

# Playtesting for Indie Studios

Pejman Mirza-Babaei  
University of Ontario Institute of  
Technology  
2000 Simcoe Street North  
Oshawa, L1H 7K4, ON, Canada  
pejman.m@acm.org

Naeem Moosajee  
University of Ontario Institute of  
Technology  
2000 Simcoe Street North  
Oshawa, L1H 7K4, ON, Canada  
naeem.moosajee@gmail.com

Brandon Drenikow  
University of Ontario Institute of  
Technology  
2000 Simcoe Street North  
Oshawa, L1H 7K4, ON, Canada  
drenikowb@gmail.com

## ABSTRACT

Creating video games is a lengthy and demanding process. Financial success for games studios often depends on making games that deliver a fun and engaging experience for a diverse audience of players. Therefore, understanding how players interact and behave during gameplay is of vital importance. Playtesting aims to assist developers to achieve their design intent and help to identify and resolve potential problem areas during development. However, playtests are not always feasible or affordable for smaller, independent game developers (indie studios) because they require specialized equipment and expertise. In addition to this, there is a lack of research on the value of playtesting for indie studios, which means most indie developers are not convinced of the value of user research and playtesting. This paper reports on our collaboration with six commercial indie developers conducting eleven rounds of playtesting session. Through these collaborations, our paper contributes to this growing domain by highlighting the value of playtesting for indie developers and discussing the user research process and approaches based on indie developers' needs and budget.

## CCS Concepts

• Human-centered computing • Interaction design  
• Interaction design process and methods

## Keywords

Games User Research; playtesting; game development; indie developers; user experience; case studies.

## 1. INTRODUCTION

Games like Grand Theft Auto V and Destiny have development costs in the \$137 million (M) and \$140M ranges respectively [23, 25] rivaling the budgets of major motion pictures which is showing large growth in the games industry. In the interest of minimizing expenses, these teams budget for playtesting and quality assurance (QA). The number of games being developed by small (and indie) teams developers on low budgets is on the rise (for example, indie games comprise 30% of 1.9 billion purchased to-date on the Steam platform). As of July 2016, over 3 million

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

*AcademicMindtrek'16*, October 17 - 18, 2016, Tampere, Finland  
Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-4367-1/16/10...\$15.00

DOI: <http://dx.doi.org/10.1145/2994310.2994364>

applications exist on the Apple App store and Google Play [26, 27]. According to Entertainment Software Association (ESA) of Canada, more than 88% of Canadian video games companies are micro or small studios (fewer than 99 employees). Additionally, Independent game developers (indie studios) have similar risks with larger game productions while having a much smaller budget. Smaller game studios often do not have the budget for arranging expensive playtesting, so there is a need to incorporate playtesting techniques for smaller teams and accommodate their budgets.

This paper reports on eleven playtest case studies on six commercial indie games. The goal of this paper is to argue why playtesting is important for small independent studios and to explore process of providing inexpensive playtests into the indie development cycle. This paper provides example of equipment and methods that can be used when playtesting with lower budget. We discuss the adaptation of previous works on user research methods involved in playtesting for the game development industry.

## 1.1 Playtesting

Playtesting is becoming a common method to improve games as an evaluation approach from Games User Research (GUR), which is building upon psychology and Human-Computer Interaction (HCI) [17, 32]. GUR aims to improve on a game's design and player experience by performing a series of usability and user experience (UX) evaluation [2, 14]. Large game corporations (e.g. Ubisoft or Electronic Arts) have in-house playtesting. Since these companies make many titles overlapping the other's production, it makes sense to have internal playtesting teams [29, 31]. In addition to having an internal playtesting team, third-party playtesting specialists are also common, which are contracted by a variety of developers of all sizes. Understanding how players experience the game is important to the development process. In order to get an idea of the player's experience, in-house and external playtesting teams use mixed methods to get various complementary yet separate sets of data.

Playtesting uses a set of techniques and tools to measure and analyze behaviors of players to retrieve information for developers in order to improve the player experience. One of the current challenges in playtesting is to effectively gather, analyze and visualize the qualitative and quantitative data recorded. This is more noticeable for indie studios as they require specialized equipment and expertise, which may not be always feasible for smaller independent game studios. Maintaining a strict budget, avoiding unneeded costs, resources, and tools is a common challenge for small indie studios. An effective playtesting approach for these studios should be applicable while remaining accessible and affordable [18].

## 1.2 Partnering with Execution Labs

With the recent proliferation of free-to-play, professional-quality development tools (such as the Unity Engine), crowdsourced fundraising (such as Kickstarter and Indiegogo), and the ascendancy of digital distribution platforms (such as Apple's App Store or Valve Corporation's Steam), 'indie' digital game developers are emerging on an unprecedented scale [3]. With the rise of these platforms and services, the game industry's overwhelming barriers to entry are becoming increasingly penetrable. Hence, the indie games are growing significantly and there are incubators and accelerators exist to support and contribute to this growing domain. For example the *Hong Kong Science and Technology Park* (HKSTP)<sup>1</sup> and the *Gothia Science Park*<sup>2</sup> in Sweden where the famous *Goat Simulator*<sup>3</sup> was developed.

In an effort to work with indie game developers, we formed collaboration with Execution Labs (XL). XL offers indie game developers funding, mentorship, playtesting sessions, analytics services, production assistance, and access to industry networks as well as a bustling indie community. XL's goal is to foster an environment in which indie start-ups can grow into stable; self-sustaining studios with aim to assist indie developers get their studios off the ground [3]. Collaborating with XL was mutually beneficial. We provided assistance with playtesting and XL provided eager indie developers.

This paper reports on our experience based on conducting playtest sessions on six games by independent game development teams. We present these studies to showcase the application of playtesting and user research methods, noting the adaptation and contributions of each method to the development process for indie developers.

## 2. RELATED WORK

One major issue that arises is the quantity of data that needs to be looked over. It can be often daunting and confusing to decipher, which may cause game developers to experience a slow, tedious task of going through the data to attempt to make improvements to their games. Thus, one of the challenges is making the interpretation of playtesting data effective and meaningful in terms of facilitating design decisions for game developers.

HCI user evaluation methods have made progress in understanding the usability of productivity applications and websites [4, 28]. However the specific characteristics of video games such as the addition of '*intentional challenge and emotion*', mean that many established methods of user research cannot be applied in the same way [15]. Current methods of evaluating games user experience commonly include subjective self-reports through questionnaires, interviews, and focus groups as well as objective reports through observational video analysis [9, 7].

In the following paragraphs we review some of the most common GUR methods and explain how we adapted these methods for indie GUR.

*Behavioural observations* are one of the most common approaches in games user research. They can provide a basis for a detailed analysis of usability [22], as well as fun and game experience [24,

8]. Observational studies is a technique where a participant is required to perform an activity (playing a game) while being recorded for future observation, or is directly observed by another individual. This technique is extremely effective since it is relatively cost effective requiring very few resources and an observer can detect issues with player progression, player challenges, and if a player gets stuck or becomes unmotivated. An observer can also examine body posture and facial expression to estimate emotions [10]. While behavioral observation is great for gathering objective data and seeing exactly what players do, it is time consuming to record and analyze the data. Another drawback to observational studies is that the observer should be experienced enough to take objective notes which is not very intuitive for novice observers. Participants may also perform differently when being observed due to external pressure, meaning the behavior observed might not be completely natural [10]. Analyzing observational data as an indication of player experience is a lengthy and difficult process, which must be undertaken with great care to avoid biasing the result [12].

*In our approach* we rely heavily on the use of observation data. The playtest area at XL does not have a one way mirror, instead it is inside a modular room where recording devices capture the observation data from the playtest. One flexible camera is positioned so it can capture the user interacting with the platform they are playing with. A game capture device is connected to the PC or mobile device streaming the footage to the observers in another room. This allows for the possibility of developers entering the observation room to view the playtest in real time without disrupting the player (see Figure 3).

*Think-aloud protocols* could be added to the observational studies where the participant being observed is required to talk about their actions and thoughts as they perform different tasks. The aim is to get inside the players' thinking processes 'in the moment', potentially revealing unobservable details and providing researchers with immediate feedback. Like observation, think-aloud can be recorded and viewed later. This technique is effective because it provides insight on the internal state of the participant providing great mental insight to gather behavioral data [11]. A downside to think-aloud is that it is unnatural for most people to talk objectively while playing without any prior training. It is argued that 'think-aloud' techniques cannot effectively be used within game testing sessions because of the disturbance to the player and ultimately the impact they have on gameplay [19]. *In our approach* we prompt players to feel free to talk and express their thoughts during the playtest. A microphone is placed near the player to record any comments during the session.

*Heuristic evaluation* provides a formal and accessible usability evaluation method, which can be used even before any code has been written. There are a number of different heuristic sets created for video game usability evaluation, including PLAY [6], and Nielsen's list [20]. Although heuristic evaluation promises to be a low-cost usability evaluation method, it suffers with problems of subjective interpretation [30]. The heuristics chosen are variable and dependent on what needs to be evaluated. Although, we did not employ a specific heuristic analysis, but we apply heuristic to suggest design changes grounded in a structured vocabulary agreed upon with the developers.

*Questionnaires* are paper based or delivered online where they can be sent to a large number of participants with little effort. If the goal is to understand what people do, a questionnaire will be a sufficient method to employ. If the researcher needs to identify why something occurred, an in-depth interview or focus group

<sup>1</sup> See: <http://www.hkstp.org/en-US/Homepage.aspx>

<sup>2</sup> See: <http://www.gsp.se/en/about-gsp>

<sup>3</sup> See: <http://coffeestainstudios.com/games/goat-simulator>

will provide more valid responses than a questionnaire [1]. Although questionnaires are very useful, due to the small number participants and primary goal of the tests (on First Time User Experience) we did not utilize this approach. Instead we conducted interviews as explained below. For future playtests we would like to explore the usage of questionnaire data.

*Interviews* are conducted on a one-to-one basis and require a large amount of the investigator’s time for transcribing and coding the data. Interviews are often time-consuming process that requires careful planning to obtain more detailed and thorough information on a topic. During the interview the investigator often follows a schedule of pre-prepared questions but can deviate when necessary in order to maximize information obtained. The interview could be flexible enough to allow jumping between questions since sticking rigidly to a structure can annoy respondents [1]. *Our approach* heavily relied on interview data for the player to self-report on their experience. *Table 1* shows what questions we asked in our interviews.

**Table 1: Interview Questions**

Interview Questions
<i>What did you have to do?</i>
<i>What did you like?</i>
<i>What did you dislike?</i>
<i>What did you find confusing?</i>
<i>What were you unsure of?</i>
<i>What was your strategy?</i>
<i>How did you feel about the controls?</i>
<i>When you died, did you feel it was your fault or the game’s fault?</i>
<i>What’s your overall feeling of game?</i>
<i>If you could change one thing, what would you change?</i>
<i>If the entire game changed and you could keep one thing the same, what would you keep?</i>
<i>What was one thing that was most memorable about your experience?</i>
<i>What was one thing that you hated/frustrated about the experience?</i>

There are several ways to interpret data from a playtest session. Some previous research has worked towards creating tools and systems that facilitate both qualitative and quantitative data recording and visualization [16]. Others have created tools that accommodate for specific measurements of data like game metrics [11]. These metrics can involve a wide range of topics such as in game purchases or in game events such as the number of times a player collects an in-game item.

When selecting playtesting methods we chose to combine different approaches to mitigate the weakness of an individual approach. We adapted Rapid Iterative Testing and Evaluation (RITE) [13] approach but aimed to focus on First Time User Experience (FTUE) evaluation [5].

## 2.1 Rapid Iterative Testing and Evaluation

A number of industry-standard usability approaches use traditional methods mentioned previously in combination; for example, RITE [13] which employs observation and think-aloud techniques with the addition of an attending software engineer to rapidly alter the design based on the findings of the usability testing. Changes can be made after observing as few as one participant, with altered designs subsequently tested on the remaining participants. Other variations include open-ended usability tasks, paper prototypes, and empirical guideline documents [22].

The RITE method is a discount usability test conducted in a fast and highly collaborative manner [13]. We also decided to focus on the FTUE evaluation. The projects at XL were in early development and we decided that focusing on the FTUE would be most beneficial for the teams (see Table 2 for more details). Based on the important factors in game evaluation by Pagulayan et al [22], the following were areas we focused on using the RITE method: Game designer intent, ease of use, basic mechanics, starting a game, tutorials and instructional gameplay, camera, in-game interfaces, mapping input devices to functions, challenge and pace. A key part of RITE is *understanding design intent*, where the general design goal of a game is to create enjoyable experiences for its players.

## 3. APPROACH

Our main challenge was selecting an effective testing methodology that is fast, simple, and affordable to setup. Since the playtesting space (see Figure 1 and 2) is shared by many teams, timeliness is important for this to be feasible. The expenses for running the test must be low to have the team be able to run several testing and improvement cycles.

The lab is equipped with an adjustable webcam set up to record the participant interacting with the platform (mobile device, controller, keyboard etc.) as well as a microphone to record participant comments while their gameplay is recorded using an HD game capture device. Researchers and developers could observe the gameplay interaction behind a room divider (See Figure 3).



**Figure 1: Mobile Setup for Playtesting at XL ©Pejman Mirza-Babaei**



Figure 2: PC Setup at XL ©Pejman Mirza-Babaei

Playtest session lengths were about an hour on average as seen in Table 2. The first half is dedicated to gameplay and the latter half a semi-structured interview. To find issues experienced by players, we record observational data from gameplay and conduct a semi-structured interview. In the interview, we asked general questions which are found in Table 1. We asked extra questions that were specific to each game, which might focus on particular areas of developer interest.

We tested six games from six different developers described in the next section. See Table 2 for an overview of playtests conducted for the six video games.

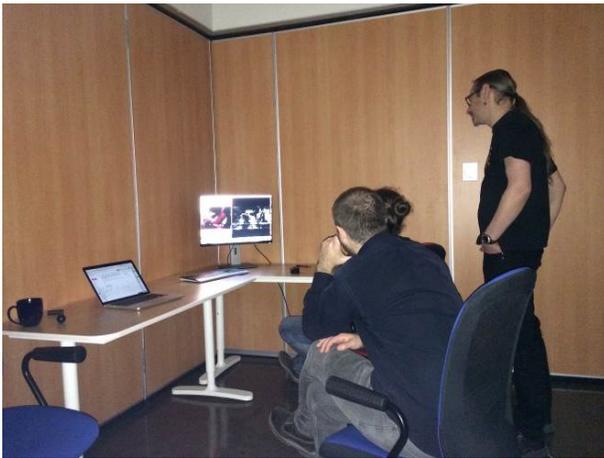


Figure 3: Observation room at XL ©Pejman Mirza-Babaei

## 4. RESULTS

In the following sections, each game is briefly explained, followed by a summary of the playtest setup and the main feedback received from the test. Differing studios expressed diverse motivations and responses to user testing.

The first three games (A, B and C) were in the early stage of development during the time of testing. The developers from *Game A* were more focused on developing content for their Kickstarter. The *Game B* developers were focused on ensuring that their core gameplay was solid. Lastly, the developers from *Game C* wanted to ensure that their platforming levels were not overly frustrating for the player.

The next three games were in the end of production stage of development, preparing for a commercial release. For these games

we focus more on preparing detailed reports highlighting issues as opposed to gaining feedback on the developers' reception to playtesting. *Game D* is a card collection game using gameplay journals as a method for testing. *Game E* is a multiplayer only top-down shooter with several characters to choose from. Lastly, *Game F* is a platformer that utilizes the PS4 touchpad as an interaction method.

Table 2: Games tested at XL

#	Game Type	Platform	Playtest Rounds	Methods	Gameplay Duration	Motivation
A	Action RPG	Console	R1	Observation, co-discovery, interview	30 min	Mechanic
B	Multiplayer Brawler	Facebook	R1	Observation, interview	30 min	Character Balance
			R2	Observation, interview	30 min	Character Balance
			R3	Analytics	30 min	Character Balance
C	Platformer	Mobile	R1	Observation, interview	30 min	Controller
			R2	Questionnaire	30 min	UX
D	Collectible Card Game	Mobile	R1	Journal study, interview	300 min	FTUE
E	Multiplayer Shooter	PS4, PC	R1	Focus group	30 min	FTUE
			R1	Observation, interview	120 min	FTUE
F	Platformer	PS4	R1	Analytics, observation, interview	120 min	FTUE
			R2	Analytics, observation, interview	120 min	FTUE

### 4.1 Game A: 3D Split Screen RPG

The first game is a 3D split-screen exploration and survival role-playing game (RPG). The game was designed for players to naturally assume roles of *Tank* and *Support* using the items and abilities provided. Generally the role of the *Tank* is to act as a sponge for taking damage and the role of *Support* is to assist the other roles using power-ups. The goal for these developers was to prepare for an upcoming Kickstarter campaign under the guidance of the specialists at the incubator.

The developers' goals for the user test were to evaluate controls, ensuring players fit into their role, and how players felt about the co-op mechanics. The playtest involved both players sharing the screen using Xbox 360 controllers to complete a demo level. Players needed to reach the end of the dungeon within the time frame or play for 30 minutes for the user test to be completed. Both players were interviewed simultaneously after the test.

When presented with testing results, the developers had acknowledged that all issues required fixing. They also expressed interest in conducting further testing, however, they indicated that the majority of future efforts would be focused on content generation, to secure crowdsourced funding. It was interesting to see that in the interest of acquiring funding or publicity, some developers may not view user testing as high-priority.

### 4.2 Game B: Facebook 2D Multiplayer Game

The second game is a Facebook 2D multiplayer brawler where the player can switch between three characters with a rock-paper-scissors style circular relationship. The developer's goals for the user tests were to explore how participants received the overall concept of the game and the controls.

Through the initial playtest results the team was motivated to conduct future tests while making immediate changes to their build before their next playtest.

The developers were inherently interested in player comments regarding controls, and the game overall. They had a polished prototype that put them in a good position to test the game frequently to make iterative changes. Additionally, since the game was already on Facebook, it allowed us to test with much larger participants simultaneously, providing more data in a shorter time frame. Having easy access to the latest build definitely increased the speed of setting up and conducting tests in multiple locations (such as a *remote usability* or appreciation test).

### 4.3 Game C: Mobile Survival Platform Game

Game C is a hardcore mobile side-scrolling platformer and survival game for touch screen devices. When presented with the report the team has expressed interest in planning more rounds of playtesting (we conducted 2 more rounds) to evaluate the resolution of identified issues while testing other levels for balancing purposes and on controls, UI, and feedback. This game is an example of a success where the team successfully integrated playtesting results in their development cycle.

The developers had designated one member of the team to be the main point of contact for everything related to user testing their game. This made it very easy to communicate the requirements for the test to the developers in terms of build requirements, player profile, and key performance indicators. Additionally, the one member attended all the tests and was able to see the improvement of the game through the iterative playtests. This helped convince the development team to hold future playtests as the team was motivated on the effectiveness of the tests.

### 4.4 Game D: Card Collecting Game

The next case is on a card collection game where the player strategically builds a deck of cards to engage in one-on-one battles. The game board is a 3 by 2 grid where each player can place up to three cards in a round. The goal is to place a better card opposite to the enemy to win the battle. The focus was on evaluating retention, FTUE and the interface. A detailed report was given to the developers outlining main issues in the FTUE and interface.

We utilized a user diary approach to evaluate this game since it allows users the freedom of picking when to play. This was the most effective way to measure retention and evaluate how the user felt about the game playing it at their own comfort.

### 4.5 Game E: Top-Down Shooter

Our penultimate case is a multiplayer top-down western style shooter. The game provides multiple character classes known as outlaws each with different stats for: damage, fire rate, stamina, range, speed, health, and character abilities.

The developers were interested in learning about their game interface flow to see if players understood their menus. The developers also wanted to evaluate their tutorial and FTUE.

We brought in four players to play through the tutorial level for the start of the playtest. After teams of two were formed, where each team played against the other for about an hour. Teams were mixed after each round was completed. After playing the game for an hour, all four participants were asked about their thoughts in a focus group setting.

## 4.6 Game F: Platformer

Our final case is on a hardcore platformer where the player controls two characters simultaneously to navigate through platforming puzzles. The game also utilizes the PS4 center touchpad to directly interact with items in the game. The touchpad is used similarly to a mouse, where the touchpad follows a one-to-one mapping with the screen.

The developers were interested in evaluating the whole player experience to understand players' reception to balancing and difficulty. In each round of testing, eight players individually played through the four playable worlds of the game totaling about four hours of gameplay. The gameplay sessions were broken down into two sessions, where the first session took place over two hours in the first two worlds. After the first session was completed, users were given a semi-structured interview based on questions from the developer. After the break, the playtest continued for the last two worlds. After the player was finished with the last session, another semi-structured interview was conducted based on questions from the developer.

## 5. DISCUSSION

For us the case studies were also of an exploratory nature, trying to understand indie development needs and highlight the contribution (and challenges) of playtesting in their development cycle.

Based on our results and observations, we discussed five key factors that influence *playtesting for indie studios*:

### 5.1 Playtesting Process

Playtesting process begins with gathering information from the developers to find what the goal of the playtesting is, all the way to following up with the developers after they have been presented the findings.

Understanding the need for a process is as important as the steps of the process itself. Having a well-defined process is a large time saver, which time is often an expensive resource for indie developers. Well-defined process would allow the time and resources allocated for the playtest to be well managed and optimized, so that most users can participate in the research. Having a process also ensures that people involved with the playtesting are well managed and since there are many people involved, this is invaluable during this stage of the game's development cycle. A process will also provide a standard to ensure there is consistency between playtests sessions. Once a standard is in place, this can allow for efficient training to new user researchers or even development team members that need to conduct playtesting in their iterative development process (see 5.5 for more details on training). Another important reason for having a well-defined process is that it gives a safety net for the team member conducting the playtesting as there will be variety of methods (see 5.2 for more details on methods) that is undergoing continuous iterations.

The process should define the overall design of the playtesting stages and will include determining many variables such as: number of participants, time playing game, and number of researchers needed. Depending on the detail of the target market and needs for the playtesting, the participant recruitment portion of this process can be time consuming so it helps to be prepared with a large participant pool to pick the proper participants from (see 5.4 for more details on participant recruitment). Once all of the details of the playtest have been planned out, the test must

actually be conducted. Past this point, the data collected (see 5.3 for more details on lab setup) needs to be analyzed and reported to the development team. The analysis and reporting can also be another resource consuming portion of this process. To reduce the time consumption during this portion of the process, it is helpful to have taken informative notes during each playtest, which may include timestamps especially if there are videos that need to be analyzed. In order to have a fast response from the developers after delivering the findings is to ensure the findings are clear and concise.

## 5.2 Playtest focus, methods

Most of the games tested focused on FTUE using RITE method. A focus was placed on FTUE since most teams only had a playable core experience for the first hour or “onboarding” phase. RITE complements FTUE well since we could evaluate the initial experience with smaller groups, providing room for more iterations. Observation and interview methods are also very affordable which are why they are the main methods used. These methods are also easy to setup and can be performed iteratively without heavy modification to the process. Our goal was to utilize smaller samples size but to focus on iterative playtests and in-depth data analysis.

However, for *Game B's playtest round 3*, an online version of the game was posted for a group of 60 online-participants to play. The focus of this round was to balance level difficulty and we utilized telemetry data (level completion and death record) for analysis. Developers who had telemetry logging in their game also benefitted from the use of an analyst (such as to create heatmaps) for the developers.

Similarly, for appreciation-based evaluation (for example in *Game C, playtest round 3*), we recruited 12 participants and only utilized questionnaires to maintain the budget.

When testing *Game D*, the card collecting game, we decided to use user dairies. The dairies were effective at getting the player to self-report on their gameplay while removing the interviewer bias. We told the players to use Google Documents to report on their experience so the researchers could access them at anytime. User diaries are also a cost-effective method of testing due to the user utilizing their own setup to play the game. Additionally, writing journals is entirely done online using free widely available tools.

Focus groups are effective at idea generation, however they are less effective at gathering feedback after a playtest. The issue with the focus group is that generally the loudest voice can sway other users into thinking a problem existed when it did not. Instead of getting an accurate self-report or discussion of issues, we found that a mass hysteria phenomenon came to fruition.

Four out of the six games used interview data as the primary method of evaluation. Interviews are extremely effective due to their ease of deployment. It was very easy to get questions from developers regarding what they wanted to ask players, and formulate them into viable interview questions. Additionally, having developers present during interviews sparked further questions during the probing period. For this reason, we found interviews very effective when conducting playtests with fewer participants.

## 5.3 Playtest Lab setup

Larger (AAA) companies have the luxury of multiple games [21] in development at different stages. For this reason having a dedicated lab inside the studio is economical. Dedicated labs are

effective for consistently testing games weekly generating more data. The other benefit is the ease of setting up a test. Since games are tested weekly, the process for gathering participants, setting up playtest machines, conducting the playtests, and gathering the data are well documented until the process is almost automated. Most in house labs follow either appreciation or usability lab setups. In-house appreciation labs generally have around 16 or more cubicles or playtest stations each fitted with gaming hardware (multiple gaming monitors, video game console, video game controller, headphones, etc.) and playtest hardware (webcam, microphone, biometrics, etc.). Usability labs try to replicate a living room environment to emulate how a game would be played in a natural setting. Usability labs often have a couch and large screen TV with a one-way mirror to observe the playtest from a separate room.

Dedicated labs generally use observation and self reporting questionnaires to gather playtest data. Observation is done through webcam and audio recordings in each playtest station or through moderators taking real time notes. Depending on what game is tested, telemetry can also be used to track in game events. The combination of screen recording software and telemetry help paint an accurate picture of what is happening in the game. Using questionnaires help provide insight into what the player is thinking. In many large labs, playtest setups have multiple monitors where one is used for gaming and the other is used for answering questionnaires.

Developers that do not have access to a dedicated lab must be more creative with the space that is available such as converting a meeting room into a testing space temporarily (this is what we did for our sessions). Once the testing has completed it is then quickly brought back to a meeting room. This can be accomplished by having a minimalistic approach to how the playtest setup is created. In order cases, we used a divider to separate an observation room and a testing room; with this temporary wall we could quickly create and remove a playtesting space.

Additional approaches to developing this minimalistic setup would include reducing equipment for example by purchasing a webcam with microphone. Another benefit of having a portable setup is it can be easily transferred to different places where it is more accessible for the users participating as well as what rooms are available in the office space the team has. Moreover, this portable setup could be very useful if the playtest need to be completed outside the office, for example in a participant's home.

The size of the testing lab is dependent on the games that will be developed at the studio. If the studio aims to develop multiplayer games, it is recommended to build a 16 to 20 player lab. For companies focusing on single player and mobile titles, a smaller lab could suffice, with the obvious limitation of testing one person at a time.

The lab setup we used was built for single player playroom inside a modular meeting room with movable walls. The meeting room can be repurposed for team meetings or a segmented testing area split in half. We felt that this was the most effective lab setup for our purposes because it was able to test a variety of different types of game with little change in setup required.

Regarding cost of equipment, the most important items that required purchasing was as follows: game capture card, webcam on flexible frame, microphone, game hardware (which are available at studios) and various cables.

The game capture card is extremely useful for capturing gameplay footage directly from a mobile or console device. The webcam is useful for capturing the user interacting with the controller, or how they hold the touchscreen tablet/phone during gameplay. The flexible nature allows it to be positioned in a variety of angles as required. The microphone was used for picking up player comments during the playtest. The game hardware entirely depends on the resources of the developer. If the developer aims to develop purely mobile games, it is recommended to have a variety of mobile devices to test with to ensure accessibility.

## 5.4 Building Participants database

A large diverse participant pool is extremely important to conducting effective playtests. In order to get useful data, it is critical to ensure that right participants (that align with the games target market) are recruited. A casual mobile gamer coming in to participate in a hardcore shooter playtest would yield significantly different results than a hardcore console gamer. An effective method for starting a participant pool is using an online form to allow players to sign up remotely.

A good example is Ubisoft's Playtest sign up form<sup>4</sup>. The form begins with basic demographic information to understand the age, gender and availability of the player. This information is important depending on the games intended audience. Having demographic data allows researchers to filter by age and gender when selecting participants to ensure the right player profile is being tested with. Other questions focus on understanding the gamer profile of a player. Examples of these kinds of questions are:

- On which platform do you usually play?
- How many hours per week do you spend playing video games on consoles or PC?
- What are your favorite gaming genres and themes?
- At which frequency do you play video games on tablets or smartphones?
- What are your favorite mobile gaming genres and themes?

Understanding the platform the player plays most on is important when making assumptions about the players experience with different controller types. Asking how many hours per week a participant plays is important for understanding their level of intensity. With the rise of mobile gaming, it is important to segment how many hours a player engages with PC and mobile games separately. This is because a hardcore PC gamer might not play any mobile games, it would be wrong to assume that player is also a hardcore mobile gamer.

## 5.5 Developers' playtesting training

One of the purposes of training playtest coordinators is to reduce biases and increase validity of evaluation result. Although, this could be accomplished by having more evaluators, but there is a potential for multiple interpretations of user testing data. There is also a risk of not evaluating correct procedures when an evaluator is not a trained expert. In larger studios there would be often a dedicated roles for user researchers, each with different levels of required experience. In our collaboration with XL, we conducted workshops for team members to learn more about playtesting and user research approaches in general. We noticed the team members who attended the workshops were more engaged when planning and conducting the playtest and were more receptive of

our playtest report. We often noticed that they took the initiative to motivate the rest of their team to act on the playtest findings. Having a better understanding of the playtest process and methods also helped them to better formulate their goals and questions for each playtest sessions.

## 6. CONCLUSION

We presented eleven playtest cases studies on six different commercial games (with 118 participants in total). Our main takeaways highlight: (1) the adaptation of playtesting process to be resource-effective for indie development, (2) the exploration of indie developers needs to showcase the contribution of playtesting in their games and, (3) the integration of playtesting in development cycles of indie studios. These takeaways highlight challenges when evaluating independent games. These are important contributions for game developers and user researchers.

In this paper we looked at adopting playtesting methods for independent developers. We stated that there is a need for indie developers to adopt GUR techniques in their development pipelines in order to ensure a high quality product to compete in the cutthroat games market. We partnered with a local incubator XL where we playtested six different games using a variety of methods. We discussed the methods we used in playtesting their games.

In addition, to better highlight important issues for developers, it would be interesting to approach the problem from a production standpoint. A producer and programmer can allocate specific resources to each feature developed. Whenever an issue arises in the playtest, a playtest could indicate how much work would be necessary to fix the specific issue. When the user researcher is preparing their report, they can then prioritize the issues by labor-intensiveness or effect on player experience, to increase the efficiency of communication with producers.

In this paper we focused on FTUE and not the whole game. Important areas such as Camera, Challenge, Pace, and Story could be the topic of future work. Additionally, we could benefit from tools that can easily sort, filter, and organize the playtesting data to increase the overall speed and efficiency of the data analysis. One important result is the need to motivate teams that are more business-oriented to polish their FTUE instead of developing more content. Future workshops training developers on developing person's, interviews, and surveys would also be beneficial.

Ultimately, studying the GUR process and iterating on that to fits indie developers would help to better integrate user research as part of the development cycle of indie games. There is also potential for future work on using a tool for assisting with the design and reporting. A potential tool would help reduce assumptions regarding the design intent where the user researcher would have the ideal designed experience in an easily accessible format when making observations during the playtest. This would make it easier to highlight blockers and areas where the player is not experiencing the gameplay segment in the way it was initially designed.

## 7. ACKNOWLEDGEMENTS

We would like to thank Execution Labs, Faculty of Business and Information Technology at the University of Ontario Institute of Technology, and NSERC for their research support. We would also like to thank Lysiane Charest for her contribution to the studies presented in this paper.

<sup>4</sup> See: <https://playtest-secure.ubi.com/profile.php>

## 8. REFERENCES

- [1] Adams, A., & Cox, A. (2008). Questionnaires , in-depth interviews and focus groups Book Chapter and focus groups. In *Research Methods for Human Computer Interaction* (pp. 17–34). Cambridge, UK: Cambridge University Press. Retrieved from [http://oro.open.ac.uk/11909/3/9780521870122c02\\_p17-34.pdf](http://oro.open.ac.uk/11909/3/9780521870122c02_p17-34.pdf)
- [2] Bernhaupt, R. (2010). Concepts and Methods. In *Evaluating User Experience in Games*. Springer.
- [3] Browne, P. (2015). Jumping the Gap : Indie Labour and the Imagined Indie Community In The Department of Communication Studies Presented in Partial Fulfillment of the Requirements For the Degree of Master of Arts ( Media Studies ) at Concordia University Montreal , Quebec , (November).
- [4] Carter, M., Downs, J., & Nansen, B. (2014). Paradigms of games research in HCI: a review of 10 years of research at CHI. *Proceedings of CHIPlay '14*, 27–36. Retrieved from <http://dl.acm.org/citation.cfm?id=2658708>
- [5] Cheung, G. K., Zimmermann, T., & Nagappan, N. (2014). The First Hour Experience: How the Initial Play Can Engage (or Lose) New Players. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '14*, 57–66. <http://doi.org/10.1145/2658537.2658540>
- [6] Desurvire, H., Blvd, W., Rey, M., & Caplan, M. (2001). Using Heuristics to Evaluate the Playability of Games. <http://doi.org/10.1145/985921.986102>
- [7] Isbister, K., & Schaffer, N. (2008). Social Psychology and user research. In *Game Usability: Advice from the Experts for Advancing the Player Experience*. Morgan Kaufmann.
- [8] Korhonen, H., Montola, M., & Arrasvuori, J. (2009). Understanding Playful User Experience Through Digital Games. *International Conference on Designing Pleasurable Products and Interfaces*, (October 2009), 274–285. Retrieved from <http://www.jbcreativgroup.com/pdf/understanding-playful-user-experience-through-digital-93744.pdf>
- [9] Lazzaro, N. (2004). Why We Play Games: Four Keys to More Emotion Without Story. *Game Developer Conference (GDC)*, 1–8. <http://doi.org/10.1111/j.1464-410X.2004.04896.x>
- [10] Lewis-Evans, B. (2012a). Finding Out What They Think: A Rough Primer To User Research, Part 1. Retrieved from [http://www.gamasutra.com/view/feature/169069/finding\\_out\\_what\\_they\\_think\\_a\\_php?print=1](http://www.gamasutra.com/view/feature/169069/finding_out_what_they_think_a_php?print=1)
- [11] Lewis-Evans, B. (2012b). Finding Out What They Think: A Rough Primer To User Research, Part 2. Retrieved from [http://www.gamasutra.com/view/feature/170332/finding\\_out\\_what\\_they\\_think\\_a\\_php?print=1](http://www.gamasutra.com/view/feature/170332/finding_out_what_they_think_a_php?print=1)
- [12] Marshall, C., & Rossman, G. B. (1999). *Designing Qualitative Research* (5th ed.). SAGE Publications.
- [13] Medlock, M. C., Wixon, D., McGee, M., & Welsh, D. (2005). The Rapid Iterative Test and Evaluation Method: Better Products in Less Time. In *Cost-Justifying Usability: An Update for an Internet Age* (pp. 489–517). San Francisco, CA, United states: Morgan Kaufmann.
- [14] Mirza-babaei, P. Biometric Storyboards: A Games User Research Approach for Improving Qualitative Evaluations of Player Experience. PhD thesis, University of Sussex, 2013.
- [15] Mirza-Babaei, P., Long, S., Foley, E., & McAllister, G. (2011). Understanding the contribution of biometrics to games user research. In *Proceedings of DiGRA* (pp. 329–347).
- [16] Mirza-Babaei, P., Nacke, L. E., Gregory, J., Collins, N., & Fitzpatrick, G. (2013). How does it play better?: exploring user testing and biometric storyboards in games user research. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 1499–1508. <http://doi.org/10.1145/2470654.2466200>
- [17] Mirza-Babaei, P., Zammitto, V., Niesenhaus, J., Sangin, M., & Nacke, L. E. (2013). Games User Research: Practice, Methods, and Applications. *Proc. CHI'13 Extended Abstracts*, 3219–3222. <http://doi.org/10.1145/2468356.2479651>
- [18] Moosajee, N., and Mirza-Babaei, P. 2016. Games User Research (GUR) for Indie Studios. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3159-3165. DOI: <http://dx.doi.org/10.1145/2851581.2892408>
- [19] Nielsen, J. (1992). The Usability Engineering Life Cycle. *Computer*, 25(3), 12–22. <http://doi.org/10.1109/2.121503>
- [20] Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 152–158. <http://doi.org/10.1145/191666.191729>
- [21] O'Rourke, P. (2014). Ubisoft Toronto is reportedly working on five unannounced projects. Retrieved June 10, 2015, from <http://o.canada.com/technology/gaming/ubisoft-toronto-is-reportedly-working-on-five-unannounced-projects>
- [22] Pagulayan, R., Keeker, K., Fuller, T., Wixon, D., & Romero, R. (2003). User-centered design in games. In *Handbook for Human-Computer Interaction in Interactive Systems*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- [23] Pham, A. (2012). Bungie-Activision contract. Retrieved June 8, 2015, from <http://documents.latimes.com/bungie-activision-contract/>
- [24] Poels, K., & Kort, Y. De. (2007). “ It is always a lot of fun ! ” Exploring Dimensions of Digital Game Experience using Focus Group Methodology, 83–89.
- [25] Sinclair, B. (2013). GTA V dev costs over \$137 million, says analyst. Retrieved June 8, 2015, from <http://www.gamesindustry.biz/articles/2013-02-01-gta-v-dev-costs-over-USD137-million-says-analyst>

- [26] Statista. (2015a). Apple App Store: number of available apps 2015 | Statistic. Retrieved October 7, 2015, from <http://www.statista.com/statistics/263795/number-of-available-apps-in-the-apple-app-store/>
- [27] Statista. (2015b). Number of Google Play Store Apps 2015 | Statistic. Retrieved October 7, 2015, from <http://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>
- [28] Tan, D., & Vanderdonckt, J. (2015). Game User Experience evaluation - Human – Computer Interaction Series. <http://doi.org/10.1007/978-3-319-15985-0>
- [29] Thompson, C. (2007). Halo 3: How Microsoft Labs Invented a New Science of Play. Retrieved June 8, 2015, from [http://archive.wired.com/gaming/virtualworlds/magazine/15-09/ff\\_halo?](http://archive.wired.com/gaming/virtualworlds/magazine/15-09/ff_halo?)
- [30] White, G. R., Mirza-babaei, P., McAllister, G., & Good, J. (2011). Weak inter-rater reliability in heuristic evaluation of video games. Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '11, 1441. <http://doi.org/10.1145/1979742.1979788>
- [31] Zammito, V. (2011). The Science of Play Testing: EA's Methods for User Research. Presentation at GDC 2011. San Francisco, CA, United states: GDC. Retrieved from [www.gdcvault.com/play/1014552/The-Science-of-Play-Testing](http://www.gdcvault.com/play/1014552/The-Science-of-Play-Testing)
- [32] Zammito, V., Mirza-Babaei, P., Livingston, I., Kobayashi, M., and Nacke, L. 2014. Player experience: mixed methods and reporting results. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14). ACM, New York, NY, USA, 147-150. DOI=<http://dx.doi.org/10.1145/2559206.2559239>