13
Graphics
Graphics are good enough but compared to other familiar games not very appealing.
Proposed future colour graphics will be much improved.
Not so much the graphics from an aesthetic point of view, but for understanding body language and facial expressions.
Sound
Background sounds requested. Soundscape would be welcomed.
Sound accompanying scrolling text is ok.



Real voices would be better for background, but still need text in case things are not clear from voices.
Idioms
Not clear what is meant by “dirty look”.
Inshallah understood, but perhaps not by all players.
Understanding of narrative
Somewhere in the Middle East. Generic scene.
Difficulty remembering data from introductory screen. Didn’t understand that Maria was a lieutenant.
Easy to lose train of conversation.
Not clear that other characters in background are listening to the conversation (overhearing) because they are static.
Not clear who the various groups in the base, with differing expertise. Or who the term “groups” is referring to – other military or NGOs, on/off base.
What does “Disagreements off base, united on base” mean?
Good to share resources on base, but good for player to know that the environment is not equal in every respect. E.g., Sauna. Boundaries.
Doesn’t understand the options in response to the Indian personnel dialogue.
Understanding characters
Not clear who is speaking at any given moment. Person greyed out is
Not clear that Mark is a local. Presumed to be.
Not clear who Capt. Novak is.
Doesn’t recognise Tove when she reappears in canteen.
Different colours for different people takes a while to associate
Understanding emotional tone and humour
Unusual that the conversation tone is so open. Normally used to orders with structures.
Not getting too friendly with locals is advisable.
Differences between nations explained in a humorous way. Different habits, etc. Clearly humour and good natured.
In military environment jokes are commonplace. The joke relating to the Indian personnel is not considered so harsh.
Understands sender of message as Maria’s sister.
Concern from home is understand, and considered normal.
Doesn’t understand meaning of “...”. Eventually assumes it means not to do or say anything.
Good sense of humour in the game.
Good approach to attaining information from informants without asking too directly. Important for intelligence gathering.
Not clear that the static characters are participating.



Understanding navigation
Would prefer to be able to navigate backwards to retrieve narrative information, when details are forgotten.
Would prefer to be able to experiment with different dialogue options going through the game.
Clicking on arrow on phone not clear. Had to be pointed out.
In market has to think about how to proceed. Assumes that must click on an individual in yellow outline.
Doesn't know how to pan across market screen. Instructed to scroll left and right to find other carpet seller.
Recognises "cog" symbol for main menu.
Not able to save game.
Not able to find menu to quit game within app. Has to be directed to do so from OS.
Not clear that the purple text affords clicking.
Training value
Good for training intelligence gathering aspects as well as cultural awareness.
Gender
Game is not perceived as gender-specific. There is increasing trend to have more females but the game is not female-centred.
Not sure most Bulgarian males will accept female role.
Game value
Entertaining, but much more value from the market scene. Much more interesting.
Best feature
The storyline goes smoothly. Good pace.
Worst feature
Not being able to understand how to scroll screen in market scene. Not being able to return back to remember who people and groups are.
Miscellaneous
Two days in the game is not enough to have a good judgement of its value.

Player #5

Pseudonym: Liliana.
Age: 59.
Gender: Female
Nationality: Polish
Experience game player: No. Very limited computer experience



Previous deployment: Not military deployment but has been abroad
Culture awareness training: Yes.
Platform – iMac OSX 10.13
Graphics
Graphics are acceptable for the game play.
Proposed colour graphics would be improvements.
Sound
Scrolling text sound distracting.
Soundscape would be beneficial.
Idioms
“No mint on pillow”, not clear.
Climbing the ladder not clear.
Finger painting
Find your bearings, not understood.
“Collection” is a strange expression when referring to people. A little negative.
Going to hell
Tug-of-war
Language needs to be simplified.
Understanding of narrative
Not sure whether the situation outside refers to the climate (sun) or the conflict.
Some of the dialogue in day one too long, particularly in relation to the motivational aspects.
Lacking some instruction about important local knowledge, e.g., checking boots for scorpions, where land mines might be present, how to be careful in these areas, etc.
Too much small talk in the canteen. Speaking too much about serious and deep issues on first meeting, before lunch.
Why only Finnish can use sauna is not understood. It should be for all. Doesn't understand the issue.
The messenger chatting is too long.
Is patrol mixed nationalities? Not clear.
Looks like Jerusalem market. Typical middle east scene.
Showing some scenery would be beneficial at the start to give some visual context of the environment.
Understanding characters
Not clear that she is CIMIC
Capt. Novak should perhaps make it clearer what his chain of command will be.
Should be explained that Mark is local.
Not clear who is speaking in canteen.



Doesn't recognise Tove in canteen. Name tags would help.
Background characters in canteen are not known.
Not clear that it is Maria back in the dorm.
Understanding emotional tone and humour
It is unusual to ask "why are you here", especially at a first meeting. Perhaps in a more informal, social setting.
Anne perceived as not being in good humour. Maria thinks about promotion but not something you speak openly about.
Hospitality aspect is missing, towards new person arriving.
Perception, that using the word "collection" indicates that Anna is in bad humour.
As different people from different countries are speaking through English they will on both sides make an effort not to assume the intention behind an expression is necessarily negative.
Initially understood the point about everyone working in harmony as serious, but learns then that it is a joke. Missing non-verbal cues.
Should be careful not to push Milo to apologise, she won't not do that as not being his commander doesn't give her the right. Still, it is useful as content to see how the reasoning of the player works.
UN "secrets" is understood as a joke.
Fruit seller looks to be hostile.
Understanding navigation
Moving from screen to screen not a problem.
Understanding and selecting dialogue options clear.
Clicking on arrow to end phone conversation is intuitive.
Doesn't understand how to navigate the market scene. How to choose the person to speak to is not clear. Interaction is different from other scenes as in before you don't select the person to speak to.
Scrolling left and right in market is not perceived. Needed to be explained.
Training value
(without having experienced the market scene) the most valuable scene is the Indian personnel scene as it is an example of potential conflict.
Game has value in general for training in avoiding conflict, including dealing with colleagues as well as locals.
Gender
Men may prefer a more dynamic game with more action. Orientated towards softer topics, with perhaps more feminine orientation, but men with experience of games may expect something more interactive.
Dialogue more feminine in content. More about feelings, men talk more about actions or things (e.g., cars, etc.).



More women characters than men.
Game value
Day one scenes are less interesting. Perhaps game should start with a more significant event, such as recounting a recent attack or accident, to hook attention from beginning.
Immersive. Possible to becoming Maria.
Best feature
Commander Novak. Very realistic, welcoming.
Worst feature
Game froze during play. Conversation with sister too long.
Miscellaneous
Gameplay was interrupted by game freezing during play. Not able to resume.

Player #6

Pseudonym: Mike.
Age: 25
Gender: Male
Nationality: Irish
Experience game player: Yes
Previous deployment: No
Culture awareness training: Yes
Platform – iMac OSX 10.13
Graphics
Graphics are good enough. No issues reported with the current quality apart from the lack of body language.
Proposed colour text will be a big improvement
Sound
Current use of scrolling text sound not of any particular value, but not annoying either.
The use of a soundscape will greatly enhance.
Idioms
No issues experienced as native speaker
Understanding of narrative
Zeroing weapon a realistic option, for preparedness.
Good understanding of what was happening.
Smooth transition throughout the game



Understanding characters
No evidence of confusion between characters. Not always clear who is speaking but learned the internal convention (greying out).
Understanding emotional tone and humour
“Trying not to judge”, implies judgement.
Use of humour is good.
The emotional/tension scenes are clear and easy to understand.
The scene with call from home is important as reassuring loved ones is necessary and how to communicate that effectively but securely is worthy of training.
Understanding navigation
No major problems encountered in terms of general navigation and dialogue with the game characters.
The “...” option with the fruit seller not understood.
Panning across the screen not understood for the market. Would have expected better capability to choose which character to speak to first. Direction arrows would have facilitated panning to display other characters.
What I really like about Day 9 is that you can choose who you speak to and in the order you prefer. That is different from day 1.
Panning of screen in market not clear.
Training value
Considered to be of high value for training in soft skills, cultural awareness and cooperation with others. Putting theory into practice. Learns a lot from more experienced colleagues but this allows some application and refining of understandings.
Should have more tactical or conflict-reduction scenes. Day one is currently too taken up with small talk.
Gender
Important that men can play female roles to understand better the experience and viewpoing of other genders.
Game value
Enjoyable game play, although day 9 much more interesting as more tactical issues apparent.
Best feature
Overall a good game, with more interesting content such as Day 9 would enhance interest in more continued play.
Worst feature



Some minor issues of navigation including the market scene panning and not understanding the “...” option with the fruit seller.

Miscellaneous

Player #7

Pseudonym: Harri.
Age: 50s.
Gender: Male
Nationality: Finnish
Experience game player: Limited
Previous deployment: Yes
Culture awareness training:
Platform – iMac OSX 10.13
Graphics
The visualisation, “I don’t see the difference between this and a powerpoint presentation”, given how static the graphics are.
Expecting richer graphics for understanding non-textual aspects.
Market scene better than Day 1 scene with richer information. Different characters.
Speech bubble would be necessary to understand more clearly who is speaking.
Name tags.
Sound
Sound for scrolling text is unnecessary. Annoying.
Soundscape would be beneficial.
Idioms
Inshallah is more than just an expression, it is meant literally, “it is up to God to decide whether there will be a tomorrow.
Understanding of narrative
There should be a formal or official navigation training.
Difficulty understanding who is speaking, between the person greyed out or not greyed out. A speech bubble would be better.
Does not understand that the person greyed out is the listener.
Necessary to keep all speakers on screen at the same time.
“Irish are the only nationality that is mentioned”
Presumes Tove is Norwegian. Not clear about nationalities.
Why are there no voices, talking heads, etc. I have to imagine everything from reading.



Important to note that the characters will have had real induction and orientation training rather than assuming that they are picking up all the relevant information from casual encounters with other personnel.
Understanding characters
Not understanding the difference between Anna and Maria.
Assumes Anna is Finnish, but then assumes it is Norwegian flag.
All characters should be present when involved in dialogue.
Unsure about the rank differences between personnel. Usually for international missions there is a first-name policy.
Unsure whether it is Anna or Capt. Novak who is speaking. Difficult to tell who the active speaker is.
Who is Mark?
Confusion in terms of who is Anna and Tove.
Doesn't recognise Tove when she reappears.
Doesn't recall who Maria/Anna are.
Name tags are necessary, on uniform.
Not clear from the beginning that this is a UN mission.
Milo comes to the conversation without introducing himself.
Not clear who the messenger from home is. "Who is this person?" "How I answer depends very much on who is asking."
Understanding emotional tone and humour
"Never lie", in response to carpet seller.
Effective in terms of understand the conflict issues, e.g., scene with the Indian troops.
Humour us clear but perhaps intentionally ambiguous.
Understanding navigation
Would like to be able to go backwards to remember how arrived at this point. Not to replay the game but for keeping up with the narrative.
Would like to go back to clarify who said what.
Unclear about the meaning of "... " with the fruit seller.
Going back to main menu confusing. Not obvious what to do.
Not possible to return to a previous scene.
"New game", does it mean play again, or for a different user, not clear which.
Pacing of the game, would like to get into the substance of the game more quickly. Too much introductions.
Training value
The ambiguity about shaking hands, "a good test".
Questions in the canteen scene quite relevant, e.g., "what is the town like"?
The first encounters with other personnel are realistic, e.g., canteen scene.



More clarity about character nationalities required.
Good that the market place dialogue options include opportunities for indirect intel gathering rather than asking straight out.
Discussion with the carpet seller ends at a point when it's becoming interesting from an intel gathering point of view.
The conversation with the carpet seller ends just when it's getting interesting.
Focus should be much more along the lines of Day 9.
Gender
Playing the role of a different gender is beneficial.
It would be useful also to have men's perspective, for the purpose of understanding how to interact with local women appropriately.
Game value
Inconsistent in terms of engagement. Day 9 was much more interesting than Day 1.
Best feature
Day 9 was particularly interesting. Good potential for exploring the relevant soft skills
Worst feature
Missing body language or facial expressions to judge reactions.
Day 1 was weak compared to Day 9. Should compress Day 1 and emphasise more Day 9 style issues.
Miscellaneous

Player #8

Pseudonym: Elaine.
Age: 30s.
Gender: Female
Nationality: Northern Ireland
Experience game player: Yes
Previous deployment: No
Culture awareness training: Yes
Platform – iMac OSX 10.13
Graphics
It would be important to have facial expression and perhaps judging situations based on non-verbal communications in some scenes.
Not being able to see what people are doing with their hands is an issue.



Sound
Background sounds missing. Will be welcomed in next version.
Liked
Idioms
If you offer a hand they take an arm. “interesting” expression to use in the context as there might be many amputees.
Understanding of narrative
Would like more time to read introductory text at the start. Static text over scrolling.
Doesn't understand what CIMIC means. Not explained from the outset.
Good that she has highlight that there's a doctor at the community centre
Would have liked to have seen the sauna developed as an “unusual” storyline, potentially valuable as training material – dealing with the conflict.
Understanding characters
Unclear who is the active speaker.
Significant that Mark is not in uniform.
Perception, from talking with Mark, that former appointees to the role were perhaps not so sensitive to the cultural differences.
Stereotyping is an interesting dimension.
Not clear from the outset who the caller is on phone (IM scene). Perceived familiarity, intimacy. “partner or friend”.
At the end, not really able to step into the character's shoes. How much prior experience she has, what language abilities etc.
Mark didn't feature much. Did I annoy him? What is his role?
Understanding emotional tone and humour
Player actively “deducing” the level of conflict from the dialogue with Mark.
Perceived context of tension with reference to the sauna dialogue
Perceiving a sense of priorities from Mark, e.g., learning language.
Don't know “whether I should smile and laugh or excuse myself”, wrt the Indian scene.
Interesting, that we're trying to put out a united front off-base but perceived that on base there is some conflict.
There is a culture where you can say whatever you like, but with consequences (in response to Indian scene).
The fact that they're having this conversation in the vicinity of other cultures is problematic, but useful as a training device.
Understanding navigation
Discovered “scrolling scross”, it's good that she can scan the scene. Not obvious but quickly discovered. Tactically important.
No significant issues.
Not apparent how to return to main menu.



Not clear who is speaking and what the colour-coding and greying-out signified.
Highlighting around the figures in the market was good.
Being able to scroll around the market made it more interactive. Other scenes should contain this element.
Not overly wordy, good use of text in dialogue.
Training value
Would like to have played it from a “devil’s advocate” perspective, to explore the consequences of different decisions.
Not clear what the purpose of the family back home scene.
Would have liked to have feedback over what transpired after market scene, e.g., reporting back what was observed and gotten feedback from commander.
An opportunity to close the issue with Milo would be important, rather than leaving it hang.
More than one right answer in a number of scenes.
Getting you to think about how verbal and non-verbal responses are significant.
You learn a lot from it. You question yourself.
Gender
Quite aware of character’s gender. Gender as an issue was very apparent.
As a woman, the sauna issues “threw me”, and the issue of shaking hands was significant, but at the market scene (tactical nature of scene).
Game value
Market scene much more enjoyable because it’s more interactive and the player has much more control.
Best feature
Good quality script.
Scenes involving actual conflict requiring de-escalation are beneficial but more of them required.
Interactivity at the market scene and the choices between the characters and the dialogue. Would have liked that in the canteen. More control.
Worst feature
There was a lot of reading. Would have preferred static text or slower scrolling for introduction screen.
Some second guessing and somethings would have chosen naturally.
Cultural issues could come out a bit more. Indian “joke” scene required closure.
Lack of choice of action in canteen.
Miscellaneous



Appendix 2 Sample Informed Consent Form

Informed consent form GAP T5.3 Usability Evaluation Basic information about the GAP game and instructions



Thank you for agreeing to help with this study.

Today we are going to evaluate the usability of a soft-skills training game called GAP. GAP is a role-playing game that is intended to help you during pre-deployment training to develop soft skills along with your other training activities in the classroom. The aim of this study is to find out how easy GAP is to use by people like yourself, the intended end-user. We want you to try to use it and help us find out what problems are there in the current GAP version and how it could be improved through better design. All suggestions from you, either while playing the game or afterwards, are welcome.

We will give you some standard tasks to do using GAP. These tasks are linked with the basic functioning of the game and its interface. We want to find out how easy it is to use, to identify any difficulties the design of the game presents to you and how they can be fixed. This is not in any way an assessment of your abilities in terms of soft-skills or your knowledge of IT or gaming in general. Please take your time and if you encounter problems or difficulties these are not your fault and we will help you through them. Finding problems is very useful information for us.

We're particularly interested in situations in which GAP encourages you to make errors in selecting commands and misleads you about what it will do. We are also interested in extra commands that would make the system easier to use. If you find yourself stuck in a position where you don't know how to progress, we want to know this and come up with ideas about what information could be provided in the game's next version to help.

To get this information we shall use a simple technique. This involves three things:

1. We ask you to think-out-loud as you do each task, telling us what you are intending to do and how you are trying to solve each task, which actions you think might be appropriate and why, and what you think the game has done in response to your commands and why. Think of this as you giving us a running commentary on what you are doing and thinking. This allows us to see whether the system has been designed with your way of thinking in mind.
2. Whenever you find yourself in a situation where you are unsure about what to do or what effect commands might have, ask us for advice. If you ask us what you need to know we will suggest things for you to try but if you get really stuck we'll explain exactly what you have to do.



3. In addition we will ask you questions about what you are trying to do and what effect you expect the actions you perform (clicking, scrolling, dragging and dropping, etc.) will have. This is simply to find out what problems there are with the system. During our conversations, we want you to voice any thoughts you have about parts of the system which you feel are difficult to use or poorly designed. While you're doing this we'll be noting down the problems you mention but in case we miss any we are going to audio tape our conversation. This recording will be anonymous and treated in confidence. Remember it's not you we're testing, it's GAP. We are interested in what you think so don't treat this as an examination. Treat it as a structured discussion about GAP. Please feel free to say whatever you think about the system and the tasks you're given to solve. There are no right or wrong suggestions.

Your rights as an evaluation participant.

You are helping us improve the GAP system as a volunteer and we thank you for your time which we understand is very valuable. At this point we would like you to be aware of the following.

- You have no obligation to participate in this session, it is purely voluntary and you can withdraw at any time, even during the session;
- We do not require from you any personal information apart from your rank, role and experience level;
- Your inputs including what you do and what you say will be treated as confidential and will be anonymised. Only GAP project personnel from TCD and NUIM who are running the sessions will be able to see your comments before being anonymised. Nothing will go back to your own organisation.
- We ask you for your permission to video or audio record your responses to allow us to take better notes. If you prefer to not be recorded we will proceed with written notes;
- You may say anything you wish or that you feel to be relevant but we are only asking you for information about how you experience this version of the game and how it could potentially be improved.
- If you wish to have a copy of the original recordings this will be provided to you.
- If after the session you feel that you no longer wish your input to be used by us then we will destroy any notes and recordings.
- We will produce a report on the quality of the GAP system which will be based on aggregated anonymous inputs of all participants. We intend to use for the purpose of illustration some useful or interesting quotes from participants or descriptions of exemplary actions or situations but we will ensure that they are not identifiable to you or any other individual. The anonymised report will be used as input for the GAP software developers to improve the game.
- We are happy to provide you with a copy of this report on request.

If you have any questions please do not hesitate to ask any of the evaluation team and when you are sure that you understand what is being asked of you today would ask you to sign the declaration below. This is not a contract and it does not indicate any obligation on your part. It simply indicates that you have received a full briefing and you are happy to proceed.



Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS: Michael.cooke@mu.ie

Participant's signature:

Signed _____

Date _____

Researcher's signature:

Signed _____

Date _____



Appendix 3 Cooperative Evaluation Task List

T5.3 Cooperative Evaluation Task List



1. Open the application
2. Commence playing the game until the end.
3. At the end of the first pass, start a new game.
4. Continue play making alternate selections to previously.
5. When asked to, please pause the game
6. When the game is paused please resume play.
7. Please pause the game again and go to the main menu
8. When at the main menu please resume game.
9. Please quit the game.