EasyChair Preprint

№ 8899

User Experience Evaluation Methods for Games in Serious Contexts

Sotiris Kirginas

October 3, 2022

User Experience Evaluation Methods for Games in Serious Contexts

Abstract

The most important factor that leads a digital game to be successful is the quality of users' experience while they interact with serious digital games. A digital game should be "fun" and "exciting" in order to provide the best gaming experience and make the players happy (Author, 2016; Author, 2017). This need is even greater when we talk about digital games for learning purposes. As there are many different types and kinds of digital games, users' experience needs to be examined in more detail by research studies. User experience is a multifactorial concept that is difficult to be measured. Users experience can be affected by many factors such as, flow (Csikszentmihalyi, 1990; Sweetser & Wyeth, 2005), immersion (Brown, & Cairns, 2004; Ermi & Mayra, 2005), frustration or tension (Gilleade & Dix, 2004), psychological absorption (Funk et al, 2003), social gaming contexts (Bracken, Lange & Denny, 2005). This chapter aims to present a variety of methods, quantitative and qualitative/objective and subjective/empirical and non empirical, that can be used to evaluate user experience in serious games, not only during game-play but also before and after the game play, as well as, to provide insights on when to apply the various user experience evaluation methods in the development cycle (Bernhaupt, 2020).

Keywords: Digital games, evaluation, experience, quantitative, qualitative

Introduction

In recent decades people all over the world have become increasingly addicted to digital games. The fact is, today, digital games compete with traditional activities like reading books, watching movies, listening to music, surfing the internet or playing sports. (Nacke, 2009: 3). Digital games regularly attract billions of players online and offline, generating huge revenue. However, digital

games also present new research challenges for many old and new scientific disciplines. With recent advances in the field of human-computer interaction (Mandryk et al., 2006; Nacke et al, 2008), new methods are available to precisely measure how people interact with entertainment technology (Drachen and Canossa, 2009; Kim et al., 2008). With new measurement tools of players' interaction, we aim to support the traditional game development process and improve game design process.

Game designer increasingly utilize user testing along with evaluation of playability for creating digital games (Pagulayan et al., 2003; Pagulayan et al., 2004; Desurvire et al., 2004; Korhonen and Koivisto, 2006). Unlike other software programs, digital games offer often unique experience, involving elements that are difficult to evaluate. User experience in digital games can be influenced by many factors such as, flow (Csikszentmihalyi, 1990; Sweetser & Wyeth, 2005), immersion (Brown, & Cairns, 2004; Ermi & Mayra, 2005), frustration or tension (Gilleade & Dix, 2004), psychological absorption (Funk et al, 2003), social gaming contexts (Bracken, Lange & Denny, 2005).

It is common knowledge that the majority of game designers' experience comes from their own observations of players' reactions to game mechanics. Due to the time it takes to produce such individual game design knowledge, it is necessary to get a quicker understanding of players' complex behavior in response to game mechanics. In order to have a more complete view of users' experience, several recent solutions have combined event logging (Nacke et al, 2008; Drachen. and Canossa, 2009) with objective and subjective player feedback. Similarly, players' behavior is modeled in order to find "optimal spots in the game and level design" (Kim et al., 2008).

The topic of user experience (UX) measurement in games and evaluating player experience has previously been covered in the body of literature. This paper aims to examine (a) the current state of player experience measurement and (b) how measurement tools contribute to designing better games. We are especially interested in the differences between traditional research and emerging evaluation of UX, such as psychophysiological data (e.g., electroencephalography, electromyography and f acial expression assessment).

Defining User Experience

According to Dewey (1938), as reported by Almeida et al. (2013) and Calvillo-Gámez (2010), experience is both the process and the outcome of a user's engagement with the environment at a particular time. It is both an interactive (the process of playing the game) and an emotional (the consequence of playing) experience — a feeling (or a combination of emotions) that arises as a result of playing. The interaction process is how players interact when playing games. It's about the interaction of the player with other playable and non-playable characters, and items in the game environment, as well as how they make decisions. The game limits this process, which is impacted by the players' backgrounds, motives, expectations, and present emotional experience, which might change throughout the game. Players' emotional states also influence interaction processes: if their anxiety is high, they may be less able to pay attention, which could affect their ability to play and win, while if they are relaxed, they may be in a Csikszentmihalyi's state of flow (1990).

This procedure will have an impact on the gaming outcome, which refers to the game's observable effects. These effects can then have an impact on the game's result. When the emotional experience is favorable, games can elicit positive feelings in the player (for example, contentment, happiness, and excitement); when the emotional experience is unpleasant, games can elicit negative emotions (anger, despair, indifference). Positive or negative repercussions may potentially influence the interaction process by altering player incentives and behaviors (expectations). This bi-directional interaction may explain why, at times, players may experience both delight and irritation throughout the course of a game (Almeida et al., 2013). Because experience is personal and subjective, evaluating it entails assessing the user's subjective opinion.

According to Roto (2007), there are three phases of player experience: (a) expected player experience (before a player interacts with a game), (b) player experience during interaction (experience that occurs while interacting with the game), and (c) overall player experience (experience that occurs while interacting with the game) (experience after game ends). The player experience during interaction is the most important of the three phases of player experience mentioned above. Examining the player experience during interaction is crucial for improving a game since features and components that deliver a positive experience, as well as those that do not, may be noticed at this phase. To comprehend player experience during the interaction phase, we must consider three factors: the human aspect (predispositions, expectations, needs,

motivation, mood, etc.), the system aspect (e.g. complexity, purpose, usability, functionality, etc.), and the contextual aspect (or environment) in which the interaction takes place (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc). (Hassenzahl & Tractinsky 2006; Reiter et al. 2014).

Methods to Evaluate UX in Games

Designing and developing digital games is a demanding and difficult process. Game developers should produce games which are rewarding, amusing, and appealing to consumers in order to secure financial success. As a result, it's critical to understand how various players behave and interact throughout games. It is critical to understand the target players and their gameplay experience throughout game design in order to produce a higher player experience and perhaps enhance game reviews and financial success.

A survey from the Entertainment Software Association (ESA) indicated that digital games have become an important feature of the game industry in recent decades. Due to their fast rising market share and a number of other variables such as new business models, broader player demographics, and unique controller interfaces and platforms, digital games are a prominent area for research (Mirza-Babaei, 2015). Despite their promise, video games nevertheless confront several obstacles. Developers should guarantee that each game is positively appreciated by the consumer, in addition to increasing production and design expenditures for large-scale releases.

As a result, the opportunity is broader; however, a deeper understanding of player demographics and platforms is required to appeal to this market. According to Mirza-Babaei (2015) single player stereotypes are generally disappearing in the industry (for instance, a picture of a teenager being addicted to digital games) to a new one, where multiple players play multiple devices at the same time. Different ways of interaction exist in modern digital games, offering more opportunities at player interaction.

Through the growing field of Games User Research (GUR), developers evaluate their games for usability and user experience to improve the gameplay experience. Games User Research borrows user research techniques from

Human-Computer Interaction (HCI) and psychology, such as behavioral observation, interviews, questionnaires, and heuristic evaluation. In spite of progress in using user research methods to understand the usability of productivity applications, researchers and practitioners are still facing challenges applying these methods to video games. Digital games provide unique characteristics that prevent the application of most conventional user research methodologies to the evaluation of the player experience.

As a result, user methodological approaches have been modified and improved to better suit the objectives of developing games. These methods seek to provide games user researchers a combination of qualitative and quantitative methods to choose from, based on their research context and their participants needs. One of the main issues confronting UX and usability evaluation in games is determining the optimum combination of various methodologies and combining the data from each into a relevant report for game creators.

Analysis of Methodologies

Several measurement and evaluation methodologies are available to measure and evaluate the users' experience in digital games. As follows, these methodologies are classified in different ways:

(α) Quantitative vs qualitative evaluation

Qualitative methods are used to explore and understand players' perceptions and interactions. Players' experiences are usually recorded in non-numerical data. In contrast, the quantitative method makes use of numerical data (Neill, 2009). Quantitative approaches show the level of engagement and interest by providing statistics, while qualitative approaches capture player experiences during gameplay. There are times when players lack emotional expression and do not speak freely when evaluating verbally or non-verbally. It is difficult for players to concentrate and discuss their experiences simultaneously while playing games. When evaluating a project, both methods should be used to achieve objective and comprehensive results.

In any research the researcher needs to make a primary but basic methodological choice between the quantitative and the qualitative approach (or their combination) to investigate his topic. The quantitative approach is the one that allows us to know "what is happening?" while the qualitative to examine "why is it happening?". The aim of qualitative research is "to discover the views of the research population, focusing on the perspectives from which individuals experience and feel about events" (Bird, et al., 1999, p. 320). In summary, qualitative assessment involves categorizing and evaluating qualitative data to assist us in analyzing and interpreting game events, users' behavior, and players' experiences. Qualitative data collection can lead us down such paths, whereas quantitative data collection cannot, particularly when it comes to user experience.

Subjective vs. Objective evaluation

Tools for uesrs' experiences measurement can be classified into two categories according to their reliability: objective and subjective.

Objective assessment tools provide accurate data, which are objective and free from any subjective judgment of the participants, because precisely they are recorded by machines (Cacioppo et al., 2007). Objective data are recorded automatically and continuously, without disturbing the participant and without affecting their physics.

On the contrary, subjective tools, precisely because they are completed by the users themselves, contain subjectivity, so they have a reduced reliability compared to objective tools. An objective assessment tool measures the expressive or phychophysiological aspect of the user's experience through facial expressions and phychophysiological data collected, while a subjective tool assesses the subjective feeling of the user's experience through self-report, rating scales, and verbal protocols.Short-term vs. Long-term evaluation

In the early stages of game development, measuring users' initial and momentary experiences is important for getting feedback (Vermeeren et al., 2010). Additionally, user experience is known to change over time (Fenko et al., 2010). Thus, it is necessary to use tools that measure the experience over a longer period of time in order to gain more reliable information about a game's

playability. In this way, a game designer can gain insight into how a player interacts with their game. At present, most user experience research focuses on short-term evaluations and, hence, primary evaluations of the initial user experience of new games. Nevertheless, the relationship between a user and a game evolves over time, so long-term usage evaluation is crucial for a game's success.

Having these different categorizations is important since the reasons we want to measure user experience can vary from survey to survey. In some instances, we may want to measure qualitative attributes resulting from the player's experience, while in others, we may want to measure quantitative attributes. Similarly, we may want to measure the player experience at a specific point in the game, such as when the player wins a significant player, or we may want to evaluate it over a longer period of time. To improve the quality of the game, we sometimes want to measure the experience of players, i.e. the subjects, so we know how well they know the game. Other times, we may want to ask experts about their experience of a game so we can improve.

Overview of the main methodologies

A variety of tools and methodologies are provided in order to uncover the quality of the experience produced by a game, either in order to improve it or to utilize the game for the purposes of education, training, awareness, changing behavior of subjects. In Table 1, all UX assessment methods are summarized along with how they fall into one or more of the above-mentioned categories.

Table 1: Overview of the main instruments and methodos

Methods/Instruments	Quantitative	Qualitative	Subjective	Objective	Short-term	Long-term
Leuven Scale		✓		✓		✓
Focus Groups		✓		✓		✓
Electro-dermal activity	✓	✓	✓		✓	✓
Cardiovascular activity	✓	✓	✓		✓	✓
Electroencephalography	✓	✓	✓		✓	✓
Electrooculogram	✓	✓	✓		✓	✓

Facial expression assessment	✓	✓	✓		✓	✓
Game Experience Questionnaire	✓			✓		✓
Game Engagement Questionnaire	✓			✓		✓
FunToolkit	✓			✓		✓
Sef-Assessment Manikin	✓			✓		✓
Emo-watch	✓			✓		✓
Memoline		✓		✓		✓
UXCurve		✓		✓		✓

(a) Observation

Observation is a profoundly qualitative research methodology that can be incorporated into a wide variety of qualitative and quantitative research projects. Researchers can collect a lot of data and information from their observations by observing users during some activity and then analyzing it. It is possible to collect valuable data to interpret the topic researchers are exploring when observation is combined with other methodologies and techniques.

The researcher is required to possess special skills, and the observation procedure involves several methodological risks, primarily regarding its validity and reliability, since the issue of objectivity and impartiality is always present. As such, it is generally preferable for novice researchers to combine this technique with another, like interviewing, in order to collect all the data he needs or to illuminate certain aspects of the research or to triangulate his information.

The Leuven Well-being and Involvement scale (Figure 1): A typical observation tool that can provide important insights into the user experience during gameplay is the Leuven Well-being and Involvement scale. The Leuven scale was developed under the supervision of Professor Ferre Laevers at the Research Center for Experiential Education of the University of Leuven, Belgium (Laevers, 2005; Laevers & Heylen, 2003), the Leuven Well-being and Involvement scale aims to assess and measure two important quality indicators for early childhood childcare. Since then, it has been widely used in other levels of education as well, capturing the "Well-being" and "Involvement" of the child during his care or teaching in the classroom.

The concept of well-being refers to how a child acts in a relaxed and spontaneous manner, without tension and nervousness. This is essential for seamless monitoring and the ability to assimilate what is happening in the classroom, since that also refers to self-esteem, adaptability, and self-awareness.

Figure 1: The Leuven Well-being and Involvement scale

The Leuven Scale for Well-being

level	Well being	Signals
1	Extremely low	The child clearly shows signs of discomfort such as crying or screaming. They may look dejected, sad, frightened or angry. The child does not respond to the environment, avoids contact and is withdrawn. The child may behave aggressively, hurting him/herself or others.
2	Low	The posture, facial expression and actions indicate that the child does not feel at ease. However, the signals are less explicit than under level 1 or the sense of discomfort is not expressed the whole time.
3	Moderate	The child has a neutral posture. Facial expression and posture show little or no emotion. There are no signs indicating sadness or pleasure, comfort or discomfort.
4	High	The child shows obvious signs of satisfaction (as listed under level 5). However, these signals are not constantly present with the same intensity.
5	Extremely high	The child looks happy and cheerful, smiles, cries out with pleasure. They may be lively and full of energy. Actions can be spontaneous and expressive. The child may talk to him/herself, play with sounds, hum, sing. The child appears relaxed and does not show any signs of stress or tension. He /she is open and accessible to the environment. The child expresses self-confidence and self-assurance.

An active involvement in the spontaneous search for knowledge is regarded as a prerequisite for the child's cognitive growth and mental development. The scale consists of five (5) levels that assess and measure both well-being and engagement. If there is a persistently low level of well-being and/or engagement, it is very likely that a child's development is at risk. On the other hand, the better a child's well-being and engagement, the more they will be able to develop. In the presence of high levels of well-being and engagement, we can be certain that deep learning takes place.

b) Focus groups

Focus groups are a form of qualitative and subjective research. During a focus group, a group of people gather in a room to engage in a guided discussion of a topic. It is a semistructured interview process in which a small group of individuals, usually six or eight, discuss a specific study topic (Author, 2016). Krueger & Casey (2000) describe the focus group method as a means to “obtain perceptions on a defined area of interest in a permissive, nonthreatening environment” (p. 5). To produce qualitative data on the research topic, the moderator regulates the discussion more or less according to its structure.

Consider a research project on the user experience of a digital game. A deeper interview with the players might be necessary, but before we do that, we want to see what types of questions will work, and if the players may raise issues we are not considering, so we can include them among our questions. Randomly asking game players about their feelings and experiences with a game would be a great way to find out what they like and dislike about it.

In a focus group, participants are selected on the basis of their relevance and relationship to the topic. As a result, they are not considered statistically representative of any significant population as they are not selected according to strict probability sampling methods. Instead, participants are selected by sampling, advertisement, or snowball sampling, depending on the type of person and the characteristics the researcher wants to include. There are several advantages of focus groups: It is a socially oriented research method, captures real-life data in a social setting, is flexible, has high validity, provides rapid results, and costs nothing to conduct. There are also some disadvantages of focus groups: Researchers have less control than in individual interviews, data can sometimes be difficult to analyze, the moderators need certain skills, and the discussion must be conducted in a conducive environment.

(c) Psychophysiological measurements

Psychophysilogical measurements rely on both quantitative and quantitative research to better understand users' experience. As users' experience during the game can have a significant impact on digital game's playability, physiological data can be really useful to assess the emotional state and performance of the players, especially when they are correlated with the subjective measurements (Author, 2021). To date findings have only been reported for First Person Shooters games (Nacke et al., 2010; Drachen et al., 2010). The question is whether physiological and subjective measures could prove equally reliable for other types of digital games. The most important methods of evaluating user's experience with physiological methods are the following:

Electrodermal activity (EDA) is perhaps the most commonly used physiological measurement. Often, the literature refers to it as galvanic skin response or skin conductance. Sweat gland secretions during gaming are indicators of positive arousal and mental activity.

Cardiovascular activity readings are important physiological measurements of human activity. Cardiovascular activity readings measure heart rate, heart rate, and heart rate variability.

Electromyography (EMG) provides measurements of muscles' electrical. When an individual is overly anxious, skeletal movements are observed as a sign of involuntary muscle contractions during intense mental activity, intense emotions, and cognitive load. (Tortora & Derrickson, 2009).

Facial expression evaluation analyzes human expressions during an activity and measures the human's basic emotional states (joy, sadness, anger, surprise, disgust, etc.).

Electroencephalography (EEG) is performed using special electrodes placed on the participant's head during the test. The brain activity is then measured using frequency wave patterns that represent different mental activities. Since electrodes are used in electroencephalography, it is strictly a laboratory measurement.

Biofeedback measuring device

A device designed and built in the Laboratories of New technologies of the department of Communication and Media Studies, University of Athens. This device consists of a sensing part that is accommodated onto a typical computer mouse, an analogue electronic circuit that feeds the processed signal to a typical home computer and finally a software component that translates the measurements into an appropriate format. STC is acquired seamlessly by contact of the thumb and ring fingers with Al-Si ring sensors located on the left and right sides of the computer mouse (Figure 2), respectively.

Figure 2: The biofeedback measuring device.

Heart rate is also detected by reflective near-infrared sensors located in the centre of the ring sensors (Fig. 5), based on the principle of reflective absorption occurring during the changes of the colouration of the skin caused by the pulsation of the blood in the tissue.

FaceReader

A software application called FaceReader was developed by Noldus Information Technology (2010). The FaceReader software uses algorithms to rate facial images according to seven basic emotional states - happy, sad, angry, surprised, afraid, disgusted, and "neutral emotional state.". These seven emotions are rated from 0 (not at all) to 100 (perfect match). FaceReader “is an effective instrument to measure emotional experience during human–computer interaction, as strongly suggest that more effective and well-designed systems elicit more positive emotions and less arousing fallings than less effective applications” (Author, 2021 p.3).

(e) Self-assessment methods

Self-assessment methods are subjective, most often quantitative and either short or long term. They provide players with the ability to self-evaluate or make judgments about their experience and the games they play based on specific self- assessment tools. Their great advantages are ease of use and the use in many situations. However, their disadvantage lies in the subjectivity of the judgments, which can be affected by a number of factors, including bias,

differences in age and gender, economic and social status, and past experiences, among others.

Game Experience Questionnaire (GEQ)

It is a tool designed specifically for young children (8-12 years old) to assess their gaming experiences. The Game Experience Questionnaire (IJsselsteijn et al., 2008) assesses 7 different dimensions of gaming experience (Immersion, Flow, Effectiveness, Intensity, Challenge, Positive emotion, Negative emotion) Each of the 7 dimensions is distinguished into 5 sub-themes rated on a 5-point Likert scale. The Game Experience Questionnaire is divided into three separate modules each of which deals with a different experience: 1) Core module which evaluates the user's experience while playing the game. 2) Social presence module which evaluates the user experience while playing a game with others. 3) Post-game module which evaluates the user's experience after completing the game.

It has the advantage of measuring different aspects of the game experience (immersion, flow, effectiveness, intensity, challenge, positive emotions, and negative emotions), assessing the experience during and after the game, and assessing social presence as well. As it covers such a large area, it can be difficult to complete by all the researchers, so many researchers only use some of the modules.

Game Engagement Questionnaire

The Engagement Questionnaire (Figure 3) was created by Nicola Whitton (2010) to assess students' degree of engagement in an activity they have completed in relation to another or other activities. The Engagement Questionnaire consists of a set of 19 questions. Each question explores one of the four concepts flow, presence, absorption, and immersion, to measure users' engagement in digital games.

1	I lose track of time
2	Things seem to happen automatically
3	I feel different
4	I feel scared
5	The game feels real
6	If someone talks to me, I don't hear them
7	I get wound up
8	Time seems to kind of stand still or stop
9	I feel spaced out
10	I don't answer when someone talks to me
11	I can't tell that I'm getting tired
12	Playing seems automatic
13	My thoughts go fast
14	I lose track of where I am
15	I play without thinking about how to play
16	Playing makes me feel calm
17	I play longer than I meant to
18	I really get into the game
19	I feel like I just can't stop playing

Figure 3: The Engagement Questionnaire

Fun Toolkit

The Fun Toolkit (Figure 4) was developed by Read & MacFarlane (2000). It consists of three separate questionnaires, the characteristics of which are described in detail below:

α. Smileyometer: It is a measurement scale based on a five-point Likert scale, with ratings from 1 “Poor” to 5 “Excellent”, as shown in Figure . The Smileyometer can be used both before and after the child's experience with a digital application be it an educational software or a website or a digital game. Using it before engaging with the application, we can gather information about the children's expectations from the game. Using it latter, we can collect information about the fun of the game or the emotional experience of the players. If we wish to evaluate several digital applications at the same time, then the Smileyometer must be used separately for each. The main advantages of the Smileyometer Questionnaire are its simplicity of use, its quick completion, its limited reading skill requirement, and the absence of a written word production requirement.

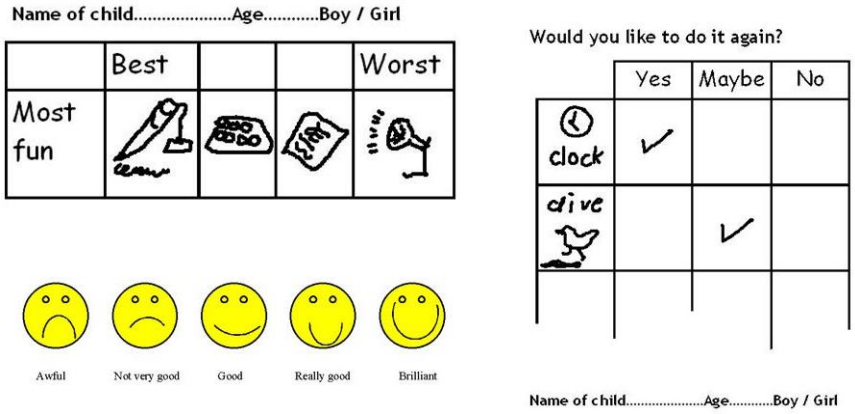

Figure 4: The Fun Toolkit

b. Fun Sorter Table: A Fun Sorter table generally compares a set of products, whether they are educational software or digital games, as in our case. For a survey on children's ratings of digital games, children compare and rank them from best to worst or from easiest to hardest or from what they intend to play again to what they intend to play less.

c. Again and Again Table: The questionnaire consists of a table in which children mark whether they experienced each activity “Yes”, “Maybe”, or “No”. The idea for this tool comes from the field of psychology where it is argued that we are more likely to return to an activity we liked again and again if we like it. In the present study children were asked "Would you like to play with the toy again?" and they had to answer accordingly.

Self-Assessment Manikin (SAM)

The Self Assessment Manikin (SAM) is a system for evaluating three dimensions of gaming experience: valence, arousal and dominance (Lang, 1985). It uses pictorial scales, illustrating cartoon creatures. Figure 5 presents a visualized version of the Self Assessment Manikin tool used in the research. The upper scale assesses the players' pleasure and the lower their arousal. Both scales are nine-point and take values from 1 to 9, with 5 representing the middle of the scale. Although it is stated that it is a weighted method, there are insufficient studies that support this claim. Its advantages include ease of completion and its ability to be used in different circumstances. The

disadvantages are what all objective assessment tools suffer from: objectivity of judgment and difficulty in matching experience with graphic.

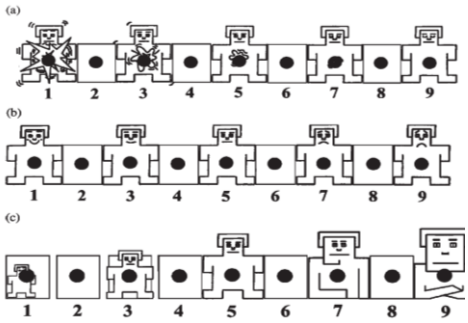

Figure 5: The Self Assessment Manikin

UX Curve

The UX Curve (Figure 6) is a tool for retrospectively evaluating user experiences. There is a timeline and a horizontal area in which the user can graph his positive and negative experiences. The advantage of UX Curve is that it allows the user to design the most immersive game experience. Nevertheless, its disadvantage is that it relies on retrospective memory from the game ratherthan reality for its completion.

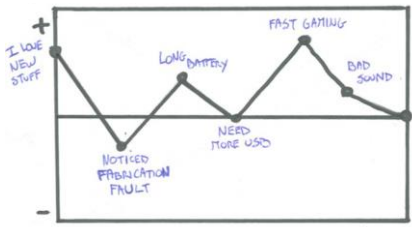

Figure 6: Example of a completed UX Curve

MemoLine

The memoLine (Figure 7) is actually a timeline that can be used to perform retrospective evaluations. There are as many frames as there are periods during

which the user plays a game. Considering that the tool is meant to be used by children, the experiences they gain are represented by three different colors: green denotes positive experiences, red denotes negative experiences, and gray indicates periods when the game is not being played, such as weekends. Users receive questionnaires for each of these game scenarios: Usability, Challenge, Quantity, and General Impression

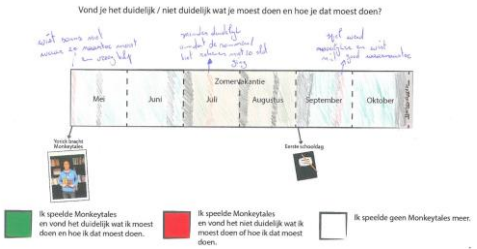

Figure 7: Example of a completed MemoLine

The questionnaires above are certainly not the only ones. There are a large number of other relative questionnaires such as Emo-watch, EGameFlow, Gameful Experience Questionnaire, Model for the Evaluation of Educational Games (MEEGA+), Game User Experience Satisfaction Scale (GUESS), iScale CORPUS (Change Oriented analysis of the Relationship between Product and USer) and many other.

Conclusions

This chapter aims to map and record the available user experience evaluation methodologies and the tools that are available for improving the user experience in digital games.

This chapter is intended to serve as a tool for digital game designers and game studies researchers to evaluate existing digital games or evaluate the user experience of digital games under development to improve the resulting experience and reach its optimum level. Since the purpose of a user experience evaluation is to record and interpret the experience experienced by users while interacting with a digital game, it is imperative that this recording is accurate and reliable in order for its results to have substance and be useful. Furthermore,

an evaluation must be able to identify the situations and factors that influence the user experience and make it more or less positive. In that case, we will be able to make the necessary adjustments and changes in order to make the user experience better. According to what has been discussed in this chapter, user experience evaluation tools can be classified into four groups, objective - subjective, quantitative - qualitative, and short-term - long-term.

To summarize, since user experience is multidimensional and difficult to measure, it is wise to use tools that have different characteristics. When measuring and evaluating user experience in digital games, it is best to employ tools from different methodologies, such as quantitative tools combined with qualitative evaluation tools or objective tools combined with qualitative evaluation tools. Taking advantage of each methodology in this way will increase the reliability of the results. Utilizing tools from only one methodology may negatively affect our evaluation efforts if we choose to leverage those tools.

Last but not least, for better understanding and in order to interpret the experience derived from a serious digital game, the methodology used to evaluate the user experience plays a very important role. Future research will evaluate digital games using different methodologies and tools. These studies will ultimately be aimed at finding the most effective combination of tools and methodologies for measuring a game's gaming potential.

References

Author, (2016). authors' reference removed for anonymity, to be restored in the final version of the manuscript.

Author, (2017). authors' reference removed for anonymity, to be restored in the final version of the manuscript.

Author, (2021). authors' reference removed for anonymity, to be restored in the final version of the manuscript.

Bernhaupt, R. (2010). Evaluating User Experience in Games: Concepts and Methods. Springer, Berlin.

Bernhaupt, R., IJsselsteijn, W., Mueller, F. F., Tscheligi, M. and Wixon, D. (2008). Evaluating user experiences in games. In CHI '08 extended abstracts (Florence, Italy). ACM, pp. 3905-3908.

Bird,M., Hammersley,M., Gomm,R., & Woods,P. (1999). Educational Research in Action/ Fragkou E: Translate in Greek. Educational Research in Practice-Study Manual. Patras.

Bracken, C., Lange, R. L., & Denny, J. *Online video games and gamers' sensations of spatial, social, and copresence*. FuturePlay 2005, East Lansing (2005).

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624–634. https://doi.org/10.1016/j.jesp.2009.02.016

Brown, E., Cairns, P.: A grounded investigation of immersion in games. In: ACM *Conf. on Human Factors in Computing Systems*, CHI 2004, ACM Press, pp. 1297-1300 (2004)

Cacioppo, J., Tassinary, L., Berntson, G. (2007). *Handbook of Psychophysiology (3rd ed.).* New York, USA: Cambridge University Press.

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience (First ed.).* New York: Harper Collins. p. 39.

Cooper, K.: Software Engineering Perspectives on Computer Game Development. CRC Press, Taylor & Francis, Boca Raton, USA (2021)

Desmet, P.M.A., Overbeeke, C.J., Tax, S.J.E.T. (2001). Designing products with added emotional value: development and application of an approach for research through design. The Design Journal, 4(1), 32-47.

Desurvire, H., Caplan, M. and Toth, J. A. (2004). Using heuristics to evaluate the playability of games. In CHI '04 extended abstracts (Vienna, Austria). ACM, pp. 1509-1512

Drachen, A., Nacke, L.E., Yannakakis, G., Lee Pedersen, A. (2010). Correlation between heart rate, electrodermal activity and player experience in First- Person Shooter games. In *Stephen N. Spencer (ed.) Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games.* Los Angeles, CA, USA: ACM, 49-54.

Drachen, A. and Canossa, A. (2009). Towards Gameplay Analysis via Gameplay Metrics. In Proc. of MindTrek (Tampere, Finland, October 1-2). ACM

Ermi, L. & Mayra, F.: Fundamental Components of the Gameplay Experience: Analysing Immersion. In: de Castell and Jenson (eds), *Proceedings of Chancing Views – Worlds in Play*. Digital Games Research Association's Second International Conference. Vancouver (2005)

Fenko, A., Schifferstein, H.N.J., Hekkert, P., 2010. Shifts in sensory dominance between various stages of user-product interactions. *Applied Ergonomics 41*, 34–40.

Funk, J.B., Pasold, T., and Baumgardner, J.: How Children Experience Playing Video Games. *In: Proceedings of the Second International Conference on Entertainment Computing* Pittsburgh, Carnegie Mellon University, pp. 1–14 (2003)

Gilleade, K. M. & Dix, A.: Using frustration in the design of adaptive videogames, 228-232. *In: Proceedings of the 2004 ACM SIGCHI* International Conference on Advances in Computer Entertainment Technology - ACE '04 (2004)

Hirschman, E.C., & Holbrook, M.B. (1982). Hedonic Consumption: Emerging concepts, methods and propositions. *Journal of Marketing 46*, 92–101.

IJsselsteijn, W. A., Poels, K., & de Kort, Y. A. W. (2008). The game experience questionnaire: Development of a self-report measure to assess player experiences of digital games, deliverable 3.3, The Netherlands: FUGA technical report, TU, Eindhoven.

IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). The Game Experience Questionnaire. Technische Universiteit Eindhoven.

Isbister, K. and Schaffer, N. (2008). *Game Usability: Advice from the experts for advancing the player experience.* Morgan Kaufmann Publishers, Burlington, MA

Jordan, P.(2000). *Designing Pleasurable Products: An Introduction to the New Human Factors*. Taylor and Francis, London

Karapanos, E., Martens, J.-B. and Hassenzahl, M. (2012). Reconstructing experiences with iScale. International Journal of Human-Computer Studies. 70, 11 (Nov. 2012), 849–865.

Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B., Pagulayan, R. J. and Wixon, D. (2008). Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In Proc. of CHI 2008 (Florence, Italy). ACM, 443-452.

Korhonen, H. and Koivisto, E. M. I. (2006). Playability heuristics for mobile games. In Proc. of Conf. on HCI with mobile devices and services (Espoo, Finland). ACM, pp. 9-16

Krueger, R.A. and Casey, M.A. (2000). *Focus groups A practical guide for applied research.* Sage Publications Inc, Thousand Oaks..

Kujala, Roto, Mattila, Karapanos, Sinnela, (2011). UX Curve: A method for evaluating long-term user experience, Interacting with Computers 23 (2011), pp. 473–483.

Laevers, F (ed) (2005) Well-being and Involvement in Care Settings: A process orientated self-evaluation instrument for care settings Leuven: Kind & Gezin

Laevers, F. & Heylen, L. (2003). *Involvement of Children and Teacher Style. Insights from an International Study on Experiential Education*. Edinburgh: Studia Pedagogica

Lang, P. J. (1985). The cognitive psychophysiology of emotion: Fear and anxiety. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 131–170). Lawrence Erlbaum Associates, Inc.

Mandryk, R. L., Atkins, M. S. and Inkpen, K. M. A (2006). Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In Proc. of CHI 2006 (Montréal, Québec, Canada, April 2006). ACM, pp. 1027-1036.

Mirza-Babaei, P. (2015). Getting Ahead of the Game: Challenges and Methods in Games User Research. User Experience Magazine, 15(2). Retrieved from https://uxpamagazine.org/getting-ahead-of-the-game/

Nacke, L. E. (2009). *Affective Ludology*: *Scientific Measurement* of *User Experience* in *Interactive Entertainment*, Blekinge Institute of Technology.

Nacke, L., Ambinder, M., Canossa, A., Mandryk, R. and Stach, T. (2009). Game Metrics and Biometrics: The Future of Player Experience Research. In Proc. of Future Play @ GDC (Vancouver, BC, Canada.

Nacke, L. E., Drachen, A., Kuikkaniemi, K., Niesenhaus, J., Korhonen, H. J., Hoogen, W. M. v. d., Poels, K., IJsselsteijn, W. A. and Kort, Y. A. W. d. (2009). Playability and Player Experience Research. In Proc. Of DiGRA 2009: Breaking New Ground (London, UK).

Nacke, L. E., Grimshaw, M. N., Lindley, C. A. (2010). More Than a Feeling: Measurement of Sonic User Experience and Psychophysiology in a First- Person Shooter Game. Interacting with Computers, 22 (5), 336-343.

Nacke, L., Lindley, C. and Stellmach, S. (2008). Log who's playing: psychophysiological game analysis made easy through event logging. In Proc. of Fun and Games, 2nd Int. Conf. (Eindhoven, The Netherlands, October 20 - 21). Springer, pp. 150-157

Neill, J.(2009, March). Qualitative & Quantitative Research, Retrieved from http://wilderdom.com/research/QualitativeVersusQuantitativeResearch.html

Pagulayan, R., Keeker, K., Wixon, D., Romero, R. L. and Fuller, T. (2003). User-centered design in games. In The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. L. Erlbaum Associates Inc., New York, NY, pp. 883-906.

Pagulayan, R., Steury, K. R., Fulton, B. and Romero, R. L. (2004). Designing for fun: user-testing case studies. In Funology: From Usability to Enjoyment. Kluwer Academic Publishers, Norwell, MA, USA, pp. 137- 150

Rademacher, U, Koschel, K-V (2006) Coming to terms with emotions. Available at: https://www.researchgate.net/publication/242558998_COMING_TO_TERMS_WITh_EMOTIONS

Read JC, MacFarlane SJ (2006) Using the Fun Toolkit and other survey methods to gather opinions in child computer interaction. Interaction Design and Children, IDC2006, ACM Press, Tampere

Read JC, MacFarlane SJ, Casey C (2001a) Expectations and endurability—measuring fun. Computers and fun 4, York, England

Read JC, MacFarlane SJ, Casey C (2001b) Measuring the usability of text input methods for children, HCI2001. Springer, Lille

Read JC, Gregory P, MacFarlane SJ, McManus B, Gray P, Patel R (2002a) An investigation of participatory design with children—informant, balanced and facilitated design. Interaction Design and Children, Shaker Publishing, Eindhoven

Sim, G., Nouwen, M., Vissers J., Horton, M., Slegers, K., Zaman, B. (2016). Using the MemoLine to capture changes in user experience over time with children, International Journal of Child-Computer Interaction, Volume 8, pp. 1-14.

Sweetser, P., Wyeth, P. (2005). GameFlow: *A model for evaluating player enjoyment in games.* Computers in Entertainment, 3(3), 1-24.

Tortora, G., Derrickson, B, (2009). Principles of Anatomy and Physiology 12th edition, *Biological Sciences Textbook.*

Vissers, J., De Bot, L. and Zaman, B. (2013). MemoLine: Evaluating long-term UX with children. Proceedings of the 12th International Conference on Interaction Design and Children, New York, pp. 285-288.

Vermeeren, A., Lai-Chong Law, E., Roto, V., Obrist, M., Hoonhaut, J., Väänänen-Vainio-Mattila, K., 2010. User experience evaluation methods: current state and development needs. *In Proceedings of the NordiCHI Conference*.

Whitton, N. (2010) Learning with digital games: a practical guide to engaging students in higher education. New York: Routledge.