Heuristic Evaluation of a Mobile Game Developed to Help Battle the Pandemic

Sami Pohjolainen
OASIS Research Unit
University of Oulu
Oulu, Finland
sami.pohjolainen@oulu.fi

Juho Mattila
Ikune Labs
Oulu, Finland
juho.mattila@ikunelabs.com

Jarkko Tuovinen
Ikune Labs
Oulu, Finland
jarkko.tuovinen@ikunelabs.com

Mikko Rajanen
INTERACT Research Unit
University of Oulu
Oulu, Finland
mikko.rajanen@oulu.fi

Leena Arhippainen
INTERACT Research Unit
University of Oulu
Oulu, Finland
leena.arhippainen@oulu.fi

Paula Alavesa
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
paula.alavesa@oulu.fi

Abstract— As the COVID-19 storms the globe, there are many efforts to battle it. This paper looks at the development of one such an effort in a form of a mobile game. The game is a 2D arcade game called Unus Terra. It is intended to be entertaining but aims to influence people to partake in social distancing through several mechanics. An early version of the game is evaluated by five usability and user experience experts using heuristics and a few other selected methods. In this paper, we present the results from these evaluations along with our findings from using this type of evaluation method. The use of heuristics is considered from the perspectives of academics and practitioners.

Keywords—behavioral change, COVID-19 pandemic, heuristic, mobile game, serious games, social distancing, usability

I. INTRODUCTION

Usability is a well-established concept in the human-computer interaction (HCI) field, but relatively young in the gaming context [1-3]. Some researchers have reported that playtesting is still the most popular method to test usability in games. Tools used by developers tend to be rather traditional and include approaches such as questionnaires, interviews, and observations [3, 4]. Game developers also use expert evaluation methods, but more sparingly, as they are less popular despite being rapid, cheap, and effective ways to identify usability methods [5-7]. The key point of using expert evaluations is to catch usability issues early in the development process when they are less costly to fix [5-7]. Typical expert evaluation methods include heuristic evaluations, expert panels, cognitive walkthroughs, and focus groups [3, 4]. While researchers have developed heuristics lists for games, developers often design heuristics to fit their own needs [3, 5]. In this paper, we use a specific heuristic evaluation method designed for mobile games [8].

A mobile game called Unus Terra started its development during the first half of 2020 This was a time when the COVID 19 had caused major social distruption but the pandemic still had many unknowns, several studies showed an increase in gaming and media consumption during the lockdowns [9-11]. This was because the interactions between people had been reduced and they started to consume a lot more games, both analog and digital variants [12]. In response, organizations, such as WHO and along with industry leaders in interactive entertainment started to promote gaming as a positive thing to do with initiatives such as PlayApartTogether [13]. These initiatives promote the importance of social distancing. It had been established that asymptomatic carriers played a critical role in the spread of the virus, which furthered the need to promote social distancing and other safety guidelines [14]. Several games have been produced that deal with pandemics. Notable ones include a board game called Pandemic from 2008, a card game called Sauchenquartet from 2018, and the Plague Inc. video game from 2012 [15-17]. These games provide a mixture of entertainment and education, as each includes knowledge about pandemics. The current pandemic has caused governmental organizations and others to fund research and development of tools to combat the outbreak. This paper looks at a game called Unus Terra that was developed through an EU-funded project with a specific intention to promote social distancing during the pandemic. These types of games are often called serious games, as they have a function that goes beyond pure entertainment value.

This paper primarily contributes to the HCI and usability research by providing information on the development and evaluation of a mobile game. The secondary contribution is on the development efforts of serious games that deal with behavioral change outcomes. The focus is on expert evaluations of an early version of the game. Five experts evaluated the game by using a tailored heuristic method. The main goal of these evaluations was to find problems with usability and gameplay. Assessing the entertainment value was not the purpose of these evaluations, but planned with user tests taking place a version later. The findings from evaluations were analyzed and summaries are reported in this paper. This includes our assessment of using the method as a part of the evaluation process. The second section focuses on the background and methodology used in evaluations. The third section describes the game development. The fourth section describes the setup and process for the evaluations. The final section provides a discussion and conclusion on the key points.

II. BACKGROUND AND METHODOLOGY

Developing serious games exists in the intersection of creativity of game development and the formality of productivity software development. Serious games can be defined in many ways but are essentially a group of games developed to be more than just entertainment [18].

Usability and user experience (UX) have gained a lot of attention recently in the game development context to the extend of becoming a distingushable field of game user research targeting spesifically player experience appose to user experience. This also includes related techniques such as game analytics, player tracking and gameplay spesific heuristics. [19-22] The increase and variety of players and

games during the last decade have pushed the focus of the game developers to develop for their specific target audiences. Such a shift in focus requires user research as well as usability processes and methods. Usability may provide a competitive edge [23], and increase the willingness of players to buy a game [24, 25]. With serious cames, usability issues can greatly impact the desired outcomes [26].

Better usability may also result in increased user satisfaction and productivity, even though these particular usability concepts have a slightly different meaning in the gaming context when compared to applications used for productivity, as playing games is voluntary by nature. Regardless, usability and the quality of the user interface are still very important for players [27]. As the competition in the gaming market has become increasingly fierce, good game usability has become a necessity. The professional game reviews and online gamer communities ensure that a game with bad usability gains a bad reputation, which will damage its sales [25, 28]. To introduce better usability into the game development context, literature proposes many different methods [5, 19, 27, 29, 30, 31, 32].

Heuristic evaluation is a widely used usability evaluation method in the software development industry [20]. It is considered the main discount usability method as it does not require much infrastructure, time, or money [20]. It involves a group of experts evaluating a software product by using a list of so-called heuristics that provide rules for the best-practice design. A single person can also perform a heuristic evaluation, although using more than one evaluator is recommended to find more problems [33]. Regarding the usage of heuristic evaluation in the gaming industry, there have been surveys indicating that the heuristic evaluation in the gaming industry is overrated in literature when compared to actual reality [7]. Most of game developers do not seem to use heuristic evaluation despite its stated benefits in the literature. Reasons often include concerns over lack of return for investment, time and other resources, such as knowledge, expertise or suitable heuristics. [34] Furthermore, many companies consider the game usability heuristics being difficult to use and not suitable for the games or genres that they develop [7]. Most of the game developers that employ heuristic evaluation prefer to use their own heuristics or an adapted version of Nielsen's list [5, 7]. There have been calls from research and practice for more usable, genre-based, and platform-specific usability heuristics. The research is starting to address this issue. [5]

Nielsen and Molich published the heuristic evaluation for user interfaces in 1990 [35]. Their method was designed with a computer interface in mind. Laitinen found that while heuristics developed for HCI were useful in finding problems in video games, they were not designed for games, and therefore lacked in comprehensiveness [36]. Other heuristic methods have been developed in very specific domains, but they still retain features from Nielsen and Molich [35]. The design objective of a game is typically quite different from a productivity application [1, 29]. This is also true for serious games [26]. However, several heuristic evaluation methods have been developed for games.

In 2002, Federoff highlighted the case for specific game-related heuristics that should also consider fun rather than pure utility [29]. This model included three categories: game interface, game mechanics, and gameplay. In 2004, Desurvire, Caplan, and Toth introduced Heuristic Evaluation of Playability (HEP) specifically intended for games [37]. Later, Desurvire and Wiberg acknowledged that while HEP was

useful, it was also limited in its approach [20]. This led to the development of PLAY, which is a list of heuristics that could be modified for a specific type of game and situation [19].

Recently, there has been a development towards a modular approach that allows researchers and practitioners to customize their heuristics for games [5, 38]. There is a general trend towards modularity with usability and UX evaluation methods in general. Standardized UX questionnaires are a good example of this trend, as seen with the modular version of UEQ and meCUE [39-41]. A similar modular approach will likely be popular, as they do not rely on a predefined set of heuristics, but can be customized for different types of games, platforms, and development phases [5, 38]. For this study, there was a specific heuristic set developed for mobile games by Korhonen and Koivisto [8]. This set includes three modules that are Game Usability, Mobility, and Gameplay. Heuristics factors in those modules are presented in table I. There is also a multi-player version of this set, but it was not relevant to this game, as the game does not include multi-player features [30].

Ponnada and Kannan used the same heuristic when evaluating four mobile racing games [34]. They used simple yes or no answers for each heuristic. While this might be suitable when evaluating multiple games, it does not provide sufficient feedback for designers or developers in a specific case. It also does not consider that issues with a specific heuristic factor might not be present throughout the game, but only apply in very specific cases or situations. Instead of using a simple answer, severity ratings for usability problems were adapted from Nielsen [42]. These ratings and their descriptions are presented in table II. This allows evaluators to score each heuristic factor.

The problem with heuristic evaluations is that evaluators tend to find different sets of usability problems. This is called the evaluator effect [43]. To address the issue, researchers give several practical recommendations [44]. Most recommendations apply to the evaluations done with the game in this study. The first recommendation outlines the importance of having domain knowledge. This means that the evaluators should be are familiar with the type of system. Second, there should be more than one evaluator who conducts tests independently, so that more issues can be detected. Third, consolidation of the severity ratings is suggested to make it easier to prioritize problems. Finally, perfect reliability is not important or even a worthwhile goal.

Research suggests using no more than five users or evaluators with usability tests. This is because 75% of usability problems can be identified with five evaluators. Additional testers will increase the number of findings, but there will be significantly diminished returns. There are of course cases where and when using more testers is beneficial, for example, when dealing with highly distinct groups of users or when approaching the product launch. [45]

III. GAME DEVELOPMENT

This section presents a serious game developed to encourage social distancing. It is targeted to adolescents and adults aged 12-35 and works on mobile platforms (Android OS and iOS). The game was initially intended for younger audiences.

The game is targeted against COVID-19. It is not all serious and the game should be also an entertaining arcade game. As the pandemic has matured, the game is likely to have a wider appeal, which means that the future development activities will not focus on the single pandemic, as it was originally conceptualized.

TABLE I. HEURISTICS FOR EVALUATION

	Heuristics used in the evaluation
Tag	***Game Usability Heuristics [8]***
GU01	Audio-visual representation supports the game
GU02	The screen layout is efficient and visually pleasing
GU03	Device UI and game UI are used for their own purposes
GU04	Indicators are visible
GU05	The player understands the terminology
GU06	Navigation is consistent, logical, and minimalist
GU07	Control keys are consistent and follow standard conventions
GU08	Game controls are convenient and flexible
GU09	The game gives feedback on the player's actions
GU10	The player cannot make irreversible errors
GU11	The player does not have to memorize things unnecessarily
GU12	The game contains help
Tag	***Mobile heuristics [8]***
MO01	The game and play sessions can be started quickly
MO02	The game accommodates with the surroundings
MO03	Interruptions are handled responsibly
Tag	***Gameplay heuristics [8]***
GP01	The game provides clear goals or supports player-created goals
GP02	The player sees the progress in the game and can compare the results
GP03	The players are rewarded, and rewards are meaningful
GP04	The players in control
GP05	Challenge, strategy, and pace are in balance
GP06	The first-time experience is encouraging
GP07	The game story supports the gameplay and is meaningful
GP08	There are no repetitive or boring tasks
GP09	The players can express themselves
GP10	The game supports different playing styles
GP11	The game does not stagnate
GP12	The game is consistent
GP13	The game uses orthogonal unit differentiation
GP14	The player does not lose any hard-won possessions

TABLE II. SEVERITY RATINGS FOR USABILITY PROBLEMS, ADAPTED FROM [42]

Rating	Description
0 None	I don't agree that this is a playability problem at all
1 Cosmetic	Need not be fixed unless extra time is available on the project
2 Minor	Fixing this should be given low priority
3 Major	Important to fix so should be given high priority
4 Serious	Imperative to fix this before the game is released

A. Hackathon and Funding

Ikune Labs participated in EIT Digital and Ultrahack Data Against COVID-19 deephack which aimed to provide solutions that would potentially save lives through innovative use of data and technology [46, 47]. In the event, Ikune Labs designed and developed the initial prototype of the Unus Terra game and won the hackathon that was held in May 2020 [48]. The initial concept and an early prototype were made during that time (Fig. 1). The team beat 145 other projects that took part in the contest. After the win, Ikune Labs partnered with the University of Oulu, Finland and Politecnico di Milano, Italy to develop the game concept further.

During the development, Ikune Labs acted as a business champion and activity leader. In this role, the company created a business model, marketing strategy, and marketing plans for the game. Ikune Labs also planned and executed the user tests together with University of Oulu, Finland. Heuristic evaluations were a part of this activity. In addition, the company is responsible for the development roadmap for the game in 2021.

Fig. 1. A screenshot of the main menu from the Hackathon version of the game

B. Game Design

The game provides a timely story where the player helps a group of undercover scientists discover new pandemic threats to the world. The task of the player is to prevent viruses from spreading. The main activity of the game involves the player selecting threats from the world map (Fig. 2A) and collecting virus samples by defeating viruses in matches at different locations. (Fig. 2B). The idea is to collect research material, find new virus strains, and advance the research. The research provides rewards to the player and works as a progression indicator. This allows the player to unlock new playable characters, enemy virus strains, and other in-game bonuses and items.

The passive gameplay revolves around social distancing. The player places a Safe Zone Beacon on the game map. The beacon is typically their home address. It creates a geo-fenced area where the game is easiest to play, and the player can maximize their progression while there. The further the player is from their beacon, the harder the game will be, and fewer points they will be able to accumulate. The player can move or even purchase more beacons, but these are intentionally costly, and the operation is not available frequently.
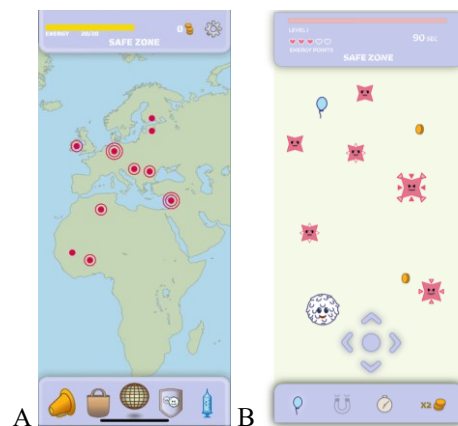

Fig. 2. A) The view of the map and B) the match (alpha)

The game design incorporates several persuasive systems design elements [49]. For example, the game is open and transparent about its behavior change intentions, as is

suggested by research [49, 50]. The game includes several persuasive features, such as primary tasks support, dialogue support, and credibility support features [49]. In the future, there is also a possibility to include social support features. Currently, the most important primary support features are tailoring and self-monitoring. Tailoring presents information and provides feedback based on the location of the player, while self-monitoring allows a player to monitor their progress over time. Dialogue support features include forms of praise, rewards, and reminders. Credibility support features focus is on surface credibility with a competent look and feel. The game also refers to authorities and uses trustworthy sources when presenting real-life information to players.

When it comes to gameplay, social distancing has a huge effect on whether the game is easy or difficult. The game has three levels of safety that social distancing effects. The zone where the player is located impacts the point accumulation. Points are important, as they are used for many purposes while playing. In short, safer the location, more points will be accumulated both actively and passively. The difficulty of matches is also relative to the zone safety. Currently, the levels are passable when in an unsafe zone, but more difficult. The player might be forced to use expensive powerups to win matches there. This makes playing the game in safer zones more attractive.

The third and the most difficult level is the danger zone. These spawn in heavily populated locations and matches will be very difficult to win there. In addition, when the player is at one of these locations, the point accumulation is very low in comparison to the safe zone.

C. Architecture and Implementation

The architectural implementation is based on a client-server paradigm (Fig. 3). It consists of the business layer, data layer, and presentation layer. The business layer contains all the application logic and communicates with external APIs. The data layer comprises data utilities such as storing user data. The presentation layer contains user interface components.

The development process was iterative and separated into several phases. The development team was responsible for internal testing. External evaluations were carried out at specific points in the process (Fig. 4). Evaluations included expert evaluations and player evaluations with the target audience.
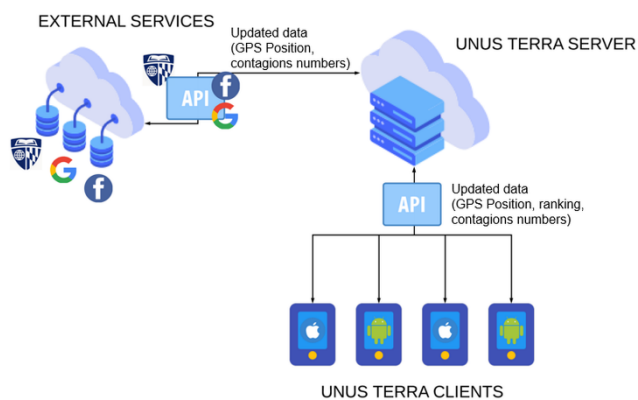

Fig. 3. Architectural implementation

An original idea from the first proof-of-concept that was developed in the hackathon was to track possible contagions between players by using Apple or Google contact tracing

technology, but it turned out that they are only doing cooperation with health authorities and governments. Another realization was that Google Maps might be too expensive for us to use and we decided to replace it with another mapping solution.

The first version of the game was ready at the beginning of October 2020. The game logic was set to be loosely based on real-life contagion data and the GPS location of the player. For example, the time limit to complete a level was set to vary based on the GPS location of the player. The location also effected the contagion level of the area. All three safety levelswere implemented in this phase.
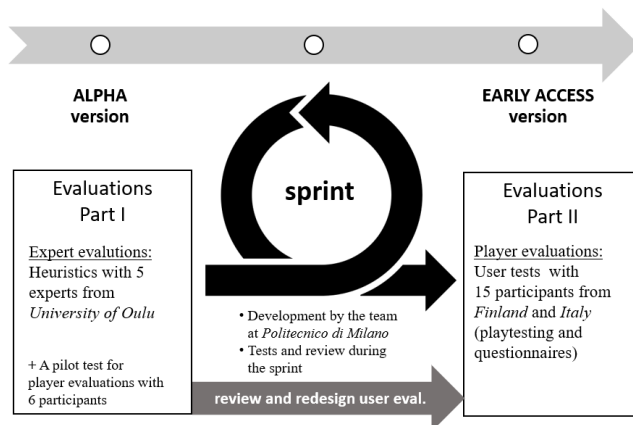

Fig. 4. A snapshot of evaluation activities during the development (Icon: Agile workflow method by Daniel Falk from the Noun Project)

Expert evaluations activities were conducted before player evaluations (Fig. 4). The game was developed further based on the feedback from the expert evaluations, as the results were utilized during the following sprint. For example, a minimap was added to matches, as it was hard to understand where the avatar was located on the map. Small improvements to the user interface were also made. This was followed by user tests that included participants that belong to the target audience of the game. The purpose of these later tests was to evaluate the gameplay further and to assess whether players thought that the game was entertaining. The expert evaluations presented in this paper center more on usability and gameplay issues.

IV. EVALUATION

A. Setup and Process

The alpha version of the game was tested by five usability and UX experts from University of Oulu. They did not take part in the design or development of the game. Two were female and three male. The five evaluators had over 50 years of heuristic experience between them, but notably two of them with over 20 years of experience with usability evaluations. Three of them had a Ph.D. and two M.Sc. from appropriate fields. Combined they had 23 years of expertise in game development. On average, they play mobile games for about five hours a week. All of them were too old to belong to the target audience of the game, but the focus of evaluations did not require this. The youngest evaluator was 36 years old at the time and others between 41 to 45. Evaluators have been sorted out by their experience in heuristics with E1 having the most experience and E5 with the least experience.

Evaluators were provided with the game file, a short-written instruction about the game and its purpose, along with a guide for installation. They were also told that the game is an alpha version and that it is missing some features such as the audio, in-game shop, and a COVID-19 related news

section. They were also provided a template for heuristic evaluation with instructions on how to use it. Evaluators were also asked to add other heuristic methods they were familiar with, if appropriate. One of the evaluators chose to carry out a small-scale user study with two children aged 11 and 15 years, respectively. Children played the game over few sessions and several questions from heuristics were translated to them. Another expert provided six extra heuristics to consider.

The instruction about the game included a brief description of the game type and what is its purpose. It was mentioned that the game has a more serious purpose focusing on behavioral change related to social distancing. The evaluation process was carried out independently and the analysis of the results was collated after all had finished their evaluations. Due to the time constraints of the development phase, we decided to forgo the planned group discussion between evaluators and developers. Identified issues and problems with the recommended solutions were forwarded to relevant parties.

B. Heuristic Evaluation

The experts were asked to evaluate the whole game, as they experienced it. This meant that specific issues or problems found may not have been present throughout the experience. For this purpose, the template provided for evaluation purposes included a lot of space to write down details on specific issues. Experts were also asked to provide possible solutions to the issues they found. For example, the E3 was only able to play the game for a limited time, as they managed to use all lives reserved for the avatar. This meant that E3 had rather limited exposure to the game in comparison to the other evaluators. This issue was not anticipated and was a good finding in itself. Another major issue was related to the installation process. The game asked for an email address that had to be verified. Some evaluators did not receive the verification email because the email service chosen by the developers was considered spam. Their evaluation had to be postponed until the issue was resolved.

Tables III to V highlight the data and relative frequencies of heuristics grounded issues. The mean of the last column relates to the issue found but does not include values that are zero. Overall, the results align with predictions derived from the evaluator effect. Many heuristic factors have high variance between experts and not all of them found the same problems. A variety of issues were also detected as feedback from experts and their recommendations were not always similar. The calculated mean is still a good indicator of a prevalence of a problem. The variance indicates whether the experts agreed on the scale of the problem when using severity ratings. Severity ratings are also subjective, as the same issue might be perceived quite differently between evaluators. However, this is one of the instances where having more than two evaluators is beneficial, as the average score provides a better indicator of the actual severity. Finally, the last column shows the percentage of experts who found at least a single issue with the heuristic.

There were only two heuristics that the evaluators found no problems with. A total of thirteen heuristics had a mean of over two. A high mean with a low variance is a good indicator for giving a heuristic a high priority for further analysis. For example, in GU02, which is about the screen, experts found many issues ranging from graphical choices to button locations. While all evaluators found issues, the variety of different issues was great. This was true for many of the heuristics used.

There are only three mobility heuristics and they all seem to have low mean value, which should indicate that no significant problems were present. However, a more detailed look at the issues raised by the experts shows that there were problems regarding loading times and occasional crashes. While these issues might not be significant at the alpha stage, they should be monitored with the later versions. This component is a good example that some issues or problems might not be present for all evaluators but might still cause notable issues if they persist.

TABLE III. RESULTS FROM GAME USABILITY HEURISTICS BASED ON FIVE EVALUATORS (E)

Game Usability	E1	E2	E3	E4	E5	mean	Variance	Issue found %
GU01	1	4	0	2	3	2,0	2,00	80
GU02	2	3	3	3	3	2,8	0,16	100
GU03	0	0	1	1	2	0,8	0,56	60
GU04	2	3	1	3	2	2,2	0,56	100
GU05	2	3	3	1	2	2,2	0,56	100
GU06	2	2	1	0	2	1,4	0,64	80
GU07	2,5	1	3	1	0	1,5	1,20	80
GU08	2	0	3	2	2	1,8	0,96	80
GU09	2	3	2	3	3	2,6	0,24	100
GU10	2	2	3	0	0	1,4	1,44	60
GU11	0	0	0	0	0	0,0	0,00	0
GU12	4	4	1	3	3	3,0	1,20	100
mean	1,8	2,1	1,8	1,6	1,8	1,8	0,03	85

TABLE IV. RESULTS FROM MOBILITY HEURISTICS BASED ON FIVE EVALUATORS (E)

Mobility	E1	E2	E3	E4	E5	mean	Variance	Issue found %
MO01	0	2	0	0	3	1,0	1,60	40
MO02	0	0	0	0	0	0,0	0,00	0
MO03	0	0	2	0	0	0,4	0,64	20
mean	0,0	0,7	0,7	0,0	1,0	0,5	0,73	30

TABLE V. RESULTS FROM GAMEPLAY HEURISTICS BASED ON FIVE EVALUATORS (E)

Gameplay	E1	E2	E3	E4	E5	mean	variance	Issue found %
GP01	2,5	4	0	3	2	2,3	1,76	80
GP02	2,5	2,5	1	3	1	2	0,70	100
GP03	2,5	2	2	3	3	2,5	0,20	100
GP04	0	3	0	0	0	0,6	1,44	20
GP05	3,5	3	0	2	2	2,1	1,44	80
GP06	3	4	0	4	3	2,8	2,16	80
GP07	3	4	1	0	2	2	2,00	80
GP08	3,5	3	0	3	3	2,5	1,60	80
GP09	3	0	2	3	0	1,6	1,84	60
GP10	3,5	3	0	3	2	2,3	1,56	80
GP11	3,5	2	3	3	3	2,9	0,24	100
GP12	0	3	0	0	0	0,6	1,44	20
GP13	3,5	2	0	0	0	1,1	2,04	40
GP14	0	0	0	0	2	0,4	0,64	20
mean	2,4	2,5	0,6	1,9	1,6	1,8	0,46	67

In table V, the variance is high with several Gameplay heuristics. This is mainly because E3 did not find issues with many of them. E3 had limited exposure to the game, as the person managed to kill off all available avatars early on, as mentioned previously. Discounting this anomaly, the values would align more comfortably with other heuristics in table III.

The serious game aspect is tackled in a few heuristics by evaluators. GP07 is about the game story supporting the game and being meaningful. Four evaluators agreed that this is an area that was lacking or that there was a disconnect. Notably, E1 felt that the pandemic is just a coat of paint on top of the game mechanics, which is a fair criticism, as this version of the game did not include the tutorial or the news section that would tie it more to the current pandemic.

C. Small-Scale User Study

The small-scale user study was conducted in an on-site setting by E2. Two test users were interviewed and observed during the gameplay. After the first gameplay, the game was explained and additional questions were asked relating, for instance, to ideas of the game and social distancing. The participants played the game for approximately 30 minutes to one hour. These tests provided further feedback on the social distancing aspect, as the adolescents (11- and 15-years old players) felt that it was exciting to play the game in less safe zones. This is the opposite that the game tries to convey as behavioral change and is a very important finding. None of the adult evaluators reported this or even considered the possibility. The preliminary indication is that the risk and reward aspect of the game needs to be finetuned to make it very unappealing to enter more dangerous zones.
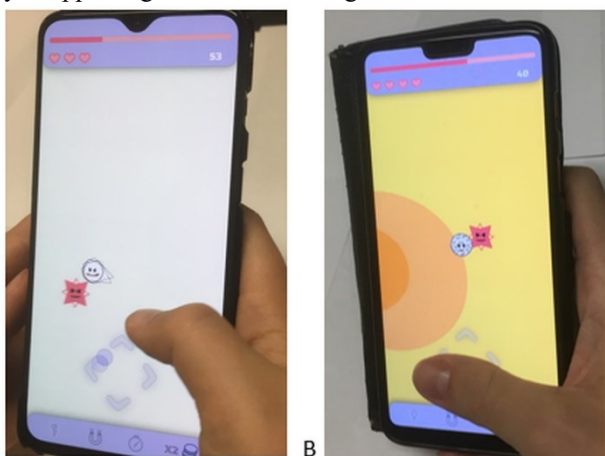

Fig. 5. A) The participant (ID1) is controlling the avatar (antigen) in the Speedy mode, B) The other participant (ID2) suggested that a character should display visual effects when taking damage

Perceptions of the adolescent players related to visual design, usability, game purpose, entertainment, and pandemic situation. For instance, the player (ID1) found it easy to understand how to control the game and understood quickly how to use different avatars to eat viruses (Figure 5A ID1 proposed that the control area for the avatar could be a circle (Fig. 5B). When registering for the game, the player (ID2) did not want to give his home location; therefore, he was in the unsafe zone when he started to play. This was another good finding from these tests, as it allowed the question to be raised in the later user test. In the unsafe zone, for instance, the background color is different (Fig. 5B). The

player (ID2) suggested that the visual design of that the avatar could twinkle or slow down when it is getting damage (Fig. 5B). These animations and corresponding sounds had not been implemented for this version.

The player experience was sometimes confusing. For instance, the player (ID2) commented: "*Why it stopped even though I had time left?*". In this version, there were several visual design and usability aspects, which had negative influences on the overall player experience. One feature that the player (ID2) particularly hoped for was a tutorial. It is important to provide a tutorial and tips for the users during the gameplay. While the game is designed to be easy to use and simplistic, it can still confuse a new player.

V. DISCUSSION

It is challenging to develop serious games that balance the entertainment value and serious aspects equally well. In this paper, the challenge was to design a game that provides entertainment but also instills the player with important information regarding the real world. Serious messages can be easily lost, so the design needs to be well-thought out. This is to make sure that the behavioral change message remains a part of the gameplay experience and is not just a coat of paint, as described by one of the expert evaluators.

It is well-established that heuristic evaluation methods are a cheap and quick way to get feedback during development. Agile teams have been shown to benefit from frequent usability and UX evaluations [51-53]. We believe that this might be especially true for games, as gameplay issues can be particularly difficult to resolve later on in the development. In our case, the development team used a version of iterative process similar to Agile. They conducted internal testing, but the external evaluations described in this paper were done fairly late in the development process, as the game had already reached the alpha state. We believe that the use of only one or two expert evaluators, but more frequently, might have been more beneficial to the development as a whole. The goal after all is not to find all issues at once, but enough actionable issues for the development team to focus on. We believe that domain experts could have been employed even earlier, for example, to evaluate mockup designs. We would also propose further study of using heuristics with different types of game development methods.

When using more expert evaluators, severity ratings can provide a good level of feedback that can be used when analyzing the data. The variance among evaluators can also be a beneficial finding. Allowing each evaluator to describe issues, comment on them and add recommendations is a good way to ensure that a wide range issues are addressed. This is an alternative, but not substitute for group discussions.

Severity of issues and possible solutions can also be discussed among evaluators providing more refined responses directly to developers. While the project described here did not have time to hold group discussions, we believe that potential discussions between developers and evaluators should focus on specific issues either clearly highlighted in the results or raised by either party. The number of issues found was high, but many of them do not require further explorations. After the analysis, some of the issues had already been noticed by the developers and had been addressed, but the heuristic evaluation did provide further acknowledgment of those issues and introduced several new issues to consider.

The five expert evaluators produced a lot of feedback. While this feedback is interesting, it might be in

overabundance, especially for smaller projects. It takes considerable effort to analyze the feedback and find the important focus points for the development team. This game is rather simple, but expert evaluators managed to reveal a lot of issues in many different areas. How to address these issues requires a lot of effort from the development team with short sprint times. Also, developers should assess whether it is better to evaluate only small segments, for example, scenarios, rather than the whole game at once. While unmoderated evaluation sessions can be beneficial as different issues are often found, the specificity might help with shorter development schedules.

One of the findings indicates that younger players may seek excitement when they see other people doing the same thing. It is not really surprising younger people seek boundaries and can be influenced to act against better judgement. This was apparent, when one of the players revealed anticipation and then excitement when they were able to play the game from more dangerous zone. While this is a clear indication that the game needs to be rebalanced so that it is less attractive to behave like this, it is also a very interesting finding that should be considered when developing similar games in the future. Is there, for example, a place for social support that would positively influence the younger players? It is also not surprising that challenging gameplay can be more appealing, rewards alone might not deter younger players trying to go against the intention of designers. A solution could be to seek balance by providing slightly more challenging gameplay even in safer zones, while increasing negative results from playing in more dangerous zones.

It is important to note that expert evaluators are not substitute for target audience Experts can be used to find issues and problems relating to usability, gameplay and game mechanics, but may not be able to evaluate the entertainment value of the game, as they may not belong to the intended user group.

The main limitation to our study is the scope. The sample size is simply too small to generalize, so we chose to forgo more detailed statistical analyses. While some of the findings from the small-case user study are interesting, the findings require further study and should only be considered as indications. We also chose to present the project in more detail for the context and focus on the heuristic evaluations, but a decision to include the data and analysis from users tests might have provided more meaningful contrast to the presented findings. As a future work, we intend to publish our findings from the user tests combined with statistical analyses.

VI. CONCLUSION

Usability is very important in games and not least because of fierce competition. Heuristic evaluations provide a cheap, fast and straightforward way to discover potential usability and gameplay issues. The method is well suited for smaller development teams using iterative approach. While using severity ratings is not a new idea, it should be utilized as it makes it easy to score issues. Furthermore, allowing expert evaluators to produce more details of issues and possible solutions for them should be encouraged.

Analyzing and communicating the results from expert evaluations can be a time-consuming task. While this depends on the level of detail provided by the evaluators, we believe that using fewer expert evaluators more frequently and also earlier in the development process is a better approach than one taken in this study.

Developing serious games with a goal of behavior change is challenging. It is also hard because measuring outcome of behavior change takes time, which smaller iterative development projects often do not have. They have to rely on best practices instead. While our results are somewhat inconclusive, they suggest that younger audience can easily go against the intention of game designers. This issue is further complicated with the fact that it might not be easy to detect this issue until the game is far along in the development. It might also not be noticed by expert evaluators. Developers designing games such as this should consider limited user tests with the target audience early to discover similar issues.

REFERENCES

[1] K. Isbister and N. Schaffer, "Game usability: Advice from the experts for advancing the player experience," San Francisco, California, Oxford: Morgan Kaufmann ; Elsevier Science [distributor], 2008.

[2] M. Rajanen and J. Nissinen, "A survey of Game Usability Practices in Northern European Game Companies," in Information Systems Research Seminar, vol. 8, p. 16, 2015.

[3] M. Rajanen and J. Tapani, "A Survey of Game Usability Practices in North American Game Companies," in B. Andersson, B. Johansson, S. Carlsson, C. Barry, M. Lang, H. Linger, & C. Schneider (Eds.), Designing Digitalization (ISD2018 Proceedings). Lund, Sweden: Lund University, p. 12, 2018.

[4] R. Yáñez-Gómez, D. Cascado-Caballero, and J.-L. Sevillano, "Academic methods for usability evaluation of serious games: a systematic review," Multimed Tools Appl, vol. 76, no. 4, pp. 5755–5784, Feb. 2017.

[5] S. Mylly, M. Rajanen, and N. Iivari, "The Quest for Usable Usability Heuristics for Game Developers," in Advances in Information Systems Development, vol. 39, A. Siarheyeva, C. Barry, M. Lang, H. Linger, and C. Schneider, Eds. Cham: Springer International Publishing, pp. 164–181, 2020.

[6] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93, Amsterdam, The Netherlands, pp. 206–213, 1993.

[7] M. Rajanen and D. Rajanen, "Heuristic evaluation in game and gamification development," in GamiFIN, pp. 159-168, May 2018.

[8] H. Korhonen and E. M. I. Koivisto, "Playability heuristics for mobile games," in Proceedings of the 8th conference on Human-computer interaction with mobile devices and services - MobileHCI '06, Helsinki, Finland, p. 9, 2006.

[9] D. L. King, P. H. Delfabbro, J. Billieux, and M. N. Potenza, "Problematic online gaming and the COVID-19 pandemic," Journal of Behavioral Addictions, vol. 9, no. 2, pp. 184–186, Jun. 2020.

[10] Y. S. Balhara, D. Kattula, S. Singh, S. Chukkali, and R. Bhargava, "Impact of lockdown following COVID-19 on the gaming behavior of college students," Indian J Public Health, vol. 64, no. 6, p. 172, 2020.

[11] T. Lemenager et al., "COVID-19 Lockdown Restrictions and Online Media Consumption in Germany," IJERPH, vol. 18, no. 1, p. 14, Dec. 2020.

[12] W.C. Kriz, "Gaming in the Time of COVID-19," Simulation & Gaming, vol. 51, no. 4 pp. 403–410, Jul. 2020.

[13] Business Wire, "Games Industry Unites to Promote World Health Organization Messages Against COVID-19; Launch #PlayApartTogether Campaign," businesswire.com, Mar. 28, 2020.

[Online]. Available at: https://www.businesswire.com/news/home/20200328005018/en/Games-Industry-Unites-to-Promote-World-Health-Organization-Messages-Against-COVID-19-Launch-PlayApartTogether-Campaign [Accessed Feb. 1, 2021].

[14] G. Syangtan, S. Bista, P. Dawadi, B. Rayamajhee, L.B. Shrestha, R. Tuladhar, and D.R. Joshi, "Asymptomatic SARS-CoV-2 Carriers: A Systematic Review and Meta-Analysis," Frontiers in Public Health, vol. 8, p. 1066, 2021.

[15] Z-Man Games, "Inc. Pandemic," boardgamegeek.com. [Online]. Available: https://boardgamegeek.com/boardgame/30549/pandemic [Accessed Feb. 1, 2021].

[16] Weltquartett, "Weltquartett," weltquartett.de. [Online]. Available: https://www.weltquartett.de/ [Accessed Feb. 1, 2021]

[17] Ndemic Creations, "Homepage - Ndemic Creations," ndemiccreations.com. [Online]. Available: https://www.ndemiccreations.com/en/ [Accessed Feb. 1, 2021].

[18] U. Ritterfeld, M. J. Cody, and P. Vorderer, Eds., Serious games: mechanisms and effects. New York: Routledge, 2009.

[19] H. Desurvire and C. Wiberg, "Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration." in International conference on online communities and social computing, Springer, pp. 557–566, 2009.

[20] H. Desurvire and M. S. El-Nasr, "Methods for Game User Research: Studying Player Behavior to Enhance Game Design". IEEE Computer Graphics and Applications 33, 4, pp. 82–87, Aug. 2013.

[21] M. Seif El-Nasr ym., "Game user research", in CHI'12 Extended Abstracts on Human Factors in Computing Systems, ss. 2679–2682 A, 2012.

[22] A. Drachen, P. Mirza-Babaei, ja L. E. Nacke, "Games user research", Oxford University Press, 2018.

[23] M. Irwin, "Cooking Up A Blockbuster Game," forbes.com, Nov. 21, 2008. [Online]. Available: https://www.forbes.com/consent/?toURL=https://www.forbes.com/2008/11/21/games-eedar-developers-tech-ebiz-cx_mji_1121eedar.html [Accessed Feb. 1, 2021].

[24] S. Rabin. Introduction to Game Development (2. ed.). Course Technology, Boston, Massachusetts, US. 2010.

[25] M. Rajanen and D. Marghescu, "The impact of game usability to player attitude", in Proceedings of 29th Information Systems Research Seminar in Scandinavia, Helsingoer, Denmark. 2006.

[26] T. Olsen, K. Procci, and C. Bowers, "Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness," in Design, User Experience, and Usability. Theory, Methods, Tools and Practice, vol. 6770, A. Marcus, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 625–634, 2011.

[27] J. L. González Sánchez, N. Padilla Zea, and F. L. Gutiérrez, "From Usability to Playability: Introduction to Player-Centred Video Game Development Process," in Human Centered Design, vol. 5619, M. Kurosu, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 65–74, 2009.

[28] P. Morton, "Learning from Fable III's UX Mistakes," uxbooth.com, Jun. 30, 2011. [Online]. Available: https://www.uxbooth.com/articles/learning-from-fable-iiis-ux-mistakes/ [Feb. 1, 2021].

[29] M.A. Federoff, "Heuristics and usability guidelines for the creation and evaluation of fun in video game," Indiana University, 2002.

[30] H. Korhonen and E.M.I. Koivisto, "Playability heuristics for mobile multi-player games," in Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts, pp. 28–35, 2002

[31] J. Paavilainen, "Critical review on video game evaluation heuristics: social games perspective," in Proceedings of the International Academic Conference on the Future of Game Design and Technology, pp 56–65, May 2010.

[32] M. H. Phan, J. R. Keebler, and B. S. Chaparro, "The Development and Validation of the Game User Experience Satisfaction Scale (GUESS)," Hum Factors, vol. 58, no. 8, pp. 1217–1247, Dec. 2016.

[33] I. J. Livingston, R. L. Mandryk, and K. G. Stanley, "Critic-proofing: how using critic reviews and game genres can refine heuristic evaluations," in Proceedings of the International Academic Conference on the Future of Game Design and Technology - Futureplay '10, Vancouver, British Columbia, Canada, p. 48, 2010.

[34] A. Ponnada and A. Kannan, "Evaluation of mobile games using playability heuristics," in Proceedings of the International Conference on Advances in Computing, Communications and Informatics, pp. 244-247, 2012.

[35] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90, Seattle, Washington, United States, pp. 249–256, 1990.

[36] S. Laitinen, "Do usability expert evaluation and test provide novel and useful data for game development?" Journal of Usability Studies, vol. 1, no. 2, pp. 64-75, 2006.

[37] H. Desurvire, M. Caplan, and J. A. Toth, "Using heuristics to evaluate the playability of games," in Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04, Vienna, Austria, pp. 1509-1512, 2004.

[38] R. Yanez-Gomez, J. L. Font, D. Cascado-Caballero, and J.-L. Sevillano, "Heuristic usability evaluation on games: a modular approach," Multimed Tools Appl, vol. 78, no. 4, pp. 4937–4964, Feb. 2019.

[39] M. Minge, M. Thüring, I. Wagner, and C.V. Kuhr, "The meCUE Questionnaire: A Modular Tool for Measuring User Experience," in Advances in Ergonomics Modeling, Usability & Special Populations, vol. 486, M. Soares, C. Falcão, and T. Z. Ahram, Eds. Cham: Springer International Publishing, pp. 115–128, 2017.

[40] M. Schrepp and J. Thomaschewski, "Eine modulare Erweiterung des User Experience Questionnaire," Mensch und Computer 2019-Usability Professionals, 2019.

[41] J. Sauer, A. Sonderegger, and S. Schmutz, "Usability, user experience and accessibility: towards an integrative model," Ergonomics, vol. 63, no. 10, pp. 1207–1220, Oct. 2020.

[42] J. Nielsen, "Severity ratings for usability problems," Papers and Essays, vol. 54, pp. 1–2, Nov. 1995.

[43] N.E. Jacobsen, M. Hertzum, and B.E. John, "The evaluator effect in usability tests," in CHI 98 conference summary on Human factors in computing systems, pp. 255-256, Apr. 1998.

[44] M. Hertzum, R. Molich, and N. E. Jacobsen, "What you get is what you see: revisiting the evaluator effect in usability tests," Behaviour & Information Technology, vol. 33, no. 2, pp. 144–162, Feb. 2014.

[45] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93, Amsterdam, The Netherlands, pp. 206–213, 1993.

[46] Ikune Labs, "Ikune – creating engaging and lasting experience for everyone," ikunelabs.com. [Online]. Available: http://ikunelabs.com/ [Accessed Feb.1. 2021].

[47] EIT Digital, "DATA Against COVID-19' EIT crisis support to digital European entrepreneurs and innovators,"eitdigital.eur, Aug. 13, 2020. [Online]. Available: https://www.eitdigital.eu/newsroom/news/article/data-against-covid-19-eit-crisis-support-to-digital-european-entrepreneurs-and-innovators/ [Accessed Feb. 1, 2021]

[48] EIT Digital, "Unus Terra, a serious game from Finland, wins top prize at the EIT Digital DATA against COVID-19 DeepHack," eitdigital.eu, May 4, 2020. [Online]. Available: https://www.eitdigital.eu/newsroom/news/article/unus-terra-a-serious-game-from-finland-wins-top-prize-at-the-eit-digital-data-against-covid-19-deephack/ [Accessed Feb. 1. 2021].

[49] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive Systems Design: Key Issues, Process Model, and System Features," CAIS, vol. 24, 2009. doi: 10.17705/1CAIS.02428.

[50] H. Oinas-Kukkonen, "A foundation for the study of behavior change support systems," Pers Ubiquit Comput, vol. 17, no. 6, pp. 1223–1235, Aug. 2013.

[51] D. Sy, "Adapting Usability Investigations for Agile User-centered Design," Journal of usability Studies, vol. 2, n. 3, pp.112-132, 2007.

[52] L. Schwartz, "Agile-User Experience Design: an Agile and User-Centered Process?," in Proc. the 8th International Conference on Software Engineering Advances, pp. 346-351. 2013.

[53] S. Masood Butt, S. Masood Butt, A. Onn, N. Tabassam, and M. Abdul Majid, "Usability Evaluation Techniques for Agile Software Model," JSW, vol. 10, no. 1, pp. 32–41, Jan. 2015.