User Experience Methods and Games: Lessons Learned

Remigius Fierley, Stephan Engl

The increasing expansion of the digital games market has led to a growing need for empirically based, high-quality evaluation of playability and game experience. For practitioners in the field of user experience the question arises, to what extent are their already established methods suitable for the evaluation of games. Based on our experiences with recent projects, we will discuss relevant aspects concerning the application of user experience methods in the games sector. Our purpose is to raise awareness of the particularities in this research field.

Evaluation Methods, Games, User Experience (UX), Best Practice

1. INTRODUCTION

Unlike other sectors of the entertainment industry the market for digital games is growing steadily. According to the Federal Association for Information Technology, Telecommunications and New Media (Bundesverband für Informationswirtschaft, Telekommunikation und neue Medien, i.e. BITKOM) an average growth rate of 13 % was achieved solely in Germany from 2005 to 2008 for the total market volume of games. Covered by this number are soft- and hardware sales of computers, game consoles and handheld gaming devices [1]. Recently, established publishers have been confronted with a slowdown of this extraordinary growth, but this decrease has been balanced by an increased rise of small companies working in innovative areas like online- and browser games [10], [11].

The development of new platforms and distribution channels has established digital games as a mainstream leisure activity. Innovative interaction concepts, such as gesture recognition, multi-touch interfaces, augmented reality applications and brain-computer interfaces have found their way into the hands of long-standing and new gamers. A growing variety of games from diverse genres has led to an increasing differentiation within the gamer population. For example, casual gamers and female gamers represent target groups which have been increasingly addressed by the game industry only in the recent years.

In the light of these developments we expect a growing need for user research and games testing. We would like to illustrate this with a few examples. An expanding market boosts competition because of a higher number of developers, publishers and a vast number of games rivalling for attention. Especially in markets with relatively low entry barriers like online and mobile markets - a large number of competitors exists today. Developers face higher barriers for market entry with console games due to higher technical and financial requirements. But even in this high profile market competition has been gaining strength and an increasing number of games are rivalling each other for the steadily growing gamer community and its increasingly differentiated needs. The integration of online services into console products will further break boundaries between the available gaming platforms and hence new products will enter the market. Similarly, technological developments in online and browser technology will allow these systems to compete with providers of traditional PC and console games.

Overall, this competitive pressure will increase the demand for the high quality development of games and especially the empiric quality assurance of games. Naturally, questions concerning the player experience from the customer’s (i.e. players) perspective will arise along these lines. Given the increased number of product developments, it must be assumed that the demand in user experience research and testing cannot be covered (or at least not fully covered) by developers and publishers alone. Therefore, adequate external service providers will be needed.

It can also be observed, that providers unfamiliar with game development are increasingly buying games to make them available on their (online-) platforms. These providers often purchase games for a specific marketing campaign or they permanently include the games in their platforms with the goal to address specific target groups over a longer period of time. In this constellation external quality control is important, because the providers usually lack the expertise to carry out quality assurance that goes
Another relevant factor that calls for sophisticated user research and testing in games is the importance of innovation in digital game development. In comparison to other products and markets, novel and innovative interaction concepts are strongly expected and demanded by players. The relative importance of innovation certainly varies depending on the platform, genre and gamer community. Nevertheless, the significance of innovation is exceptionally high for digital games. With regard to novel forms of interaction and innovative interface designs, experiences from prior projects and best practises can only be applied in limited ways, if at all. Therefore, it is important to iteratively test the games with actual gamers feeding back the results into the development process.

Most developers of digital games employ classic software-testing methods such as modul tests, stress tests, and compatibility and localisation tests. Likewise, more often metrics are used to measure in-game events of players during gameplay, in order to statistically analyze this data afterwards. Actual users are usually only involved in the development process for expanded bug tracking during Alpha or Beta testing phases.

If game development is oriented towards a positive game experience, however, more user (i.e. player) insights are needed, be it for the evaluation and further development of early concepts and designs, to optimize interactions and processes, or to track motivation and player experience in the long run.

Usability and user experience research provides a variety of established research methods and procedures able to generate insights for classic products such as software, websites or mobile applications. Given the special characteristics of ‘user experience’ in games, the question arises as to which extent established methods can be applied to the evaluation of games? To which extent traditional user experience methods have to be adapted to suite the peculiarities of games?

Based on our own experiences we will discuss some of the special characteristics of games and their implications for the application of user experience methods in game testing. Before that, we want to highlight the role of usability and user experience in digital games.

2. USABILITY AND USER EXPERIENCE IN GAMES

Computer and video games represent a distinct kind of digital product. To distinguish games from other products within the field of human-computer interaction, special phrases have been coined. Functional aspects of usability, such as navigation, controls, design of the interface etc. are equally important in games as in other digital products. For the classic usability aspects, like effectiveness, efficiency and satisfaction, the term ‘playability’ is often used in the game sector instead (see [8], [22]). However, the users’ experiences while interacting with a digital product go beyond the range that classic productivity measures can capture. In Human Computer Interaction a product is not only considered in terms of usability but also in terms of user experience. Therefore, it is equally important to optimise the users’ subjective experience by a design that emphasizes the non-instrumental qualities of the product, such as beauty, originality and self-expression (see [12], [13]). In the field of games the actual subjective user experience is even more important for the success of a product than in other areas. While using a word processing program or an e-commerce website, a user can achieve his respective objectives satisfactorily, even when the products do not provide an outstanding user experience. However, the user experience itself is the overarching goal when playing games.

Different concepts, such as Flow (complete absorption in an activity resulting from the optimal balance between skill and challenge, see [4], [6], [23]) or Immersion (reduced self-perception due to an intriguing and challenging virtual environment, see [3], [7], [16]) are used to describe central elements of player experiences. The terms ‘player experience’ or ‘gameplay experience’ therefore emphasise different dimensions of a players’ experience, such as motivation, challenge, excitement, fun, aesthetic appeal and other related factors (see [5], [9], [17], [19], [20], [21]). To sum it up, game experience is a user experience, which clearly is more multifaceted and complex than the many other experiences users may have while interacting with products.

Considering the specific nature of player experience, the question arises as to which extent our established user experience research methods can be applied to games.

3. USER EXPERIENCE METHODS AND THE EVALUATION OF GAMES

In the following section we want to discuss a selection of practical experiences, which we have gained in past projects with games and which we consider relevant to the application of user experience methods in this area. The description of these aspects addresses external service providers as well as in-house professionals who have a ‘classic’ usability and user experience background. The following description shall raise awareness for the special characteristics of games as a test subject. We will discuss the test setting, data collection, free interaction, thinking aloud, standardized questionnaires and iterations.
3.1 Test Setting

As usability and user experience professionals we take the context of use into account when designing tests. In most cases, however, the context of use is of minor relevance when designing the test setting. Regular Business to Customer (B2C) or Business to Business (B2B) products and applications are mostly used by a single user at his work place or his personal computer at home. For the evaluation of these products it is usually sufficient to arrange the test room as a study or work place. In contrast to this, games require much more consideration of the test surroundings (although, naturally in some cases the surroundings are more important than in others, depending on game platform, genre and gamer community).

The evaluation of so-called social or party games is a prominent example for the importance of the test setting. In one of the studies we conducted, a first functional prototype of such a party game should be evaluated. Besides the comprehension of the game’s interface, the test focused especially on questions concerning gameplay experience, such as the specific group dynamics and the fun had by the participants while playing. The test design (amongst other things) included one hour of open, unmoderated playing and a subsequent interview. To allow group dynamics and immersion to occur as naturally as possible the test setting was designed to reproduce a ‘typical Saturday night with friends’. Furniture and the arrangement of the test room matched a common living room. Special attention was paid to the appropriate lighting of the room and to the quality of the sound. The players should experience a ‘party atmosphere’, not an office or a ‘usability lab’ situation. For the same reason the participants were additionally provided with snacks and beverages during their playing. For this test, small groups of friends and families were recruited. We observed that the test groups very quickly to unfold. Only if participants are able to become absorbed in the game for a certain amount of their experiences. An interviewer or the presence of the interviewer interferes with their immersion in the game.

3.2 Data acquisition

Direct observation of behaviour is one of the core methods for user experience testing. Therefore, trained interviewers are a crucial element to ensure the quality of a test. In standard tests the requirements are already high for the interviewer: the observation of gestures, facial expressions and interactions with the product, attentiveness towards verbal expressions of the subjects and their emotional reaction to the test object and the test situation, as well as in-depth knowledge of the test object and the system.

And yet, these demands can increase significantly when games are being tested. Additionally, the complex and rapid interactions of the player with the game interface, multi-player contexts and group dynamics (see above), or other relevant aspects of game experience (e.g. the influence of acoustics and sound) have to be observed and documented. Therefore, the amount of information and data relevant for the evaluation of a game is substantially higher during a test session.

This task is easier when two user experience professionals work together as a moderator and a second observer (similar to test settings employed for complex software products). The increased utilisation of post video analysis can be another solution or also the additional use of adequate tools, which automatically track game interactions. Also, the implementation of a specific notation and structure for written recordings has already proven to be a helpful instrument when used in tests of more complex games or multi-player games.

3.3 Free Interaction

Classic user experience tests usually follow a structured or semi-structured procedure, where phases in which the participant works on a task alternate with phases in which the participant is questioned about his/her experiences. An interviewer accompanies the participants during the different test scenarios as well as during the interaction with the product. In the appropriate moments and situations the interviewer poses questions concerning the relevant aspects of the product and the product-associated experiences.

This semi-structured approach is unsuitable for games testing. Undisturbed engagement with the game is necessary to allow immersion, flow and the other important dimensions of game experience to unfold. Only if participants are able to become absorbed in the game for a certain amount of time without disruption, can psycho-physiological responses related to their game experience be observed, whereas communication with the interviewer or the presence of the interviewer interferes with their immersion in the game.

We therefore recommend a different test procedure be applied in game tests, with extended phases of free play framed by phases of intensive questioning. With certain test objects and study designs, the interviewer can additionally leave the room during
these phases of free play. Thus the participants can devote themselves to the exploration of the game, without any disturbances. Likewise, in short and fast games the participants should be given the opportunity to start and play the game several times before an interview takes place. There lies a significant difference to the use and explore technique which is mostly applied in regular qualitative usability tests (which focus strongly on the first use of a product or service, requiring intuitive and self-descriptive interfaces). The aforementioned party game study is an example demonstrating that the relevant and interesting aspects of game experience can only be identified by allowing extended phases of free and undisturbed gameplay. The form and extent of the free play varies of course, depending on the platform being investigated, the genre and the state of development of the game.

3.4 Thinking Aloud

Thinking aloud is an established technique in usability testing and user experience evaluation. By constantly voicing the (conscious) decision and thought processes, the reasons and rationales of the interaction with the product can be documented and analyzed. What do participants think during the interaction? Why do they choose certain options and ignore others? Why are certain solutions chosen and certain decisions made? While this technique can be widely and easily applied to the functional evaluation of games, it can affect the experience of the test subjects negatively when used to investigate player or game experience. The individually different facets of flow, immersion, excitement and fun can’t emerge if participants are instructed to continuously verbalize their thoughts. The subject would have to continuously split their attention between game and interview situation (one can imagine attempting to watch an exciting movie while simultaneously trying to explain one’s thoughts to someone who is present but uninvolved). Therefore, the thinking aloud method is only partially applicable to the evaluation of non-functional qualities of a game. A suitable alternative can be found in retrospective thinking aloud [14]. In this method the thinking aloud is conducted at the end of a test, when the participant and interviewer analyze a video of the test together. However, while this method allows the interviewer to pose highly experience-specific questions, experience has shown that for many participants it is very difficult to verbalize emotions and experiences after the event. Since retrospectively thinking aloud significantly increases the time spent on one test session, it is recommended (depending on test material and goal of the study) to completely abstain from thinking aloud. Instead of that, spontaneous comments, reactions, gestures and facial expressions, which participants show while playing the game, should be recorded and explored by the interviewer after the game in more detail. The pros and cons of thinking aloud also depend on the game concept. For complex and time-consuming games thinking aloud is not recommended due to the aforementioned reasons. In small, quick casual games, where an immersive game experience is not so much important, or to test predominantly functional game features, thinking aloud is certainly still a practical method.

3.5 Standardized questionnaires

The measurement of subjective satisfaction regarding the usability and user experience of a product is a quasi standard for system evaluations, and standardized questionnaires are frequently used for this purpose. Participants are usually presented with the questionnaire at the end of their interaction with a product. A good number of standardized questionnaires evaluating usability and user experience exist now, to some extent even platform-specific questionnaires (e.g. System Usability Scale [2] or WAMMI [24]). Regardless of whether the questionnaires are self-developed or ‘standard’ questionnaires, in the area of games testing they are only partially suitable, or not at all. While these established instruments can still be used effectively to evaluate the functional properties or features of games, these questionnaires do not have the specific ability to measure game experience. If the holistic gameplay experience has to be examined, even complex user experience questionnaires such as the User Experience Questionnaire [18] or the AttrakDiff2 [13] are not able to adequately cover the game-specific dimensions of user experience, such as excitement, immersion and flow. To assess the specific dimensions experienced in games, one possibility is to customize already existing user experience questionnaires to incorporate questions about the dimensions of game experience. Another possibility is to develop questionnaires individually for certain projects, which is costly, and most of the time not transferable to other projects. However, for the future we expect standardized questionnaires to be available, designed specifically to capture the digital game experience. A promising example is the (so far unpublished) Game Experience Questionnaire by Ijsselsteijn et al. (see [15], [21]).

3.6 Iteration

The iterative evaluation and optimisation of interaction concepts is a cornerstone of user-centred design. Still, in many projects only a few iterations are conducted, or the product is tested only one time. For this kind of singular testing optimisation
recommendations are integrated directly into the conceptual or completed product, without further evaluation loops. This approach can often be justified by the fact that the tested applications are based on conventional interface elements and principles or that they are used already in a common context of use. Thus implementing the optimisations can be guided by extensive experience, knowledge of the users’ needs and best-practices.

In game experience testing, neglecting the iterative procedure can be fatal for playability and game experience. On the one hand, it is impossible for innovative designs and interactions to rely on former products. On the other hand, the number of relevant variables to be considered is higher (comprehension of the interface, game speed, flow, immersion). Thus the number of variables affected by the alterations is higher and the mutual calibration more difficult. In ‘traditional’ usability testing, optimisation normally focuses on fewer dimensions: simple and quick accomplishment of the tasks and consequently a high level of subjective satisfaction. In the area of games, however, even small changes to the interface can destroy a subjectively pleasant flow experience or the gameplay mechanics of a game. Consequently, iterative testing is mandatory.

As a brief illustration we want to share the experience we gained during a study for an online game. The game enabled the player to make fast, sophisticated moves with his game character in a novel manner while at the same time playing tactical manoeuvres, alternating with his opponent. This combination of speed and tactical interaction with the opponent was appraised by the test players as being the highlight of the game and the specific genre. At the same time we observed that the players did not fully understand all the different moves possible in the game. In parts of the game this led to confusion. As the tested game was a casual game, it was one of the requirements that it could be learned within the first few rounds. Consequently, one of the optimisations suggested was to insert additional options (via mouse click) prior to every move in the game to foster a better understanding of the game. In the re-test of this modification we saw that this one additional click took away the feeling of speed and game flow. The actual highlight of the game was destroyed (and an alternative solution to the problem had to be found). Without an iterative process the first optimisation step would have had widespread negative consequences on the game experience.

Based on this example, we recommend a consistent iterative testing and development process to achieve the optimal playability and game experience. Size, number and timing of the evaluation loops can also be adapted to short and fast development processes, e.g. using a Rapid Prototyping method.

4. DISCUSSION AND FUTURE DEVELOPMENTS

In this article we have illustrated several practical issues which we experienced in game experience testing and which we found relevant to share in the discussion related to the adoption of user experience methods for games. We hope that we have been able to show that games are indeed a very specific test object for user experience research, but that this does not mean a fundamentally new research approach is required. We can use our established user experience methods, but we have to adapt them to the specific characteristics of games. We have explained how this adaptation could work for several key aspects of the testing process.

While participant observation and interviews will still constitute the methodological basis of user research for games, promising new technologies for the analysis of game experience should receive more attention in the long run. Technical advancements in sensor technology already allow the use of psycho-physiological methods to objectively measure the cognitive and emotional responses of users. So far the biggest disadvantages of these methods are the confounding nature of the instruments as well as the complicated and lengthy analysis of the collected data. However, the success by companies employing this kind of instruments in the consumer market, such as NeuroSky1, Emotive Systems2 and EmSense3, could mean that the psycho-physiological methods in the future can be applied for a lower price and with fewer disturbances for the participants. And above all, the data could be analyzed and evaluated much faster by providing Application Programming Interfaces (API’s) and analysis routines. Thus these methods could become suitable for application in game testing, as they can effectively supplement observation and questioning methods.

Other interesting questions concerning the applicability of user experience methods and approaches in game testing remain open. For example, the question arises as to how to deal with the issue of insufficient motivation of participants in the test situation. In extreme cases, subjects can be strongly discouraged or frustrated after they have played a game for the first time. This demonstrates the difference between goal-oriented tests (e.g. productivity software) and event-oriented tests (e.g. games). In games, the fun derived from interacting with the game is the main goal. Asking a participant who is already frustrated after a short game session to

try again is of little help. What is plan B in such cases? Accept the test result? Send the participant home and waste the 'expensive' interview and recruitment time? Or keep an alternative questionnaire ready for this case, which explores more general questions concerning the game's concept or the assessment of alternative concepts? Questions also exist regarding the recruitment, another key criterion for the quality of user research studies. Given the many different and relevant behavioural dimensions, is the recruitment of gamers more differentiated and more specific than the recruitment of "standard" test participants? How can all the criteria – game genre, game category, player type, player motivation and motive, player experience – be adequately covered, also taking into account the frequency, duration of play and the respective combinations of the mentioned criteria (If someone plays casual games 5 hours a day, is it appropriate to label him/her as a casual gamer?). Another topic of interest is longitudinal studies: several weeks of field tests to evaluate the long-term motivation and to identify problems outside of the first use context are not uncommon in the area of user research. But their importance is significantly larger in the games sector. In more complex, long-running games the emergence and evaluation of the game flow, of the emotional responses to the game or of the motivation are just some of the central aspects, which would require long-term observation to be analysed. Such longitudinal study designs may include methods such as longer field tests, diaries (including online-diaries), accompanying online forums, online surveys or focus groups. However, which combination of methods will produce the most optimal results in the most time- and cost-efficient way, has yet to be discussed. However, it is evident that a time-limited evaluation in the lab can only provide limited results, of course. Our experience with games so far has taught us that not only the development of games can benefit from user experience research, but also user experience research can learn from games research. To pay more attention to the non-functional qualities of a product is recognized by other consumer markets, too. Therefore, we believe that the knowledge and concepts gained during games research (e.g. flow, immersion, an understanding of the relevance of emotions and aesthetics) can be profitably transferred to the domain of user experience for all kinds of products, while taking into account the playful nature of interaction in general.

5. REFERENCES


www.bitkom.org/de/publikationen/38338_60729.aspx, Last visit: 04.03.2010.