Playability and player experience
Proceedings of the Fun and Games 2010 Workshop

The development of systems for establishing playability and player experience, as well as implementing such measures earlier in the game design cycle can be expected to optimize the game development process as well as the final game design.

This book contains the 7 papers that were selected for the Playability and Player Experience workshop that was held in conjunction with the Fun and Games Conference 2010, hosted by the Centre for User Experience Research of the Katholieke Universiteit Leuven in Belgium. All these papers were positively refereed by an International Programme Committee consisting of 12 qualified experts with a wide variety of backgrounds and expertise domains.

The focus of the Playability and Player Experience workshop was on the development and use of all kinds of playability and player experience measures (like quantitative metrics and data harvesting) to evaluate digital games. Because playtesting metrics are standard practice in larger game development studios, we decided to predominantly focus the workshop on casual games playability measures: playtesting metrics are very much in demand yet not often available within the casual gaming industry.
NHTV Expertise Series

In the context of contributing to the development of knowledge in the subject areas of digital entertainment, hotel and facility, city planning, logistics and mobility, tourism and leisure, NHTV Breda University of Applied Sciences has launched the NHTV Expertise Series. These publications have a logical link with NHTV's strategy and lines of research and they will contribute to stressing NHTV's distinct features as a knowledge institute.

The publications in the NHTV Expertise Series are:

1. 'Coast tourism: a tour d’horizon'
   Research report by Martijn Smeenge

2. 'Education, research and the art of creative thinking'
   PhD thesis by Paul Delnooz

3. 'The global and the local: thinking inclusively about cultures in Breda and the rest of the world'.
   Speech Associate Professorship Cross-cultural Management by Vincent Platenkamp

4. 'Systemic Constellations Work in Organizations'
   Dissertation by Joseph Roevens

5. 'Understanding the behaviour of cultural tourists; towards a classification of Dutch cultural tourists.'
   Dissertation by Rami Isaac

   A pilot-project in cooperation with NBTC-NIPO by Kim de Bruijn, Rob Dirven, Eke Eijgelaar and Paul Peeters

7. Het talent van Brabant. De rol van diverse actoren in talentontwikkeling binnen de sport
   Research report by Theo Hutten

8. Voices in Tourism Development | Creating Spaces for Tacit Knowledge and Innovation
   Conference proceedings edited by R. Isaac, V. Platenkamp and A. Portegies

9. Customer Relationship Management in Hospitality | A theoretical introduction& guidelines for applying the CRM-7-18 model | MSc. Olaf Hermans and Dr. Dan Mount

Colophon
Playability and player experience: Fun and Games 2010 Workshop proceedings

Copyright © 2010, NHTV Breda University of Applied Sciences
All rights reserved. Nothing may be published, or reproduced without prior consent of the authors.

NHTV Breda University of Applied Sciences
Academy for Digital Entertainment
P.O. Box 3917
4800 DX Breda
The Netherlands
T +31 (0) 76 533 2203
F +31 (0) 76 533 2205
www.nhtv.nl
Organizing Committee

Licia Calvi, NHTV Breda University of Applied Sciences (the Netherlands)
Stefano Gualeni, NHTV Breda University of Applied Sciences (the Netherlands)
Koos Nuijten, NHTV Breda University of Applied Sciences (the Netherlands)
Lennart Nacke, University of Saskatchewan (Canada)
Karolien Poels, University of Antwerp (Belgium)

Program Committee

Hans Bouwknegt, NHTV Breda University of Applied Sciences (the Netherlands)
Yvonne de Kort, Eindhoven University of Technology (the Netherlands)
Robbie Grigg, University of Portsmouth (UK)
Wijnand IJsselsteijn, Eindhoven University of Technology (the Netherlands)
Steven Malliet, University of Antwerp (Belgium)
Christof van Nimwegen, K.U.Leuven (Belgium)
Marianna Obrist, University of Salzburg (Austria)

ISBN 9781616278175

© All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or made public in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the author.

This is a NHTV Expertise Series publication
PLAYABILITY AND PLAYER EXPERIENCE

NHTV Expertise Series 10
# Table of Contents

Foreword ......................................................... 7  

Subliminal Advertising in Shooter Games: Recognition Effects of Textual and Pictorial Brand Logo .................................................. 9  

* Koos C.M. Nuijten, Anouk de Regt, Licia Calvi  

“Let’s sweep some Mines together”: Social Interaction & Competition in Casual Games .................................................. 19  

* Brian J. Gajadhar, Henk H. Nap, Yvonne A.W. de Kort & Wijnand A. IJsselsteijn  

Gameplay experience testing with playability and usability surveys – An experimental pilot study .................................................. 31  

* Lennart Nacke, Jonas Schild, Joerg Niesenhaus  

Towards a Serious Game Experience Model: Validation, Extension and Adaptation of the GEQ for Use in an Educational Context ....................... 47  

* Frederik De Grove, Jan Van Looy, Cédric Courtois  

Evaluating Educational Game Experiences in a Classroom Context - Implications for Qualitative Research .................................................. 63  

* Steven Malliet, Niels Quinten, Veerle Van der Sluys  

Critic-Proofing: Robust Validation Through Data-Mining ....................... 81  

* Ian J. Livingston, Lennart E. Nacke, Regan L. Mandryk  

Let’s Start Playing Games! How games can be less about complying and more about playing .................................................. 95  

* Menno Deen, Ben A.M. Schouten  

Notes on Contributors .................................................. 109
The authors that have contributed to this volume all recognize the importance of playability of digital games. In order for a game to engage, move, persuade, or entertain the game player – in other words: to be successful as a game – playability is a key factor. Therefore, on a day to day basis, game developers and game designers aim to deliver work that optimizes playability and game experience.

The enormous production costs of commercial games have made playability testing a regular practice in the development pipeline of larger game studios. However, as Livingston, Nacke, and Mandryk correctly mention in this issue, for casual game development the implementation of playability and game experience measures is somewhat restrained as a result of time and budget limits. For casual games, overcoming these restraints by developing methods for establishing playability and player experience, as well as implementing such measures earlier in the game design cycle, can be expected to optimize the game development process as well as the final game design.

With this argumentation in mind, researchers of the Academy for Digital Entertainment at NHTV Breda University of Applied Sciences (the Netherlands) initiated a workshop within the context of the 2010 Fun and Games conference, hosted by the Centre for User Experience Research of the Katholieke Universiteit Leuven (Belgium). The workshop focused on the development and use of all kinds of playability and player experience measures to evaluate digital games – with an emphasis on casual games.

This volume contains the 7 papers that were incorporated in that workshop after a positive review by the Program Committee. The papers' topics vary from games for fun to games for classroom contexts, from papers that can be qualified as position paper to papers that reflect empirical field phase research, and from papers that take a more fundamental scientific position by developing conceptual models to papers that aim to readily add to the toolkit of playability tests available for game producers. This
diversity reflects a multi-disciplinary field that deserves to develop strongly in the near future. Our aim in initiating our workshop is to add to that development.

We like to thank all authors, the organizing committee, and program committee for their support and participation.

Licia Calvi, Koos Nuijten, Hans Bouwknegt
Subliminal Advertising in Shooter Games: Recognition Effects of Textual and Pictorial Brand Logos

Koos C.M. Nuijten*, Anouk de Regt & Licia Calvi

International Centre for Experimental Media Effects Research at NHTV Breda University of Applied Sciences, the Netherlands

Abstract

An array of studies has shown that games can be suitable for marketing communication – but that not all game genres can be expected to be equally successful in generating advertising effects. Since the game task in shooter games requests players to strictly focus on objects that can pose a threat, that game genre seems especially problematic for in-game advertising. Subliminal communication can theoretically be expected to overcome (part) of that problem. In order to test this, we designed an experimental study (N = 143) which focuses on marketing communication effects of subliminally presented brand logos in shooter games. Our aim was to find out (a) whether subliminal marketing communication causes recognition effects and (b) whether pictorial logos differ from textual logos in the size of the effect they generate. This way we can shine a light on the message processing mechanisms that are foundational to subliminal message effects.

The results of our study show that subliminally presented brand logos do have a communicative potential: the recognition effects are significant. Besides, our study indicates that pictorial logos have a greater propensity to subliminally communicate than textual logos.

Introduction

Over the last decade, quite an extensive number of studies focused on the effectiveness of in-game advertising (Acar, 2007; Chaney et al., 2004; Grigorovivi
and Constantin, 2004; Kuhn et al., 2007; Lee and Faber, 2007; Mau et al., 2008; Nelson, 2002; Sharma et al., 2007), showing that games can be suitable for marketing communication. To what extent in-game advertisement effects are reached seems to be dependent on the interaction between brand and game player – i.e., mere brand presence is less effective than brand interaction (Schneider and Cornwell, 2005), the telepresence a game player experiences – i.e., higher telepresence leads to stronger effects (Nelson et al., 2006), and the graphic environment that the game is set in – i.e., environments that resemble the genuine world support advertisement effects better than environments that do not resemble the genuine world (Nuijten et al., 2010).

From these results, we may infer that some game genres would be more usable for in-game advertising than others: not all games aim for the highest telepresence possible; some game attractiveness is quite inherent to an estranged game setting; moreover, the effectiveness of in-game advertisement is likely to be dependent on the type of task that the game evolves around (e.g., navigating, racing, shooting and surviving). Experimental studies concerning these issues are problematic, mainly because creating credible games that manipulate telepresence, game setting, and game task usually entails a series of changes in related variables (like game environment, game difficulty, game likeability) that would hamper any causal conclusion.

Especially the genre of shooter games (e.g., Modern Warfare, Battlefield, Medal of Honor) can be expected to be problematic for in-game advertising. Although shooter games can lead to high telepresence (Farrar et al., 2006), the game setting (i.e., battlefields, war zones, demolished cities) is generally estranged. Moreover, because the game task in shooter games usually requests the game player to focus strictly on objects or agents that can pose a threat, in-game advertisement or product placement in this game genre is faced with a dilemma: either the advertisement does not pose such a threat – but then it is likely to be actively ignored, or it does pose a threat – but then it is likely to have unwanted brand attitude effects.

Our study tries to overcome this dilemma for shooter games in-game advertising by exploring an advertising form that is not commonly used. Previous in-game advertising effectiveness studies predominantly tested commercial messages that were embedded in virtual contexts in much the way such messages could also be
embedded in genuine environments. Race track games advertising studies (Chaney et al., 2004; Nuijten et al., 2010; Schneider and Cornwell, 2005), for example, contain advertisements next to the track, on safety fences and crash barriers, much like race tracks in the real world do. Instead of context inherent in-game advertising, our study focuses on an advertising form that depends on the media platform: subliminal advertising.

Subliminal communication consists of messages that are presented for just a fraction of a second: just short enough to not be consciously detected by the media user but still potentially long enough to lead to advertising effects. Ever since the late 1950s, subliminal communication has been issue to (predominantly fiercely negative) public debate. This negativity is related to the covert character of subliminal communication: it is considered sneaky and therefore reprehensible. Much of the debate assumes that subliminal communication is over powerful because media users that do not know they are being communicated to, cannot employ defence mechanisms (like cognitive counter arguing) to ‘neutralize’ persuasive effects. However, subliminal effects research, from Adams (1957) to Lodge and Taber (2005), notoriously shows indecisive or mixed results, at least when mediated political or commercial communication is concerned. Subliminal persuasive communication is not by far as powerful as the public debate assumes.

We do not aim for an exhaustive or conclusive answer concerning the ethical acceptability of commercial subliminal communication in video games: ethical decisions are not within the realm of empirical science. Our research aims to establish and understand subliminal message effects in order to support an ethical debate that is grounded in facts rather than in fear.

Theory and hypotheses

According to Lang (2000), media messages can be conceived of as an ongoing stream of auditory and visual stimuli. To make sense of media messages, within the head of the media user a continuous process of coding (decoding and encoding), retrieval, and storage takes place. But due to the fact that the media user’s capacity to simultaneously run these processes has a ceiling, the human information processing system is
equipped with stimuli selection mechanisms. These mechanisms consist of automatic and intentional attention processes that select certain stimuli for processing and filter others. According to Lang (2000), the system is prone to support processing of evolutionary important or task-relevant stimuli. This explains why product placement or embedding brand logos in a virtual environment in a shooter game might not really be successful: the commercial messages are not part of the task-relevant stimuli, nor do they trigger evolutionary selection mechanisms. The result is that these messages are likely to become immediate victims of our processing ceiling: not selected for processing – game over.

Subliminal messages may be able to avoid immediate filtering, however. When a stream of stimuli is shortly, abruptly, and unexpectedly interrupted, the human information processing system autonomously reacts with an orienting response – OR (Lynn, 1966). The OR is an evolutionary mechanism associated with attention that can also be described as the ‘what is this?’-reaction. Stimuli that elicit an OR therefore can be expected to have a communication potential because they automatically attract attention, even if they are not task-relevant. Subliminal messages by definition only allow for low elaboration: higher level cognitive processing (like reasoning, counterbalancing arguments) is not possible if a media user is not consciously aware of the message. Therefore, we may only expect quite basic subliminal advertising effects – recognition rather than recall, let alone brand attitude change. Our experimental hypothesis is then:

| H1: Logos subliminally presented in a shooter game lead to recognition effects. |

The implications of this hypothesis may seem a bit limited. If it is only capable of causing recognition effects, subliminal marketing communication does not seem to be a very powerful tool. To what extent longitudinally repeated and extensive exposure to subliminal communication enhances the effects is not clear, however. Empirical research usually does not allow for subjects to play for hours a day, weeks in a row – whereas in the everyday environment that is exactly what a lot of shooter game players do. In other words, while empirical research may be able to only establish recognition
effects, the effects in the natural context may be stronger (e.g., in the realm of recall effects).

Furthermore, empirical clarity is lacking concerning the link between lower level media effects (like recognition, or recall) and higher level media effects (like brand attitude, or buying intention). Priming theory argues that short term impact of media exposure could have effects on subsequent judgments or behaviours (e.g., Roskos-Ewoldsen et al., 2009). Subliminal advertisements might be able to trigger associative pathways related to our network of memories and thus lead to higher level advertising effects, especially in situations of high exposure (such as with shooter games in the natural setting), even if experimental empirical research is not able to establish such long term effects.

In all, our first hypothesis can be considered to forecast a minimum requested effect: if our study is not able to establish immediate recognition effects, we may assume that higher level effects, even in situations of high exposure, are unlikely. In case our first hypothesis is rejected, the marketing communication potential of subliminal messages is next to none.

Additionally, we can expect a difference within different types of marketing communication. Pictorial marketing communication (pictures, drawings, figures) may be symbolic, but it demands less coding than textual communication (letters, words, sentences), that requires the abstract-symbolic and cognitive task of reading. Therefore, the effects of pictorial subliminal marketing communication can be expected to be bigger than the effects of text-based subliminal marketing communication. We hypothesize:

H2: Pictorial logos subliminally presented in a shooter game will have higher recognition effects than textual logos that are presented under the same conditions.

Method

For our study, we used a third person shooter game that was especially programmed and designed for our purposes. In the game, a pirate ship orbits the galaxy and is attacked by various flying objects, some of which just pose the threat of collision, while others actively fire on the pirate ship. The game was designed to both
graphically and auditory present each impact (collision or hitting gunshot). Game features entailed playing time in seconds, an indicator of damage-to-game-over and a high-score list.

The game was programmed to generate 60 still frames per second to simulate movement. When the game timer hit 30 seconds, at a pace of an average of 6 per minute, the computer started replacing single game frames (1/60th of a second) by frames that contained a total of three brand logos (in random order). In order to test our second hypothesis, the three brand logos were selected to be fully pictorial (Shell), partly pictorial and partly textual (Heinz), or fully textual (MTV). To rule out bias from logo colors and to make sure all logos would equally stand out, the graphic environment all logos were presented in different shades of blue.

Since our study aimed at establishing recognition effects of subliminal messages embedded in shooter games, and at finding individual differences in such effects, we tried to rule out as much ‘noise variance’ (originating, for example, from computer experience – and therefore sense of telepresence or involvement) as possible by achieving a homogeneous sample of respondents. Our sample consisted of 143 respondents (50-50 gender division) in the age range between 17 and 28 (mean age 21), all undergraduate students of a media and entertainment educational program at our (international) university.

All respondents were invited to participate in a memory task where they were asked to memorize a series of brand logos. For this memory task, we created a black and white logo overview containing brand logos of well-known multinational companies. All logos (some more textual, other more pictorial) were presented with about the same size. Besides, we made sure that the companies that these logos represented fitted in an array of market branches (automotive, fast food, FMCGs, media companies, nonprofit organizations) in such a way that mnemonic chunking could not easily be employed. The overview contained 26 logos, but due to a cloud-form presentation, the amount of logos was not to be easily established.

Each respondent was informed about the test protocol: 45 seconds to memorize the logos, then playing our computer game as distraction task, followed by a logo recognition task. All respondents consented to our study without incentives. The mean playing time for the distraction task was about two minutes.
For the logo recognition task, we presented each respondent with a black and white list of 49 brand logos (all well known multinational companies within the same market branches, all neatly organized in three columns). We told our respondents that the logo cloud they tried to memorize at the beginning of the experiment contained 29 logos and we requested each respondent to tick all 29 – even if that would entail some guessing. After that, we asked them whether they noticed something strange with the game in order to check whether the subliminal messages were consciously noticed. Less than five percent of our respondents claimed having seen anything peculiar, none of the respondents reported having seen flashes of brand logos. At the end of the experiment, all respondents were fully informed about the purpose of the study.

Results

Our first hypothesis concerned undifferentiated subliminal logo presentation effects. To test this hypothesis, we calculated the mean recognition rate for the three subliminally presented logos and for the logos that were not presented in the logo cloud nor in the game. Then, we used a t-test to establish whether the mean recognition rate for the subliminally presented logos was significantly different from the mean score for the non-presented logos. Our results show that the subliminally presented logos have a significantly higher recognition rate than the non-presented logos. And although the recognition rate of the subliminally presented logos is not as high as the rate for the logos that were presented as part of the memory task (the logo cloud), our first hypothesis is supported.

Our second hypothesis concerned differences in effect size depending on the pictorial or textual nature of the logo. As described in the method section, in the game we subliminally incorporated a picture based logo (Shell), a picture-and text logo (Heinz), and a text logo (MTV). Consequently, according to our second hypotheses the recognition effects for the Shell logo should be highest; the Heinz logo recognition effects should be somewhat in the middle; and the recognition effects of the MTV logo should be the lowest. In order to test this second hypothesis, we respectively employed t-tests to compare the mean recognition rate for the Shell and the MTV logo, the Shell and the Heinz logo, and the Heinz and the MTV logo. Our results show
that the mean recognition rate for the Shell logo is the highest, i.e., significantly higher than the mean recognition rate of both the Heinz logo and the MTV logo. Moreover, the mean recognition rate for the Heinz logo is significantly higher than for the MTV logo. In all, the pictorial logo has the highest mean recognition rate, the combined textual-pictorial logo has a mid-score, and the full textual logo has the lowest mean recognition rate. Our second hypothesis is therefore supported by our data.

Conclusion and discussion

We have performed a study showing that subliminal communication is not a powerless communication form. We have argued that the impact of subliminal communication can be expected to be limited, although the marketing effects of subliminal communication could be bigger with high exposure and in case ‘traditional’ in-game advertising is combined with subliminal messaging.

The most important conclusion of our study is that messages that are designed to go unnoticed still generate marketing communication effects. Our results indeed indicate the lower limit of the potential of subliminal communication, i.e., recognition. The occurrence of recognition effects is a prerequisite to higher level communication effects: if a message does not cause (cued) recognition that by and large means that the message did not enter the message processing system at all (Lang, 2000). However, our study shows that they do, and that means that higher level impact at least may occur.

When considering the media exposure time, the results of our study were found after a mean game playing time of approximately two minutes. Since, in a natural setting, shooter games are typically high exposure media, in a natural setting the communication effects of subliminal messages will be higher than in our study.

The second conclusion of our study is that pictorial logos have a stronger propensity to communicate in subliminal messages than textual logos. This means that subliminal message processing is a task that is highly sensitive to coding (decoding and encoding) demand. Our results confirm this by showing that stimuli that demand a more complex coding task (textual coding/reading) generate less powerful results than stimuli that require less complex coding tasks (pictorial logos).
References


“Let’s sweep some Mines together”: Social Interaction & Competition in Casual Games

Brian J. Gajadhar* & Henk H. Nap,
Yvonne A.W. de Kort, & Wijnand A. IJsselsteijn
Game Experience Lab, Human-Technology Interaction Group, Eindhoven University of Technology, the Netherlands

Abstract

Casual games distinguish themselves from other digital games in their easy access, simple rules, minimal time commitment and simple interface. These games are known for their quick and short rounds, which make them ideal to play in mini-breaks between other activities. Recently, many social networking sites have started offering opportunities to chat and play multiplayer casual games with other users. Studies on digital games have shown that social characteristics of play settings have a strong impact on players’ in-game experience. As little attention has been given to casual games in literature, in this paper we discuss the important role of social factors in understanding and measuring player experience in casual digital games.

Introduction

Casual games are digital games developed for the mass consumer, even those who do not regard themselves as “gamers” (Casual Games Association, 2007; Kuittinen, Kultima, Niemelä & Paavilainen, 2007; Li & Counts, 2007). These games distinguish themselves from other (complex) games by their easy access, simple rules, and straightforward interface (see Figure 1). They are known for their quick and short rounds, which make them ideal to play in mini-breaks during work, chores at home, or school time. In spite of these short rounds, many players play casual games for substantial amounts of time in a row.

*b.j.gajadhar@tue.nl
Since the introduction of the Internet, all sorts of games can be accessed easily on one’s PC. As they are available in almost any genre topic (e.g., puzzle, strategy), casual games are played by all sorts of players, regardless of their age, gender or nationality (Casual games Association, 2007). Yet, the typical casual gamer is a woman who is in her early 40s and likes to play when she has the house to herself. Her favorite game genre is puzzle games, followed closely by card games (Partridge, 2007). Findings indicate that casual games are mostly played for stress relief and taking a break (Casual Games Association, 2007; Kuittinen et al., 2007; Li & Counts, 2007; Bogost, 2004).

In digital gaming literature, relatively little attention is paid to casual games compared to studies on more complex games (Kuittinen et al., 2007). Studies regarding the motivational pull and the experience of digital play have more often used popular games, such as Half Life 2 (van den Hoogen, IJsselsteijn & de Kort, 2008), Super Monkey Ball Jr. (Ravaja et al., 2006), and Neverwinter Nights (Weibel et al., 2008). Developed play-testing metrics have proven useful for these types of games, but have not been used extensively for casual games. As player experience evaluation in (casual) game design cycles becomes more important (Martins et al., 2010), adapting advanced methods for measuring players’ feelings and emotions in playing casual games is highly recommended. Despite the successful progress in developing useful measurement tools, one factor has reached insufficient attention when players’ experiences are measured. Especially for the field of casual gaming.
which recently has suffered an extension in available play styles (Bogost, 2004), the social factor becomes more relevant.

At first sight, social context might seem to play an insignificant role in casual games, since these games are mostly played in solo-settings to pass time and for stress relief (Casual Games Association, 2007; Kuittinen et al., 2007). However, casual games recently have gained popularity on social media such as Facebook, Hyves and MSN Messenger, where they can be played with other players (USA Today, 2010). As a result, existing reasons to engage in playing casual games – i.e., pass time and stress relief – may have been extended with new motivations such as social interaction and competition. In line with research in other multiplayer games (Gajadhar, de Kort, IJsselsteijn & Poels, 2009c), these new factors may also have an impact on players’ in-game experience.

(A)synchronous multi play
Multiplayer casual games can be played in two ways: synchronously and asynchronously. An example of an asynchronous multiplayer casual game is Mob Wars hosted on the social networking website Facebook (see Figure 2). This game is played with others in sequence, at different moments in time (Bogost, 2004). In Mob Wars, players start out as a thief with the goal to get higher in rank, up to the level of Mafia godfather. The

Figure 2. Print screen of the casual game Mob Wars on the social networking site Facebook.
The rise of available multiplayer games may have an effect on the motivations for people to engage in casual game play. In addition to passing time and finding relief...
from stress, the social factor – i.e., social interaction and competition – may become important. Recent studies on complex games have shown that social context has a great (and positive) impact on the motivations to engage in digital gaming and on players’ in-game experiences.

**Social interaction & competition**

Many studies reported that people enjoy playing digital games together or watching others play, as players can demonstrate their skills and enjoy the feedback of enthusiastic bystanders (de Kort & IJsselsteijn, 2008). A recent field study (Gajadhar et al., 2009c) revealed that playing together online or in co-located settings provides gamers the feeling of inclusiveness and belonging to a group, i.e. social connectedness. This social function is seen as the key factor that affects the motivation for a player’s choice of co-play setting. In line with (Jansz & Martens, 2005), the study furthermore revealed that social interaction and (social) competition are among the main motivations to engage in digital gaming.

A subsequent study (Gajadhar, de Kort & IJsselsteijn, 2008a; 2009a) was conducted to empirically testify the role of social interaction and competition on player experience. An experiment was performed where people played a digital game in three types of co-play configurations: virtual, mediated, and co-located (Gajadhar, de Kort & IJsselsteijn, 2008b). The Game Experience Questionnaire (IJsselsteijn, de Kort & Poels, in preparation) was applied in the study and results indicated that playing side-by-side significantly adds to fun, challenge, perceived competence, flow, boredom and immersion in the game as compared to playing against a distant or virtual opponent. For most components the effect of social context was mediated by the level of social presence (the feeling of being together, although physically apart; Biocca et al., 2003), which was measured with the Social Presence in Gaming Questionnaire (SPGQ) (de Kort, IJsselsteijn & Poels, 2007). Furthermore, player experience in terms of positive affect, competence, challenge, frustration and flow was significantly influenced by players’ in-game scores.

A subsequent study (Gajadhar, de Kort & IJsselsteijn, 2009b) investigated why co-located co-play was more positively experienced than mediated play. Therefore the influence of additional social communication channels – such as webcams and
headsets – on player experience and social presence was tested. The data revealed that player experience components were significantly influenced by the availability of social cues, especially by talking and laughing. Analyses revealed that the level of social presence depended on the availability of audio cues in digital game settings. Again, in-game achievements appeared to be highly important for players’ in-game experience.

These findings illustrate that social context is an important determinant of player experience in complex games; especially when there is room for conversation. By applying the SPGQ in all studies, differences in most player experience components were explained by the feelings of social presence. Interestingly, many player experience components in both studies revealed highly significant effects on player performance. The latter proves that – next to the opportunity for social interaction – competition is a major factor that influences the player experience.

Age, gender, familiarity
The importance of social interaction and competition has been investigated not only for the stereotypical male adolescent gamer, but also for senior gamers and for female players (e.g., Gajadhar, Nap, de Kort & IJsselsteijn, in press). Percentage-wise, senior gamers (50+) play casual games far more frequently than younger players; and multiplayer casual games have also found their way to the 50+ demographic (Casual Games Association, 2007). Yet, the older segment of the population perceives and experiences multiplayer gaming differently than younger players. A recent explorative study (Nap, de Kort & IJsselsteijn, 2009) showed that in contrast to young adults (Gajadhar et al., 2009c), senior gamers have negative perceptions about social play over the Internet and prefer to play single or in co-located settings. These findings are supported by an experimental study (Gajadhar et al., in press) in which it was found that senior gamers – in contrast to young adults – experienced online co-play as least enjoyable compared to solo play or co-located co-play. Moreover, seniors’ sense of social presence did not increase from solo play to online play; in other words, they felt as alone playing against a computer as playing against a distant, mediated other. The authors concluded that to maximize player enjoyment for seniors, less focus has to be on social competition and opportunities for social interaction should be enabled.
Demographic data also show gender differences in casual game play and how people experience social play. Female gamers report higher liking of casual games (Lucas & Sherry, 2004) compared to male gamers and are typically overrepresented in casual gaming and multiplayer casual gaming (Casual Games Association, 2007). However, motivations to play multiplayer casual games differ between both sexes. A recent study (Vanden Abeele, Gajadhar & Schutter, 2009) quantitatively and qualitatively revealed that females particularly enjoy the social interaction in digital games, while men enjoy the social competition. Similar to seniors, multiplayer casual games for females will therefore be more enjoyable when a game’s focus is less on competition and more on social interaction.

The aforementioned studies reveal that – in line with a framework presented by de Kort & IJsselsteijn (de Kort & IJsselsteijn, 2008) – social characteristics of play settings are highly important for players’ in-game experience. In particular due to opportunities for social interaction and presence of competition, digital games can be experienced differently by people.

Discussion

Casual games have been played mostly to pass time and to relieve stress (Casual Games Association, 2007; Kuittinen et al., 2007). These games can be played satisfactorily in a mini-break at school or at work, and have more often been played in solo settings. Recently, social media offer the opportunity to play multiplayer casual games (USA Today, 2010). These multiplayer casual games can be played asynchronous – i.e. in sequence over a substantial amount of time (Bogost, 2004) – and synchronous at the same moment in time while chatting. Due to the possibilities of playing against others, “social interaction” and “social competition” are factors that need to be accounted for in player experience studies of multiplayer casual games.

Since casual games are most popular among seniors and females, who in general dislike competitive digital games (Nap et al., 2009; Gajadhar et al., in press; Vanden Abeele et al., 2009), it is unclear whether social factors will also become important motivators; as is the case for other types of games (Gajadhar et al., 2009c). Moreover, results of studies on complex games have revealed that social context is
an important determinant of player experience (Ravaja et al., 2006; Weibel et al., 2008; Gajadhar et al., 2008b; 2009a; 2009b). In addition, studies on interpersonal differences – in terms of age, gender and familiarity – showed the importance of social interaction and social competition in how players experience digital gaming (Gajadhar et al., in press; Vanden Abeele et al., 2009; Gajadhar et al., in preparation). Since social elements have gained popularity in all type of games, these findings may also hold for the player experience in casual games.

Therefore, we emphasize to control for social context effects in measuring player experience in casual games, as has been done in recent studies on complex games (e.g., Gajadhar et al., 2008a; 2009a; 2009b). To do so, we recommend future experimental designs to include the Social Presence in Gaming Questionnaire (de Kort et al., 2007) to control for effects of social interaction between players. These outcomes should be used in the data analyses regardless of the way player experience has been measured; e.g., observations (Gajadhar et al., in preparation), questionnaires (e.g., Gajadhar et al., 2008a; 2009a; 2009b), physiological measurements (e.g., Ravaja et al., 2006). Furthermore, to take account of the influence of players’ in-game achievements on player experience, we put forward to also include players’ performances – i.e., in-game scores, achievements, winning vs. losing – in data analyses as has been done in recent related studies (e.g., Gajadhar et al., 2008a; 2009a; 2009b). In sum, these regulations will give more insight in the understanding of player experience in casual games.

Conclusion

The importance of social context effects has already been testified in player experience measurement in digital play with complex games. Social interaction and competition are factors that have a decisive impact on players’ feelings and emotions while playing digital games. Since casual games currently are more frequently played with other players, also social effects have to be considered in player experience metrics for this type of games. These measures will enhance current methods for measuring player experience in casual games, which will result in a more valid and comprehensive understanding of the casual game play experience.
Acknowledgments

Support from the European Games@Large project is gratefully acknowledged.

References


Vanden Abeele, V., Gajadhar, B.J., & de Schutter, B. (2009). ‘Gaming naturally is more fun together: The influence of controller type on player experience.’ International

Gameplay experience testing with playability and usability surveys: An experimental pilot study

Lennart Nacke*
Department of Computer Science
University of Saskatchewan
Canada

Jonas Schild, Joerg Niesenhaus
Department of Computer Science and Applied Cognitive Sciences
University of Duisburg-Essen
Germany

Abstract

This pilot study investigates an experimental methodology for gathering data to create correlations between experiential factors measured by a gameplay experience questionnaire and player quality measures, such as playing frequency, choice of game, and playing time. The characteristics of two distinct games were examined concerning the aspects of game experience, subjective game quality, and game usability. Interactions within the three aspects were identified. The results suggest that gameplay experience dimensions flow and immersion are similarly motivating in different game genres, which however might not be equally enjoyable. On the one hand, usability ratings may be positively influenced when a game provides immersion and flow or on the other hand, flow and immersion may be negatively influenced by poor usability ratings. These results emphasize the need for an approach to classify games based on correlation patterns involving game experience, quality, and usability.

Introduction

Prior studies of digital games have often focused on the negative effects of digital gaming, such as violent content and its impact (Carnagey, Anderson, & Bushman,

* Lennart.Nacke@acm.org
2007; Gentile & Stone, 2005) or addiction to playing (Grüsser, Thalemann, & Griffiths, 2007). However, there has been a recent focus on trying to understand aspects central to gameplay experience (Nacke, 2009b; Poels, de Kort, & IJsselsteijn, 2007). For example, Fernandez (2008) proposed a gameplay experience model, which focused on temporal influences before, during, and after gameplay experience in player-game interaction. Fun is the main component of player experience in this model. It further proposes that game evaluation should concentrate on emotional and cognitive player reactions. IJsselsteijn, Poels, and de Kort theorized that immersion, tension, competence, flow, negative affect, positive affect, and challenge are important elements of gameplay experience and developed a game experience questionnaire (GEQ) to assess these elements (IJsselsteijn, Poels, & de Kort, 2008). It is the goal of this study to investigate correlations between the experiential factors measured by the GEQ and player quality measures such as playing frequency, choice of game, and playing time.

This test also requires investigation of whether the underlying usability of a game implementation has an influence on gameplay experience. As Nacke (2009a) notes, usability research on the other hand has taken the ISO 9241-11 standard, defining usability as “effectiveness, efficiency, and satisfaction in a specified context of use” (ISO/IEC 9241-11, 1998). Relating more directly to game developers, Sánchez, Zea, & Gutiérrez (2009) tried to map usability to playability for evaluating UX in games by deconstructing playability and integrating methodological considerations from game development practice. Following this argumentation, one could see digital games merely as software with the same interaction requirements as other products. However, the interactive experience in games is focused on the progression inside the game system rather than the outcome achieved by playing the game. This is one essential difference in interaction design for digital games and software tools: You play games for the experience itself, thus you have creative freedom in designing the experience itself, while in software tools you are trying to design a pleasant way of achieving a goal efficiently. Game development practice must account for this interaction design aspect. If we see experiential factors like flow, immersion, and enjoyment as constructs of game experience, then these can be facilitated by:
1. Choices in gameplay design – which is essentially the social, psychological and cognitive construction of an enjoyable, interactive, goal-driven experience – or
2. Underlying technical prerequisites for this interactive experience to unfold – this relates to the usability of the technology, interface, and interaction devices, which facilitate gameplay.

While in general, both of these factors contribute to overall game usability, here we use the term game usability to refer to factor 2: technology, interface, and interaction. We also assume for our pilot study that we can measure game usability of games using a modified system usability scale (SUS) (Brooke, 1996) – which will be introduced in the methods section below – as well as measures like playing time, playing frequency, and game quality evaluation. The idea of using playing time and frequency as additional measures for this study came from the discussion of usability metrics in Seffah, Donyae, Kline, & Padda (2006), where behavior over time was discussed as a usability metric. For games, it is especially interesting to look at frequencies and play-session times, because these metrics could indicate a preference of a certain game, just in the same way as one would prefer software tools that take the least amount of time for achieving certain tasks (Seffah, et al., 2006). However, game preference may also come from aesthetic factors that could enhance how people perceive the quality of a game, which is why we also chose to measure the quality of gameplay with a separate questionnaire item. In summary, we hope to gain a richer picture of the experiences and preferences evoked by gameplay with this combination of usability metrics, game quality evaluations, and subjective experience assessments. The main contribution of this paper is therefore more in the discussion of its methodological approach than in its initial results presented.

Our approach is the examination of game experience ratings’ impact on quality measures such as rated quality, playing time and playing frequency. More specifically, we used the following research questions as a basis for our hypotheses: “Do immersion and flow influence play behavior? What is the effect of usability aspects on these quality measures?” Finally: “How do such aspects differ for games in different genres with different interaction and play styles?” In order
to investigate these research questions, we formulate the following experimental hypotheses:

\[ \text{H1: The two games from different genres invoke a different gameplay experience measured by the GEQ.} \]
\[ \text{H2: Game quality correlates with game usability as measured by the SUS.} \]
\[ \text{H3: A high game quality correlates with longer playing time and/or frequency.} \]

**Material and Methods**

*Participants*

Participants were 12 right-handed Swedish young adults (between 20 and 33 years old). On average, participants played video games 17 hours and 30 minutes per week \((M = 17.5, SD = 21.62)\). Of the total, one third \((N = 4)\) were female and two thirds \((N = 8)\) were male. On average they have played digital games for 16.18 years \((\text{Min} = 3, \text{Max} = 27, SD = 6.56)\) in their life. Only one third of the participants preferred to play games in multiplayer mode (MMOG1, Local Multiplayer or Clan) in contrast to two thirds who favored single player mode (either alone or with other people in the room). The most popular genres were adventure games (including action-adventures) \((33.3\%)\) and role-playing games \((33.3\%)\). One participant decided to abort the experiment after the initial session and was excluded from analysis.

*Games Used in the Study*

In this experiment, we used two games of similar style and quality, but from different genres and using different interfaces. The games were chosen to be equally pleasurable for advanced and novice players, without factoring in the date the games were developed or released. One game was a remake of a classic commercial game, while the other was a downloadable commercial game:

- Maniac Mansion Deluxe [MMD] (originally released in 1987 by Lucasfilm Games) (LucasFan Games, 2004) – Adventure game, see Figure 1.
- Zuma [Zuma] (PopCap Games, 2003) – an Action Puzzle game, see Figure 2.

---

1 *MMOG* is short for Massively Multiplayer Online Game.
Figure 1. Screenshot from the adventure game Maniac Mansion Deluxe [MMD] (LucasFan Games, 2004)

Figure 2. Screenshot from the action puzzle game Zuma [Zuma] (PopCap Games, 2003)
The games were chosen with the two genres being very different from each other so that there should be a clear preference for one of them for each player. Both games are in a similar comic style, non-realistic and simplistic in their interface. While one is more complex in its narrative, the other focuses on the timely solution of puzzle challenges, thus challenging the player in a different way, which could lead to different game experiences. Concerning the usability, the different forms of interaction in both games lead to the question if the usability criteria of the SUS apply well for both games. Maniac Mansion Deluxe uses a command system to control the game, which is shown in a separate graphical user interface (GUI) window. This type of GUI becomes outdated as many current games, like Zuma, use game elements, graphical and acoustic feedback mechanisms within the game to give feedback to the user. Despite the different interfaces, both games have received excellent reviews indicating an equal quality of the very different gameplay mechanics.

For example, Maniac Mansion Deluxe was hailed by Computer Gaming World as “a clever and imaginative game” (Ardai, 1988), receiving an average reader rating at IGN.com of 9.6\(^{\text{II}}\), and Zuma has a Metacritic\(^{\text{III}}\) score of 77% and an average reader score of 9.2.

Methods
We used several different questionnaires for assessing tendencies of game players to experience flow or immersion during gameplay. For the initial assessment of game experience (when playing the games in the laboratory) we used the GEQ. These results were later compared with logged playing time and frequency data and a game quality and game usability assessment completed by each player for each of the games. After the three-week experimental period, an assessment of game quality was made using our own questions and questions from the SUS. The scale consists of ten items, which are then used to derive a total usability score in a range from 0 to 100. We altered these questions to account for game systems instead of general systems, by replacing

\(^{\text{II}}\) IGN.com is an online magazine focused on games, see also http://pc.ign.com/objects/006/006749.html

\(^{\text{III}}\) Metacritic combines review scores from a carefully-screened group of well-respected critics into an overall grade. Its game section is available at http://www.metacritic.com/games
the word “system” with “game system” and the word “use” with “play.” Regarding
the assessment of game quality, we asked the participants to rate overall quality
of each game on a scale from 1 (worst quality) to 10 (excellent quality). In addition,
participants were asked to rank their games (1st and 2nd place) according to their
playing preference.

Procedure
We recruited participants from a gaming background, also from within a Master’s
program in Digital Game Design. An initial assessment questionnaire was sent out to
all participants willing to take part in the experiment. All participants were invited
to a game laboratory. After a brief description of the experimental procedure, each
participant filled in a compulsory “informed consent” form (with a request not to take
part in the experiment if suffering from epileptic seizures or game addiction). Each
participant had to complete an initial demographic and psychographic assessment
questionnaire prior to the experiment, which was checked for completion. The first
phase of the experiment consisted of a game session in the laboratory. The participants
played each of the games for 15 minutes with a laptop computer. After each game,
participants reported their game experience (using the GEQ). The games were played
in a counterbalanced order. Then, the functionality of the game logging software was
explained to each participant individually and the software and games were handed
out via direct transfer, USB-stick or secure download. Each participant was asked to
play the two stimulus games, emphasizing that they can freely play the games over a
period of three weeks (i.e. freely choose which game to play for any play session, for
how long, and how many play sessions they would like to play), with the exception
that each of the two games should be tried at least once. It was also explained that it
is preferable to play only those games contained in the study during the three weeks.
Each participant was thanked for taking part and escorted out of the lab. In the second
phase of the study, the participants played the two games at home. Their game-playing
behavior (game selections, playing time) was recorded using simple, custom-made
software that is used to launch the games and to record starting times, ending times
and durations to a log file. After the three-week experiment period, participants were
allowed to keep the games and asked to send back the log file by email.
Data Reduction

Data entries showing two minutes or less of playing time before a longer playing session were deleted from the logs. This was done after discussing the playing experiences with the participants and after many of them indicated that they had trouble starting the games through the logging software for the first time. It was checked whether the data collected was distributed normally using the Kolmogorov-Smirnov test. This was important in order to evaluate whether non-parametric methods had to be used for the analysis if the distribution was not normal.

Results

Game experience questionnaire (GEQ) results after initial play session

The data were parametric except for the component tension under the MMD condition, D (11) = 0.36, p < .0001. A dependent t-test showed significant differences between Zuma and MMD for flow (t(10) = -2.23, p < .05), positive affect (t(10) = -2.67, p < .05), competence (t(10) = -3.96, p < .01), and immersion (t(10) = 2.68, p < .05). Differences for challenge and negative affect were insignificant. Significance for tension was checked using a Wilcoxon signed-rank test and no significant differences were found between the games. As expected, both games scored low in tension (MMD: M = 2.4, SD = 1.06; Zuma: M = 2.2, SD = 0.89) and negative affect (MMD: M = 1.8, SD = 1.04; Zuma: M = 1.6, SD = 0.80), suggesting that both titles were equally suited for longer gameplay with this participant group. This could indicate that MMD would be more suited to immersive gameplay than Zuma, which would be more suited to flow gameplay, MMD scored significantly higher on immersion (M = 2.34, SD = 0.70) than Zuma (M = 1.41, SD = 0.79). In addition, Zuma scored significantly higher on flow (M = 2.55, SD = 0.77) than MMD (M = 1.71, SD = 0.86). Interestingly, this does not affect the ratings on challenge, which have almost equal average values: for Zuma, M = 1.76, SD = 0.81, and for MMD, M = 1.73, SD = 1.0. In addition, we have to mention that there was a strong positive correlation between flow and immersion scores, r = .62, p < .05, potentially suggesting a mutual influence of these concepts for the games that were measured in this study. For Zuma, we also found that the ratings for positive affect (M = 3.12, SD = 0.90) and competence (M = 2.82, SD = 1.01) were much higher than for MMD (PA: M = 2.06,
39

SD = 0.85; C: M = 1.18, SD = 0.78). Figure 3 shows a comparison of all game experience scores for each game. In summary, these differences support our first hypothesis (H1) that the two different games lead to distinct game experiences.

Playing time and frequencies during a three-week period

Both games were played almost equally frequent, Zuma being preferred a little (53.6%, N = 45) over MMD (46.4%, N = 39). On average Zuma was played 4.09 times (SD = 0.81) and MMD 3.55 times (SD = 1.40). The frequency for MMD was nonparametric (D(11) = 0.37, p < .0001) and there were no significant differences in playing frequencies between MMD and Zuma. The mean playing time for Zuma was 1748.16 seconds (~29 minutes), which was not significantly different from the mean playing time invested in MMD of 1905.08 seconds (~32 minutes). When looking at correlations for Zuma,

IV Neither the dependent t-test nor the Wilcoxon signed-rank test showed significant differences.
we did not find any correlation between flow and playing time, between immersion and playing time, between flow and playing frequency, and between immersion and playing frequency using Pearson’s correlation coefficient \( r \). In contrast to this in MMD, we found a strong positive correlation between flow and playing time, \( \tau = .58, p \text{ (one-tailed)} < .01 \), and between immersion and playing time, \( \tau = .46, p \text{ (one-tailed)} < .05 \). Both flow and immersion were correlated with playing frequency (flow: \( \tau = .53, p \text{ (one-tailed)} < .05 \); immersion: \( \tau = .48, p \text{ (one-tailed)} < .05 \)). This suggests that high flow and immersion ratings correlate with longer gaming sessions in higher frequencies. Similarly to what we observed in the Zuma condition, there was a strong positive correlation between flow and immersion scores in the MMD condition, \( r = .62, p < .05 \).

**System (Game) Usability Scale**

After the subjects had played the games for three weeks, we used the slightly modified SUS to assess the perceived usability of the games. According to Tullis (2008), an average SUS score under 60% is relatively poor and one over 80% can be considered as good. The average SUS score for MMD was quite low (\( M = 58.41, SD = 24.32 \), Cronbach’s \( \alpha = .92 \)) whereas the SUS score for Zuma can be considered to be very good (\( M = 86.59, SD = 9.83 \), Cronbach’s \( \alpha = .56 \)). In addition, we let the participants rate the overall quality of each game on a scale from 1 (bad quality) to 10 (excellent quality). Again, Zuma scored relatively high on this scale (\( M = 7.32, SD = 1.19 \)) in comparison to MMD (\( M = 5.09, SD = 2.43 \)). When asked to rank the games in order of preference, 72.7% ranked Zuma and only 27.3% ranked MMD as their favorite game. When we looked for a correlation between quality rating, SUS score, flow and immersion scores for Zuma, we found a positive correlation between flow rating and SUS score \( (r = .67, p < .05) \) as well as a positive correlation between immersion score and quality rating \( (r = .72, p < .05) \), but no significant correlation between quality rating and SUS score. Correlations with playing time were also insignificant. When looking at the same correlations for MMD, we found a positive correlation between SUS score and quality rating \( (r = .60, p < .05) \).

\[ V \] Which was non-parametric as indicated by the Kolmogorov-Smirnov test.
Discussion

For one of the games (Maniac Mansion Deluxe) a significant positive correlation, between flow and immersion on one hand and playing time and frequency on the other, was found. This indicates that, at least for such a narrative-based game, the amount of time invested in the game as well as the frequency of playing it have a positive relationship with the self-reported feelings of flow and immersion. Another factor for this positive relationship regarding time and frequency is the fact that playing an adventure-genre game usually requires more time investment per session than a puzzle game does. The completely different game mechanics of these different genres lead to different playing times. However, this does not explain the correlation with playing frequency. Despite, our results for Zuma do not replicate this effect which is interesting since Zuma was ranked as the more favorite game for many participants, which would led one to believe an increased correlation of flow and playing frequency. A possible reason is that Zuma’s game concept of challenge-based action puzzle gameplay might contain other aspects with a strong impact on frequent play besides flow and immersion.

As for our second hypothesis (H2), we did find a significant correlation between game quality rating and usability score for the game MMD. Thus, the lower quality ranking of this game might be explained by the usability (SUS) score of the game, suggesting that aspects of interface and functionality lead the game to be perceived as being as of less quality than the Zuma game. It could also be that the SUS rating indicates a poor gameplay experience although its questions mainly focus on interface and functionality. There is obviously a need for a different measure that does discriminate more clearly between usability issues and gameplay quality than our modified version of the SUS can account for. In addition, no such correlation between game quality rating and usability score exists for the Zuma game. For this game, the flow and immersion ratings both had a strong relationship to the reported usability score, suggesting that either when a game is perceived as providing an experience of immersion and flow, this does affect how the game’s usability is rated or that a game with poor usability does not support flow and immersion. At least for an action puzzle game this is a very interesting relationship between game experience and usability. As a result, we draw the conclusion that – while standard usability criteria may apply
in studies, which focus on games with a clearly separated GUI (like the command system in MMD, which is separated from the game screen) – other games with more hybrid forms of interfaces and feedback mechanisms may be better evaluated by using evaluation methods for playing quality instead of applying technical usability standards.

Although the described difference in quality ratings of the two games was not statistically significant, we do conclude a higher quality rating of Zuma based on the explicit preference of nearly 73% of the players. Interestingly, the participants’ perceived differences in quality and experience between the two games did not lead to either game being played less often. Thus, we have to reject our last hypothesis (H3) that the preferred game would be played longer and more often. The games may have complementary strengths and weaknesses: Zuma shows a clearly higher flow score along with positive affect and competence, while MMD scores higher in sensory and imaginative immersion. This suggests that flow and immersion are two experiential constructs, which may not be equally enjoyable (as indicated by the different quality and positive affect ratings), but nevertheless prove to be equally motivating to play the game. It also leads us to question whether playing frequency and time are suitable as dependent measures to investigate positive gameplay experience and subjective quality of a game.

In summary, these results show clear differences in how the various aspects of game experience, quality, and usability influence each other for the two different game genres. This raises the question if these characteristics remain constant for games that are each very similar to the two tested games and whether these differences apply to other game genres. With this being the case, the combination of methods in the present study, GEQ, SUS and quality assessment through gameplay logging, could lead to patterns of correlations that help to support game genre classifications. For further investigating this topic, we suggest to create groups of similar games in a larger evaluation setup involving more subjects and a broader variety of genres.

Conclusion and future work

In this study, we examined the characteristics of two games in distinct genres concerning the aspects of game experience, subjective quality, and game usability.
This combination of methods led to the identification of strong interactions within the three aspects. Gameplay experience dimensions flow and immersion may not be equally enjoyable, but seem to be equally motivating to play a game. On the one hand, usability ratings may be positively affected when a game provides experiences of immersion and flow or on the other hand flow and immersion may be negatively affected by poor usability ratings. These results strongly differ for the two different games, implicating an approach for game genre classification based on correlation patterns involving game experience, quality, and usability. This pilot study provides a basis for more comprehensive future research defining the mutual influence of gameplay experience, quality, and usability. By formalizing gameplay experience in coherence with other aspects of gameplay, we might be able to better evaluate, categorize, and design games in the future.

Acknowledgments

We would like to express our gratitude to our colleagues Sophie Stellmach, who helped to conduct this study, and Craig Lindley, who contributed with helpful discussions. We would also like to thank Ian Livingston and Regan Mandryk from the interaction lab at the University of Saskatchewan. Part of this research was funded by the European Commission under the 6th Framework Programme: New and Emerging Science and Technology (NEST), Project FUGA - The Fun of Gaming: Measuring the Human Experience of Media Enjoyment (Contract: FP6-NEST-28765). We also thank all our participants.

References


Towards a Serious Game Experience Model: Validation, Extension and Adaptation of the GEQ for Use in an Educational Context

Frederik De Grove* - Jan Van Looy - Cédric Courtois

MICT-IBBT Ghent University

Abstract

In this paper, we present the results of game experience measurements of three design stages of the serious game Poverty Is Not a Game (PING) using the FUGA Game Experience Questionnaire (GEQ) extended with a Perceived Learning (PL) module. It is hypothesized that subsequent design stages will evoke a more positive game experience and higher PL. In a first step the factor structure and convergent and discriminant validity of the existing GEQ modules are tested yielding disappointing results. Next an adapted version is proposed yielding more acceptable results. Based on this model, the different design stages are compared failing to yield significant differences either for most GEQ dimensions (except for challenge and competence which is probably related to usability issues) or for PL. Significant differences were found between classrooms however pointing to the importance of taking into account context in future research. Moreover, game experienced proved to have an effect on the experience of learning.

Introduction: Serious Game Experience

Although game experience has become an important concept in recent academic research concerning video games (see e.g. Ijsselsteijn, de Kort, Poels, Jurgelionis, Bellotti, 2007), attempts to clearly define it are scarce. This can probably be attributed to the complex, subjective and dynamic nature of the idea of experience (Buchenau & Suri, 2000). Ermi & Mäyrä (2005) describe game experience as ‘an

*Frederik.DeGrove@UGent.be
ensemble made up of the player’s sensations, thoughts, feelings, actions and meaning-making in a gameplay setting.’ This definition gives us a notion of the concept by implying a relationship between game, gamer and context. However, it also uncovers its problematic nature by referring to a variety of agent-dependent, hence subjective processes making clear that ‘The experience of play comes in so many forms that creating a single catalogue that takes all of them into account would be an impossible task’ (Salen & Zimmerman, 2004, p.314).

It could therefore be more interesting to approach game experience from a different angle than an ontological one. By reframing the question as why the concept of game experience has become a topic of interest for academic research, it can be narrowed down to a workable proportion. As such, game experience is directly related to the motivational aspect of gaming which is in turn connected to the question of what makes video games enjoyable (see e.g. Vorderer, Hartmann & Klimmt, 2003; Sweetser & Wyeth, 2005; Shen, Wan & Ritterfeld, 2009). In that respect, game experience can be considered as the underlying mechanism that makes video games intrinsically motivating and fun.

A central concept in relation to enjoyable experiences is that of flow or optimal experience which can be described as ‘an optimal, intrinsically motivating experience induced by an activity in which one is fully absorbed’ (Csíkszentmihályi, 1990). Although Sweetser & Wyeth (2005) adapted this construct into a game-specific model with eight correlating dimensions (concentration, challenge, skills, control, clear goals, feedback, immersion and social interaction), attempts to thoroughly operationalise the multifaceted construct of game experience are scarce as most attempts are limited to using either one dimension of the game experience or employing a narrowed-down version of the flow concept (see e.g. Douglas & Hargadon, 2000; McMahan, 2003).

From 2006 to 2009, however the “Fun of Gaming” (FUGA) project funded by the European Community has worked towards measuring the human experience of media enjoyment (see IJsselsteijn, van der Hoogen, Klimmt, de Kort, Lindley, Mathiak, Poels, Ravaja, Turpeinen & Vorderer, 2008). One of its core tasks was to develop a Game Experience Questionnaire (GEQ) to measure the game experience. This self-report measure consists of three modules: a core module, a social presence module
and a post game module and has been used in several studies. For instance, Gajadhar, de Kort & IJsselsteijn (2008) used the GEQ to determine the effect of social context on video game enjoyment and found social presence to be a mediating factor. Nacke, Stellmach, Sasse & Lindley (2009) on their turn used the core module of the GEQ to assess the game experience in a gaze interaction game while Nacke & Lindley (2008) used the core module in combination with psychophysiological measures and video recording to explore flow and immersion in a First Person Shooter. Except for the initial construction of the GEQ however, no information is available regarding the validity of the GEQ. Moreover, as far as we know, no attempt has been made as yet to validate the application of the GEQ to the domain of serious or educational games. Considering these issues, this leads us to formulate our research question.

**RQ1: Do empirical data confirm the construct validity of the GEQ?**

In this paper we report on the testing of the alpha, beta and release candidate versions of the game PING. Since the GEQ was primarily constructed to be used for measuring the experience of commercial games, its use in a formal learning environment requires the adoption of another experiential dimension, c.q. the experience of learning itself, which has in previous research been conceptualized as a flow effect (Kiili, 2005). The only attempt we know of to link perceived learning with game experience is that by Fu, Su and Yu (2009) which consisted of a scale based on Sweetser & Wyeth’s Gameflow Model (2005) adapted to the specificity of what they call e-learning games. However, as this measure is only applicable to video games with clearly defined learning outcomes, it was decided that it was unfit to be used for a video game under development of which the primary aim was to raise consciousness about poverty. It was therefore decided to build a perceived learning scale that takes into account the taxonomy used by Rovai, Wighting, Baker and Grooms (2009) (including affective and cognitive learning) and the first two levels proposed by Kirkpatrick (1998) to evaluate training courses, namely Reaction (level 1) which pertains to the affect of the respondent towards the learning method and Learning (level 2) which corresponds with the affective and cognitive learning dimensions used by Rovai et al. (2009).
As enjoyment in games is strongly intertwined with aspects of usability (see e.g. Federoff, 2002), it is not unreasonable to assume that improvements in the design of a video game will have their effect on the game experience.

H1: The game experience will become significantly more positive over the three design stages of PING.

Furthermore, it can be expected that the game experience and the experience of learning are positively correlated (Kiili, 2005).

H2: There is a positive effect of the game experience on perceived learning.

Since we expect the game experience to become more positive over the different design stages of the game, as a consequence, perceived learning will likewise be affected.

H3: Perceived learning will rise significantly over the three design stages of PING.

The Game

PING (Poverty Is Not A Game) was commissioned by the King Baudouin Foundation and is an initiative as part of the European Year for Combating Poverty and Social Exclusion (2010). Its primary aim is to raise consciousness in adolescents concerning poverty and social exclusion in a way that is close to their everyday lives. The game takes place in a three-dimensional environment which represents an average Western European city. Players can choose between a male or female avatar. Although the decision to play with a certain avatar has an impact on the storyline, the central message the game wishes to convey stays the same. It hopes to raise consciousness concerning the mechanisms underlying poverty and is specifically aimed at what is sometimes referred to as the fourth world.

No detailed accounts exist concerning what has changed between the different design stages of PING (Alpha, Beta and Release Candidate). An
informal conversation with the game developers revealed that changes primarily pertained to the story of the game and to the navigation in the game. The Alpha stage only consisted of a rudimentary storyline while orientation in the 3D game world was a challenge as no maps were readily available. In comparison, the Release Candidate (RC) provided a fully developed story which could be finished in about 50 minutes. Navigation was facilitated by a mini-map with GPS functionality. Similar in all design stages is the fact that PING had no sounds or music.

**Method and Procedure**

*Data Collection*

Data were collected by testing PING in a total of 22 classrooms. To measure game experience, an online survey consisting of several blocks was used. A first block was the core module of the GEQ and inquired how the player felt whilst playing the game (92 five-point Likert scale (FPLS) items: Not at all, Slightly, Moderately, Fairly and Extremely). A second block was the social presence module and consisted of 25 FPLS items. Then a first part of the perceived learning scale followed (10 FPLS items). Next are the post-game module (21 FPLS scale items) and the second part of the perceived learning scale (10 FPLS items). A subsequent block of the survey explores the gaming behavior of the respondent. The three final questions asked about sociodemographic characteristics (gender, age and education level). In total, the survey was filled out by 373 respondents of which 340 were retained after data cleaning.

*Population and Sampling*

As PING has been specifically developed for use in a classroom, the obvious consequence is that it needed to be tested in the classroom as well. For the sake of uniformity it was tested in third and fourth grade classes only (age: 14-16 years). Furthermore, participation was limited to General and Technical educational levels. Our population thus consists of all third and fourth grade students attending a General or Technical class in a school certified by the Flemish Government.
Sampling was performed on the basis of a database of schools listed on the website of the Flemish Ministry of Education (http://onderwijs.vlaanderen.be/) and on the basis of a database of schools provided by the King Baudouin Foundation. Schools were picked out at random and were sent an e-mail asking for their cooperation. Schools that did not respond were contacted by telephone.

Procedure
The general procedure was the same for all test phases and for all classes. The teacher in charge of the class was asked to give a brief introduction about the video game and about the subject matter (poverty). After this introduction, students could start playing. Due to the status of the game, during the Alpha phase, students were allowed to play for approximately 25 minutes while students were allowed to play for 50 minutes during the Beta and Release Candidate phases. After playing, students were asked to fill out an online survey.

Scale Validation
We first assessed the construct validity of the GEQ to evaluate its appropriateness in the context of a pre-launch serious game in a formal learning environment. Construct validity was accounted for by testing the factor structure of the core module, the social presence module and the post game module and by using measures accounting for convergent and discriminant validity. To test the factor structure we made use of confirmatory factor analysis which ‘is a way of testing how well measured variables represent a smaller number of constructs’ (Hair, Black, Babin, Anderson and Tatham, 2006). As proposed by Hair et al. (2006) we checked convergent validity by means of the coefficient alpha and the average percentage of variance extracted (VE). Discriminant validity was examined by comparing the VE of two constructs with the squared correlation of those two constructs. The rationale behind this is that a latent construct should explain the variance in its items better than that it explains another construct. To complement the GEQ a perceived learning scale was constructed and tested for its construct validity.
Results

Construction of a Serious Game Experience Model - Core Module

Goodness-of-fit indices (N=330, $\chi^2$/df = 2.86, CFI = .78, TLI = .75, RMSEA = .075, CI90 = .071, .080) were not satisfactory as CLI and TLI scores suggest that the proposed model did not fit our data. This is confirmed when checking for convergent and discriminant validity as none of the core module’s dimensions has an acceptable VE. This ‘indicates that on average, more error remains in the items than variance explained by the latent factor structure imposed on the measure’ (Hair et al., 2006) which points to a problematic scale. When checking for discriminant validity, six concepts show considerable similarity with other concepts, i.e. Competence, Immersion, Flow, Positive Affect, Challenge and Annoyance. More specifically Competence, Immersion and Flow explain each other better than they explain the variation in their own items. The same is true for Annoyance and Challenge and for Positive Affect and Competence. Only the concept Negative Affect proves to be different enough.

GEQ Social Presence Module

Goodness-of-fit indices of the social presence module were similar to that of the core module (N=330, $\chi^2$/df = 3.68, CFI = .84, TLI = .81, RMSEA = .090, CI90 = .081, .099), which indicates that the data do not fit the proposed model. Moreover none of the dimensions has an acceptable VE statistic while coefficient alphas seem to suggest reliable scales. Discriminant validity was not satisfactory for Empathy and Behavioral Involvement and for Negative Affect and Behavioral Involvement.

GEQ Post Game Module

Goodness-of-fit statistics show that our data do not fit the proposed model (N=330, $\chi^2$/df = 4.44, CFI = .84, TLI = .81, RMSEA = .102, CI90 = .093, .111). Concerning convergent validity, all dimensions yielded an acceptable coefficient $\alpha$ but only Tiredness explained sufficient variance. Discriminant analysis for the post game module showed that three of the four dimensions are not different enough. It concerns Negative Experience which relates to Tiredness and Returning to Reality.
Construction of a Perceived Learning Scale

Of the total item pool (20 items), nine items were retained to construct the perceived learning scale. Two items composed the construct of Affective Gaming which assesses how one responds to receiving education through video games while the construct of Learning (7 items) explores to what extent the respondent thinks they have learned something on the topic of poverty. This model was inputted in AMOS and resulted in a good fit (N=330, $\chi^2/df = 1.79$, CFI = .99, TLI = .98, RMSEA = .049, CI90 = .025, .071) while both Convergent and Discriminant validity proved to be acceptable.

Towards a Serious Game Experience Model

As the current structure of the GEQ was considered as inadequate for our further exploration (RQ1), it was decided to build our own serious game experience model in which the item pool and dimensions of the GEQ served as a starting point. Furthermore, we aimed to incorporate our own perceived learning scale.

The first decision we made was to leave out the social presence module since this module was not deemed fit to assess the complex and rich nature of social interactions that emerge during gameplay in a formal learning environment. Moreover, the construct of Tiredness of the post game module was omitted as it proved to be highly context-dependent (being tired was not due to gaming but to the moment of testing).

All three modules were reviewed and adapted based on item distributions, coefficient alphas, corrected item-total correlations and the one-dimensionality of the dimensions. This eventually resulted in a model in which game experience and post game experience were conceptualized as second order constructs. Game experience is composed of eight first order constructs: Competence (2 items), Vividness (3 items), Negative Affect (2 items), Positive Affect (3 items), Immersion (2 items), Challenge (2 items), Affective Gaming (2 items) and Learning (7 items). Post game experience consists of Positive Experience (3 items), Negative Experience (3 items) and Returning to Reality (2 items). Annoyance was omitted since this construct proved to be impossible to build from the applicable data. Furthermore, the construct of Challenge did not result in an acceptable scale either but was retained on the basis of theoretical and practical considerations.
Fit indices of the proposed Serious Game Experience model yielded an acceptable fit (N=330, \( \chi^2/df = 1.84 \), CFI = .93, TLI = .92, RMSEA = .050, CI90 = .045, .056) This model (see Figure 1) served as the foundation for our further exploration of the serious game experience of PING during its subsequent design stages. It shows that Vividness (.94), Positive Affect (.93), and Immersion (.76) are strong predictors for Game Experience while Competence (.37) and Affective Gaming (.42) are moderate predictors. The fact that Learning has a standardized regression weight of .50 confirms our hypothesis that there is a positive effect of the game experience on perceived learning (H2). Remarkable is the fact that Game Experience has no effect whatsoever on Challenge (.01). As for Post Game Experience, Returning to Reality (.97), Positive Effect (.65) and Negative Effect (.74) are strong predictors. Finally, by connecting the error terms of Affective Gaming and Learning, of Competence and Challenge and of Negative Affect and Positive Affect, we wanted to examine if those constructs shared variance that was not explained by the Game Experience construct. The rationale behind this was that it could be expected, on theoretical grounds, that those constructs share other common causes for their variation. This was confirmed for Competence and Challenge (-.57), Affective Gaming and Learning (.36) and for Positive Affect and Negative Affect (-.63).

The Evolution of Game Experience - During Subsequent Design Stages

To compare how the different dimensions behaved during subsequent design stages, we analyzed the variance within groups and between groups (ANOVA, power = .98, Effect size = .025). Results indicated that, over the three different design stages, only Competence (p < .005, F = 6.03, df = 335) and Challenge (p < .001, F = 5.37, df = 330) differed significantly. Post-hoc tests (Scheffe) show that these differences are to be found between the Alpha stage and the RC stage. This applies to Competence (p < .003) as well as to Challenge (p < .021). On average, Competence scores were lower during Alpha testing (M = 2.73, S.D. = .13) than during RC (M = 3.26, S.D. = .09) testing while scores for Challenge were higher for Alpha testing (M = 2.05, S.D. = .11) compared to RC testing (M = 1.72, S.D. = .07). Learning is marginally significant (p < .065, F = 2.80, df = 329) but differences for the design stage are situated between Alpha and Beta where average Beta scores (M = 2.95, S.D. = .08) were higher than Alpha scores (M = 2.79, S.D. = .09).
**Classroom Comparison**

To check if our sample size was big enough to execute an ANOVA with 22 groups, we performed a power analysis. With a power of .77 we were only capable to reliably detect large or, to a lesser degree, medium differences. Notwithstanding our relatively low power, a considerable number of dimensions proved to differ significantly between classrooms (Competence, p < .001, F = 2.317, df = 335; Vividness, p < .012, F = 1.880, df = 330; Challenge, p < .020, F = 1.784, df = 330; Negative Affect, p < .000, F = 3.428,

*Figure 1. Serious Game Experience Model*
df = 338; Positive Affect, p < .000, F = 2.684, df = 335; Affective Gaming, p < .000, F = 2.234, df = 329; Learning, p < .011, F = 1.900, df = 329; Negative Experience, p < .000, F = 2.642, df = 329). Only Immersion, Positive Experience and Returning to Reality do not differ. Regrettably, we did not have enough data at our disposal to identify differences between individual classrooms. Furthermore, classroom is a level 2 unit. By using an ANOVA we do not know whether the differences we found are related to the different compositions of the classrooms (e.g. more females in some classrooms) or if they were caused by level 2 variables (e.g. a positive group atmosphere).

**Conclusion / Discussion**

The testing of PING yielded some interesting results. A first remarkable outcome is the fact that there is a strong effect of the game experience on perceived learning which confirms that a positive, enjoyable game experience contributes to the experience of perceived learning (H2).

When it comes to the game experience and the different design stages, only Competence and Challenge differ significantly between the Alpha and RC stage. When we add the fact that, first of all, variation in Challenge was not explained by the game experience but shared unexplained variation with Competence (-.57), and secondly, that one of the two major changes during the design stages pertains to the usability of the game (navigation), some interesting assumptions can be made. On theoretical grounds, both Competence and Challenge can be connected to usability issues in terms of effectiveness and efficiency. As such, usability can be considered as a prerequisite for a good game experience but it is not equal to it. Hence, it is possible that Challenge (or the way it was operationalised) is actually a measure of usability. This would explain why it failed to fit in our game experience construct while still being related to it. Furthermore, this could also explain some of the error variance of Competence. As such, we did not find a significant positive change in the game experience during the different design stages (H1) but we did find a significant change in the experienced usability of PING.

Although there is a positive effect of the game experience on Learning, the change is only marginally significant between the Alpha and Beta stages of the
game. This significance is probably due, however, to the fact that the Beta stage had some atypical distributions. When checking for interaction effects with Gender of Educational level by means of a multivariate analysis the difference in Learning ceased to be (marginally) significant. The fact that perceived learning does not change positively during the subsequent design stages (H3) is surprising. Especially because the storyline was one of the major changes in the design flow. Furthermore, students were allowed to play the Beta and RC versions longer than the Alpha version which could have resulted in a better learning experience. On the other hand, this finding is not illogical if we take our serious game experience model into account. As most of the dimensions do not vary between the different design stages, it is logical that learning does not vary either. More specifically, if the storyline would have changed enough, this would have been reflected in the concept of Vividness. Consequentially, the experience of perceived learning would have changed too. This indicates that the changes in the subsequent design stages of PING were not large enough to evoke an improved learning experience. Considering the pre-launch status of the game, an interesting starting point could have been to use the GEQ in combination with a validated usability measure. That way, improvements in usability could have been linked to game experience. As such, it would be interesting if future research on video games that are under development would incorporate usability as well as game experience measures.

Finally, it is interesting to see that there seem to exist strong differences between classrooms on most of the game experience dimensions. Perhaps the most remarkable result is that, when gaming, only Immersion does not differ significantly between classrooms. A possible explanation could be that social interaction during gameplay prevents Immersion to go above a certain level while the absence of sounds or music could have been a decisive factor in stimulating social interaction. Equally intriguing is the fact that constructs such as Competence, Vividness, Challenge and Learning seem to have a collective component. With our current dataset, however, we could not explore this further. Future research could consider using focus groups to explain these findings. Another approach might be to use an experimental design in which the content-related variable is manipulated.
As such, using the GEQ for a serious game in a pre-launch status in a classroom context might not be ideal as some of the items are highly context dependent (Flanders versus the Netherlands, education level of respondents, etc.). However, our own proposed model is susceptible to criticism as well. With the exception of Learning, the constructs of our model contain only two or three items. This is barely enough to cover the concepts they intend to measure hence resulting in operational narrowing of the concepts involved. Results also indicate that the concept of Challenge does not fit its theoretical content (conceptual displacement). Moreover, we built the model from the available data. Since our data ensued from one video game only, the possibility exists that our model cannot be generalized. Further testing will reduce uncertainty. However, future attempts to construct a game experience model should try to include concepts such as Concentration, Feedback, Clear Goals and Control while a model aimed at measuring experiences in a classroom context should try to incorporate a measure that is able to take context effects into account such as the rich diversity of social interactions. Be it in a quantitative way in a serious game experience model or in a qualitative, ethnographic way.

References


**Play.** Selected papers of the 2005 Digital Games Research Association’s (DiGRA) Second International Conference, 2005.


Evaluating Educational Game Experiences in a Classroom Context: Implications for Qualitative Research

Steven Malliet*
Media, Arts & Design Faculty, University College of Limburg
Centre for Media, Policy and Culture, University of Antwerp

Niels Quinten, Veerle Van der Sluys
Media, Arts & Design Faculty
University College of Limburg

Abstract

Although quantitative methods for assessing digital game playability have been extensively documented, research on the use of qualitative evaluation methods remains scarce. In this chapter we explore, based upon practical experiences obtained during the development of a number of educational micro-games, the methodology of qualitative playability research. More specific we focus on the use of qualitative evaluation methods in a classroom context. We formulate practical guidelines in a number of areas: duration of evaluation, role of the observers, number of evaluations and observer/participant ratio.

Introduction

In the past years researchers have developed a strong interest in the issue of digital game playability. In a media literacy context, the term has been used as a counterpart to the readability of novels or textbooks (Kücklich, 2003), and as a means to identify a gap between digital natives and digital illiterates (see Prensky, 2001a). From a cultural perspective, scholars have developed several models to analyze the gameplay

---

*Steven.Malliet@ua.ac.be

vi This investigation was performed in the context of the Gimme, Gimme, Gimme a Game project, in collaboration with the centre for Education & ICT (ED+ict), University College of Limburg. Part of this research was financially supported by the PWO (Projectmatig Wetenschappelijk Onderzoek) program.
mechanisms that constitute the text of a digital game (Aarseth, 2003; Konzack, 2002; Consalvo & Dutton, 2005; Malliet, 2007). Others have argued that, in order to explain why games can be engaging, moving, persuasive or fun, issues of rule-based gameplay should be considered in addition to issues of representation (Frasca, 2003).

The playability of digital games has been an important concern for game developers and game designers as well. Not only has the game industry become a multi-billion dollar business that generates returns comparable to those of other mainstream media industries, but also the development of a contemporary commercial game title involves enormous production costs. Research regarding the elements that can make the difference between commercial success and failure has become invaluable in the game development cycle - to such an extent that, in most contemporary game productions, 8% to 12% of the total budget is spent investigating the experiences of targeted players (Schafer, 2007). Methods of performing player research range from organizing co-design workshops and performing focus group conversations with Prospected players, to the use of formal methods to evaluate the flow and the player-friendliness of a game. These latter methods are commonly classified under the denominator of playability research, a field that has evolved from usability research in Human-Computer Interaction.

In this paper, we will focus on a number of methods for playability evaluation that have recently been developed, and we will complement existing research with insights that were gained through our own experience, during the development of a number of educational micro-games. We will argue that, thus far, existing playability methods have focused, either on strictly formal aspects of game content (Desurvivre et al., 2004; Desurvivre & Wiberg, 2009), or on user-experience related aspects such as presence (de Kort et al., 2007). Researchers have also highlighted the mechanisms of interaction between content and player by means of biometric or psychophysiological measures (Nacke, 2009). While the use of quantitative evaluation methods has been well documented in the past years, the use of qualitative game evaluation methods has received significantly less attention. Nevertheless, in user-centered and design-oriented research, qualitative methods are generally considered a useful, and often necessary addition to the objective methods derived from psychological or physiological research (Laurel, 2004). The main goal of this paper is to fill this gap, and
to formulate, based upon practical experience, a number of guidelines for observing
and evaluating player-game interactions. More specific, we will focus on the evaluation
of player experiences in a classroom context. Qualitative methods have traditionally
been considered useful for the investigation of contextual meanings, experiences in
a natural context and collective practices (Denzin & Lincoln, 2005; Guba & Lincoln,
2005). These aspects take a central role in the study of educational gaming, given the
relevance of issues such as didactic context (Egenfeldt-Nielsen, 2005), group dynamics
(Van Eck, 2006) and situated learning (Gee, 2003).

**Playability and related issues**

Initial research on the method of digital game playability was largely inspired by the
evaluation heuristics that have been developed to assess the learnability, efficiency
and memorability of interfaces in Human-Computer Interaction research (Schafer,
2007). This has resulted in a number of adaptations of existing usability and likeability
questionnaires, applied to the specific needs of digital game design. Within these
initial methods, several elements of game content have been identified as constructive
to the overall experience of fun. Federoff (2002) made the distinction between play
elements (such as AI, storylines, or audiovisual elements), game mechanics (such as
involvement, or intuitiveness) and game interface elements (such as feedback and
error prevention). Similarly, Järvinen et al. (2002) identified 4 types of playability:
structural playability, functional playability, audiovisual playability, and social
playability. More recently researchers have extended this framework, and have
constructed a set of standardized, formal research tools. Desurvivre et al. (2004)
and Desurvivre and Wiberg (2009) constructed the Heuristics to Evaluate Playability
(HEP) tool, consisting of 50 items that correspond to a wide range of commonly used
design patterns. Schaffer (2007) identified over 50 design guidelines that should
enable the game developer to improve the playability of a game during the initial
stages of the development cycle.

The studies mentioned above mainly serve to investigate the content and the
programmed code of digital games. Only recently the concern has been raised that, in
order to sufficiently evaluate the immersive potential of a game, a central role should
be given to the assessment of player experiences (Nacke, 2009). Nacke et al. (2009) argue that not only the relationship between game content and game design should be investigated, but also the relationship between game design and user experience. The authors point out that, in a number of related fields (such as interactive art design or multimedia design), experience design has become a common practice, and experience research has become a conventional phase in the development cycle. Following this perspective, researchers have used psychophysiological methods to evaluate the relationship between game content and play experience. Biometric measures have been applied in the evaluation of aspects such as affective level design (Nacke, 2009), the immersive use of game audio (Grimshaw et al., 2008; Nacke, 2009), or the relationship between game controls and flow (Nacke 2009). At the same time, researchers with a background in media research and psychology have developed scales for performing post-hoc tests of experience-related aspects such as social presence (de Kort et al., 2007), involvement (Calleja, 2007), or perceived realism (Ribbens & Malliet, 2009).

Although playability was initially developed as a qualitative method, allowing experts and designers to make temporal and iterative evaluations of games-in-progress, current research strongly focuses on the use of quantitative metrics. A few authors have addressed the need for complementary methods that address the same issues from a qualitative point of view (Poels et al., in press). Methods such as focus group interviews or collaborative design frameworks have been forwarded as valuable additions to the objective methods derived from psychology or psychophysiology, especially when the contextualized and collaborative experience of the prospected players is considered an important point of attention. Qualitative inquiries into the relationship between player and game have thus far mainly been based upon general methodological literature on focus group interviewing, or on research-by-design.

Setup of this study

In the project described in this paper, extensive use was made of collaborative design methods and qualitative user experience tests, during different phases in the development of a number of educational micro-games. Because these methods were specifically applied to the educational context of testing in a school environment,
and because the practical application of these methods depended strongly on the properties of the games that were developed, a number of practical challenges were met that have not been documented in traditional textbooks on qualitative research or on research-by-design. In the remainder of this paper we will give an overview of these challenges. Based upon a critical analysis of the practical solutions that were developed, we will formulate practical guidelines that can be used by researchers who intend to make use of similar methods in future research.

**Development context**

The games were developed by master students in the context of a game design module at the MAD-Faculty, University College of Limburg. The goal of the module was to translate educational content into game play mechanisms, guided by literature on serious game design and on the learning models associated with digital game play. In collaboration with a local high school, three courses were selected: Biology, Mathematics and French. Within each course, the content of one lesson was selected, and was translated by the students into a digital micro-game. Each team was given creative freedom with respect to the technologies and game play mechanisms that were used. The central focus of the module was on interaction design, and on the adaptation of traditional textbook material into a meaningful player experience, as described by Salen & Zimmerman (2003).

As is demonstrated in table 1, the approach of the module resulted in the development of different types of games, as well as in the composition of different types of development/evaluation teams. With respect to the communication of educational content, different approaches were taken by the different teams. In two games (G2 and G5) the decision was made to convey information by using a quiz mechanism similar to the traditional assessment methods in the classroom. The developers of one game (G3) made extensive use of narration and dialogue in order to improve the learners’ knowledge of French vocabulary and syntax. Finally, in two games (G1 and G4) the educational context served as a background against which an engaging game play experience was elaborated. In these games knowledge of the course content served as an aid to make progress, and knowledge acquisition was implemented as an iterative process of trial and error. As such, with respect
to the balance between play elements and educational elements that is propagated in the field of digital learning (Prensky, 2001b), each team took a specific approach and made well-considered decisions.

In addition, different development teams made use of specific interfacing technologies and play mechanisms. In two games (G1 and G5) a physical interface was implemented, resulting in a game play experience that relies heavily upon direct player-game interaction. In one game (G2) the combination of a complex rule system and a team-based setup resulted in a tactical and collaborative game play mechanism. In one game (G3), narration and character design were used as the main vehicle for message construction. Finally, in one game (G4) character design, narration, puzzle solving and strategic game play were integrated into a hybrid online game.

**Evaluation context**

Player experience evaluations were performed during three phases in the development cycle. First, in the stage of paper prototyping, a co-design session was organized with

<table>
<thead>
<tr>
<th>Educational topic</th>
<th>Game play mechanisms</th>
<th># Developers</th>
<th># Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1 Biology</td>
<td>Physical interface - driven game. Player-game interactions against background of human skin structure.</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>G2 Biology</td>
<td>Strategic quiz game against background of human skin structure. Complex rule structure to instigate tactical game play.</td>
<td>3</td>
<td>4 – 8</td>
</tr>
<tr>
<td>G3 French</td>
<td>Narrative game. Learning French through conversations, with multiple possibilities for correcting mistakes and looking up correct answers.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G4 Mathematics</td>
<td>Hybrid game, relying upon character design and memory-based missions to learn using statistical summary methods.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G5 Mathematics</td>
<td>Physical interface – driven game. Learning about statistical methods through physical responses to quiz questions.</td>
<td>3</td>
<td>2 – 6</td>
</tr>
</tbody>
</table>
students from the cooperating school. These sessions were carried out according to the principles of participatory design described by Muller (2007). Second, a mixture between collaborative design and qualitative playability evaluation was applied to a first electronic prototype of each game. Third, the final version of each game was evaluated. During these final evaluations a post-hoc playability questionnaire was used, based upon the scales developed by Desurvivre et al. (2004) and Desurvivre & Wiberg (2009). The use of this quantitative method was complemented with the use of qualitative interpretive methods. The evaluators made ethnographic notes, observing player responses, player choices, and game progress, according to the principles of ethnographic observation in a school context outlined by Kantor et al. (1981). Finally, post-hoc in-depth interviews (in the case of single-player games) and focus group interviews (in the case of multi-player games) were performed. These interviews were conducted in a structured way, complementing the quantitative playability evaluation.

The target group were students in the 4th grade of General Secondary Education. The games were tested in three classes (total N=57) during three class moments of one hour. As is demonstrated in table 2, this resulted, with the exception of one game being tested with only three students (G3), in each game being tested with 10-20 persons. The number of evaluations performed and the number of students involved in one evaluation session, depended upon two aspects: the single-player / multi-player characteristic of a game (with multi-player testing involving more than one participant), and the duration of a playing session. The duration of play testing varied from 15 minutes (G3) to 50 minutes (G4).

Table 2: Overview of evaluation procedures

<table>
<thead>
<tr>
<th># Evaluators</th>
<th># Players / session</th>
<th># Evaluations</th>
<th>Mean duration of evaluations</th>
<th>Evaluation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>5</td>
<td>1</td>
<td>15</td>
<td>20 min.</td>
</tr>
<tr>
<td>G2</td>
<td>3</td>
<td>4 – 8</td>
<td>12</td>
<td>25 min.</td>
</tr>
<tr>
<td>G3</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>15 min.</td>
</tr>
<tr>
<td>G4</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>50 min.</td>
</tr>
<tr>
<td>G5</td>
<td>3</td>
<td>2 – 6</td>
<td>5</td>
<td>30 min.</td>
</tr>
</tbody>
</table>
Although all evaluations were performed following the same principles of ethnographic research and playability evaluation, the choice was made to have different teams emphasize different aspects of the game content during the evaluations. Two teams (G1 and G2) made the choice to emphasize one specific aspect of the game content (i.e. interface, narrations, dialogues, graphic design) during a playing session. The participants were given specific assignments, which resulted in a procedure similar to the task-oriented approach taken in traditional Human-Computer Interaction research (Dumas & Redish, 1999). With two teams (G3 and G5) the players were encouraged to adopt a personal playing style, and to make their own choices throughout the game – a procedure that is in accordance with the characteristics of digital game play as an experience involving exploration and emergence (Juul, 2005). Finally, in the context of one game (G4), a hybrid approach was taken, where the participants were given specific assignments in addition to being given the freedom to follow their intuition and make personal choices at certain points – a procedure resulting in a significantly longer duration of each evaluation session. In the results section, the benefits and drawbacks of each procedure will be discussed, in relation to the context of play generated in each game.

Results: challenges and responses

Evaluating collective experiences

A first challenge was met during the evaluation of collaborative games that involve multiple players (G2 and G5). Because the evaluations took place during school hours, many participants perceived the sessions as moments of distraction in between working time. One participant noted: ‘Cool, we get to play games instead of attending classes!’ Many other participants made similar remarks. While the use of the current procedure involves the benefit of evaluating educational games in an educational setting, researchers should watch to it that the evaluation process remains focused upon educational aspects, and that the participants are not distracted from the main focus of the game. A few user experience tests resulted in very low degrees of information. Based upon other experience tests that yielded more significant results, we observed that, in order for a playability test to proceed properly, at least one
evaluator is needed to guide each separate participant. In addition, at least two more evaluators are needed in order to make field notes that cover the wide range of activity taking place in a multiplayer game. Specifically, the use of audio and video recordings of evaluation sessions generated contextual data that allowed the evaluators to frame the user responses more accurately. Finally, during a few experience tests, the teachers of the course were included in the evaluation process. This resulted in a stronger focus of the participants on the main objectives of the games, and in richer degrees of participant responses.

Similarly, the use of physical interfaces (G1 and G5), and the use of competition elements (G1, G2 and G5) had a similar effect on the participants’ attitudes towards the testing procedure. Respondents were easily distracted from the educational content of the game, and directed their attention mainly towards elements of play. Consequently, while this enabled us to generate a good understanding of the ‘fun’ dimension of the game play, often only moderate information was received on the educational content of the games. We concluded, once again based upon experience tests that proceeded more successfully, that the use of scenarios in order to structure the evaluation procedure, may help significantly in overcoming this difficulty. Moreover we observed that the use of multiple evaluators to simultaneously evaluate the players’ behavior may result in a stronger research outcome.

Balancing between education and fun

While the think out loud method, wherein participants are demanded to comment on their perceptions of game elements, as well as on the strategic decisions made within the game, helped enhance the quality of the field notes that were taken, a number of limitations were encountered. Especially in the context of games that thrive on the creation of a flow experience and on a rapid succession of game stimuli and player responses (most notably, G1), players found it difficult to continue providing verbal cues. At moments, this made it very difficult to assess the players’ relationship to the educational contents of the game. This difficulty was partly accounted for in the in-depth interviews that were taken afterward with the respondents.

A comparable challenge was met with games that are narration-driven (G3) and games that involve puzzle solving (G4). Higher degrees of participant immersion
in the games made it difficult to evaluate the balance between play elements and educational elements. Very often, the coders had obtained a good understanding of whether a game was fun or not, but only a moderate understanding of the participants’ liking of the educational aspects. A scenario-based or walkthrough-based approached is advised, where the attention of the players is alternately directed towards play elements and instructive elements.

_providing feedback during game play_

In usability research the guideline is formulated that the evaluator should interfere as little as possible in the progress of the user through an interactive text (Dumas & Redish, 1999). The main rationale behind this guideline is that the error rate while performing an assignment is a good indicator of the user-friendliness of a procedural system. While educational games can in essence be considered procedural systems (Bogost, 2007), nevertheless a number of difficulties arose from a strict application of this method. In the context of games where repeated play is considered a necessary condition for a learning effect to take place (G1, G2 and G4), the developers may not dispose of the time to let the participants play for several consecutive sessions, nor can they expect the participants to have acquired the level of expert from the first session on. Evaluating these games often proved a time-consuming assignment. Only after a relatively long period of having the participants explore the game rules, interesting insights could be distilled with respect to the game’s replayability. In these cases, it proved a useful strategy to have the evaluator participate more actively in the evaluation process, and to provide detailed feedback at moments when the player encountered difficulties proceeding through the game. In addition, the manipulation of the level of difficulty by the developers proved a very useful technique to attune the evaluation test to the level of experience of the participants.

Balancing between structured play and virtual experience

Although thus far we have mainly emphasized the benefits of performing assignment-based user tests, nevertheless in a few cases a free play-based approach instigated rich and significant evaluation results. With respect to every game, the evaluation procedures had enabled the developers to discover playing styles and player
responses that had not been accounted for on beforehand. Only one developer (G4) had made the choice to adopt a hybrid method, resulting in a longer duration of the user tests, but also in rich, varied and relevant research results. Based upon the outcome of this study, the conclusion is drawn that a well-considered measurement should be made on beforehand of the balance between different evaluation emphases, as well as between the duration of an evaluation session and the number of evaluations that are made.

In table 3, the most important findings are summarized, providing the researcher with guidelines that should be useful in making this measurement. As

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Response</th>
</tr>
</thead>
</table>
| Enhanced group dynamics may result in loss of focus with participants    | • evaluator/participant ratio should be larger than 1/1  
  • strict assignment of evaluator roles is advised  
  • scenario-based evaluations work better than free evaluations  
  • involving teachers in evaluation process enhances quality of testing |
| Use of physical interfacing in combination with competition elements may result in distraction of participants | • scenario-based evaluations work better than free evaluations  
  • number of evaluations should be larger than 5  
  • evaluation methods should be complemented with player interviews afterward or with field notes and video/audio recordings of playing session. |
| Participants tend to consider game evaluations a moment of distraction in between classes. | • scenario-based evaluations work better than free evaluations  
  • specific testing of education-related aspects is advised  
  • quantitative post-hoc questionnaire helps complement qualitative insights |
| Providing evaluator feedback during evaluations may result in loss of focus with participants | • number of evaluations should be larger than 5  
  • combination of free evaluations and scenario-based evaluations provides significant results |
| Both free evaluations and scenario-based evaluations result in significant insights | • longer duration of evaluation sessions is advised  
  • complementing observations with structured post-hoc interviews is advised |
became apparent throughout the investigation, most of these aspects are dependent of the types of game that are being tested, as well as on the number of evaluators that a team possesses of. Nevertheless, with respect to the number of participants, the participant-evaluator relationship, and the structure of a user test, a number of general conclusions can be drawn.

Conclusions

While quantitative methods have recently proven useful in order to analyze player responses to digital game content, qualitative methods may serve as an important addition and complement to the heuristic and psychophysiological measures that have become extensively used and documented. Especially in the context of educational games, where aspects such as collaborative work, contextual learning and motivation have been strongly emphasized, researchers may benefit from the development of a qualitative approach on playability research. With this paper we wish to draw attention to the possibilities and challenges included in performing qualitative playability evaluations. Whereas the literature review exposed a number of benefits of this approach, in the results section we highlighted a number of challenges and proposed solutions, based upon a systematic analysis of the evaluation procedures that were applied in the development of five educational micro-games. Although some of the proposed solutions are related to the properties of specific games that have been tested, the analysis enabled us to formulate a number of general guidelines with respect to the methodology of qualitative playability research. More specific, conclusions are drawn in five areas: observer/participant ratio; role assigned to the evaluators; duration of the evaluation sessions; number of tests; and type of evaluation.

Observer / participant ratio

In the context of all games that were developed and tested, we observed that at least one observer/evaluator is needed for every participant in the game play experience. When multiplayer games are tested, using a observer/participant ratio that is larger than 1:1 is highly recommended, because making field notes may become a highly
complex assignment, and because specific characteristics of the classroom context demand an enhanced coordinating effort of the researchers.

Evaluator roles
Based upon this study the guideline is formulated that different researcher roles should be assigned to the members of an evaluation team. Taking field notes may become a challenging assignment, given the limitations of the ‘think out loud’ method in the context of educational game testing. We advise using audio and/or video recordings in order to assess player responses that are overlooked during the playing session. In addition, evaluating games in a classroom context may result in a reduced focus with the participants, and as such in an enhanced need for structuring a user test.

Use of scenarios / cognitive walkthroughs
Although the educational potential of a digital game resides partly in properties such as freedom of action or virtual exploration, this study unveiled a number of benefits associated with performing player evaluations that are assignment-based. Most importantly, in a classroom context the use of scenarios and cognitive walkthroughs may help maintain the focus of a session on the educational aspects of the games being tested.

Duration / Number of tests
Finally, the researcher should aim for a balance between well-timed player tests, and sufficient amounts of player tests being performed, since performing player tests in a classroom context is often restricted by time limitations. Our analysis unveiled that the most significant results were obtained with evaluation sessions that had a duration of 30 minutes or longer. Especially in the case of games that aim for a high replayability, this duration should be considered a strict minimum. On the other hand, unlike is the case with traditional usability procedures, testing an instructive game with only 5 participants did not always prove sufficient. As a consequence of the variety in player activity, performing player evaluations should be considered a costly, but highly useful phase in the development cycle of an educational game.
References


Critic-Proofing: Robust Validation Through Data-Mining

Ian J. Livingston*, Lennart E. Nacke & Regan L. Mandryk

Department of Computer Science University of Saskatchewan Saskatoon, SK, Canada

Abstract

Critic-proofing is a modified heuristic evaluation technique, specifically designed to provide a fine-grained, prioritized list of heuristic violations. The critic-proofing technique weights the severity of a problem based on the frequency that similar problems are found in similar games. The severity ratings are calculated using data collected from game reviews, and the severity assigned to a problem during the heuristic evaluation process. However, heuristic techniques have had limited adoption within the video game industry. One reason for this is the perceived lack of validity and robustness of game specific heuristic principles. In this paper, we introduce and outline a new data-mining project designed to validate game-specific heuristic techniques, especially the critic-proofing technique by using the popular game-review aggregation website Metacritic.

Introduction

In the gaming industry - especially in casual game development - the need for a quick turnaround from playtests often does not allow for the implementation of sophisticated and advanced biometric measurement techniques (e.g., Mandryk, Atkins, & Inkpen, 2006; Nacke, Grimshaw, & Lindley, 2010), because they require personnel training and expensive tools. A possible alternative for getting reliable and robust data in playtesting is the utilization of easy-deploy inspection methods, some of which could be considered discount evaluation methods because of their low cost.

*ian.livingston@usask.ca
and short time requirements, making them ideal for application in the casual games industry. One example of a discount evaluation method is heuristic evaluation, where expert evaluators examine system actions and features that violate heuristic principles. After the inspection has been completed, a severity rating is assigned to each violation. The evaluators’ ratings are then combined. Jakob Nielsen developed the most well known heuristic principles (Nielsen, 2005). His heuristics were designed for software usability, and since have been shown to be both valid and effective (Nielsen, 1994). In recent years, many heuristics collections have been developed for game specific purposes. Generally, these game heuristics are designed to evaluate usability, playability, or both (Federoff, 2002; Korhonen & Koivisto, 2007; Korhonen & Koivisto, 2006; Pinelle, Wong, & Stach, 2008a; Pinelle, Wong, Stach, & Gutwin, 2009). In some cases, attempts have been made to combine different collections of heuristic principles into more comprehensive collections of heuristics (Desurvire, Caplan, & Toth, 2004; Koeffel et al., 2010). However, as with many advanced evaluation techniques the industry uptake of these heuristics has been lackluster.

Recently, Livingston, Mandryk, & Stanley (2010) published work that used video game review data (Pinelle, Wong, & Stach, 2008b) to develop a new heuristic evaluation technique called Critic-Proofing (CP); an evaluation technique that assigns a rating to heuristic violations (called a genre rating) based on the game’s genre and the heuristic being violated. The genre ratings are combined with severity ratings assigned during the heuristic evaluation to form a prioritized list of heuristic violations.

The technique was tested in a case study, where it was used to evaluate the game usability of Capsized (AlienTrap Software, 2010), which is currently under development. Feedback was collected from the development team, who found the CP results to be much more valuable than those from traditional heuristic evaluations. In addition, the easy deployment and quick turnaround time of these heuristics would make them ideal for evaluating casual games, since shorter development times are common there. What CP currently lacks is a larger database of reviews upon which CP weightings can be based, since it currently uses only reviews from GameSpot. The natural next step to improve CP would be to use a larger aggregation of reviews. One website dedicated to aggregating and accumulating reviews for games is the website called metacritic (see Figure 1).
Metacritic is a review aggregation website that collates articles and scores of games, movies, TV shows, DVDs, and music. The scoring mechanism implemented by the website converts each review into a percentage before taking a weighted average and then listing different numbers of reviews. For reviews with no explicit scores, metacritic staff manually evaluates the tone of the review before assigning a subjectively fitting grade. Weighting is applied to reviews depending on the public
impact of the review source, so that reviews from major periodicals may have a
greater effect on the average than niche ones. The weight numbers are however
not publicly available (“About Metascores,” n.d.).

Game publishers and developers are currently using metacritic in a variety
of different ways. Some are using the system in contract negations as a metric
for bonus distribution; others are using it as a management metric - in some cases
using metascores to influence developer royalties, or measure of desired quality
for the final game product (“Is Metacritic Damaging the Games Industry? | Edge
Magazine,” n.d.).

In this paper, we will discuss our ideas for using the review base of metacritic
as the foundation for a tool that validates CP and other game heuristic methods based
on this large review database.

Requirements Analysis

Before we start the development of our game heuristic evaluation tool, we need to
consider that it needs to meet certain requirements of a low-cost, easy-to-deploy, and
fast evaluation tool for the game industry. For any type of evaluation technique to be
valuable in the game industry the technique must meet five requirements based on our
past experience in contractual work for game companies:

1. It must have demonstrable value.
2. It must be cost-effective.
3. It must be valid.
4. It must be easy to integrate into the current pipeline.
5. It must have clear and interpretable results.

The heuristics developed by Nielsen (Nielsen, 2005) meet all of these requirements.
However, these heuristics were developed for the evaluation of traditional desktop
user interfaces. Video games are more complex (e.g., they value process over outcome),
which means that Nielsen’s heuristics can only be directly applied to game features
that are similar to desktop interface features.
Game Heuristics

Some heuristics have, however, been developed specifically for games. These heuristics fall into one of three categories: (1) usability, (2) playability, or (3) a combination of both. In contrast to Nielsen’s heuristics, which were developed for user interface analysis, game usability heuristics apply to the usability of the gameplay, such as camera control or artificial intelligence issues, which are fundamentally different problems than those found in traditional software development. Playability heuristics apply to the dramatic and formal elements of games. Playability heuristics might apply to problems with the game narrative or to player fatigue.

Federoff (Federoff, 2002) did pioneering research on game heuristics. Through observations and conversations at a San Francisco game company, and a comprehensive review of literature, Federoff was able to compile a list of heuristics,

Figure 2. The technique was used to evaluate Capsized (AlienTrap Software, 2010), a 2D platform shooter game
which focus on three game areas: (1) interface, (2) mechanics, and (3) play. Federoff’s heuristics are quite broad, while we are specifically interested in problems pertaining to game usability.

In 2004, Desurvire et al. (2004) developed a set of heuristics called Heuristic Evaluation for Playability (HEP). Although similar to Federoff’s heuristics, Desurvire provided a simpler description and organization of the heuristics. Recently, Desurvire has published a newer version of the HEP called PLAY (Heuristics of Playability) (Desurvire & Wiberg, 2009). The PLAY heuristics attempt to make the underlying principles more generalizable by evaluating the principles across three different genres.

Other forms of heuristics have been developed to fill different needs of game evaluation: Korhonen et al. created heuristics for mobile multiplayer games (Korhonen & Koivisto, 2007); Pinelle et al. generalized this idea with a set of heuristics for all multiplayer games (Pinelle et al., 2009). Many of these heuristics overlap and share similarities, and some attempts have been made to compare techniques. Korhonen, Paavilainen, & Saarenpää (2009) compared two different sets of playability heuristics – HEP and their own mobile games heuristics (Korhonen & Koivisto, 2006) – to examine the relative strengths and weaknesses of each approach. Koeffel et al. (2010) attempted to create a comprehensive list of heuristics from those found in literature, and to assess the collection’s effectiveness using a comparison to video game reviews.

Pinelle et al.’s (2008a) heuristics list is specific and short—there are ten principles—making the evaluations simple and quick to perform. In addition to this, prior work has been extended in the exploration of usability specific to game genres (Pinelle et al., 2008b). These heuristics are different from others because they were developed through the analysis of game reviews found on the game review website GameSpot. According to Larsen (2008), it is reasonable to consider game reviewers as unscientific user experience evaluators with expert domain knowledge. Thus, the approach of Pinelle et al. already provides an implicit measure of validation because their heuristics are based on user reports.

However, all of these heuristic principles - with the exception of Nielsen’s - are not in use within the video game industry. This is primarily because they do not
fulfill the five basic requirements outlined at the start of this section. Most importantly, game specific heuristics have not been shown to be valid, clear and interpretable, or demonstrably valuable to game developers. Conversely, Nielsen’s heuristics have been shown to be valid in other software fields, the results are clear and interpretable, and the value of improving user interface elements is obvious. By providing means to validate review-based heuristic techniques, such as CP, we are able to make a first step to scientifically support these game evaluation techniques, making them more robust. Once the support for CP is established, interface problems that have historically caught critics’ attention can be identified early in the development process. This is especially critical for casual games, since a game’s success depends largely on how easy its interface is understood and whether it allows for easy-in/easy-out gameplay cycles, which are most common in this field.

*Game Critics*

The voice of a professional critic is powerful. A single influential critic can have a significant effect on any product, influencing consumers’ purchasing decisions. Larsen (2008) explored how a critic’s review is in essence an unscientific user experience evaluation. Larsen identifies that game reviewers generally provide both a qualitative (textual review) and quantitative (score) component. Some reviewers evaluate different aspects of a game separately. In many cases, these aspects are comparable to hedonic and pragmatic aspects of user experience research.

Koeffel et al. (2010) also offers support for the use of the reviews written by game critics. In their recent study, they compare results of heuristic evaluations to the average critic scores of games, and suggest that a similar trend between review scores and the number and severity of usability problems exists.

The term critic-proofing has recently been used to describe the technique used at BioWare in MassEffect 2. In an interview with Gamasutra (Nutt, 2010), producer Adrien Cho described how the development team mapped critic and player feedback from the first game onto the design objectives for the sequel.
Game reviews provide us with an excellent way of dealing with the limitations of current game heuristics. Unfortunately the current review-based heuristics (Pinelle et al., 2008a) were limited and biased in their scope, focusing on only 6 game genres and 108 reviews from GameSpot (Pinelle et al., 2008b). Also, the entire evaluation was conducted manually, with researchers visually searching the reviews for criticisms of usability problems. Since metacritic provides an aggregation of a large sample of game reviews, it is perfectly suited to provide a basis for improving and extending review-based heuristics. metacritic has also gained traction in the industry as a metric used by publishers to determine the success of a title and the potential return from a given game. We plan to use metacritic similarly to how Pinelle used GameSpot, to a) validate existing heuristics, b) use a tool already used by the games industry to provide understandable evaluation results, and c) demonstrate the potential value of using game specific evaluations.

Implementation Strategy

Examining game reviews manually is impractical on a larger scale. Instead, the process should be automated as much as possible so a much larger review sample set can be utilized. To develop a more automated process we have developed a three-step implementation strategy (Figure 3).

Web Crawler Functionality

Before any analysis can begin, review data must first be collected. This entails two types of data in addition to the more specific textual qualitative reviews and game scores (as well as metascores). The web crawler will collect complete game reviews from the same websites used by Metacritic in the calculation of their ‘metascores’. The web crawler is an automated system that will run continuously, collecting new data as it’s posted and periodically rescanning to ensure that the data collected is up to date.

Data Collection Strategy

Metacritic is a rich source of data. In addition to the actual text reviews, we will also collect: platform, genre, release date, publisher, developer, number of players, rating,
number of reviews, reviewer’s name, metascore, and any relevant or potentially important data from the specific game review sites we mine. We will store this data in a database for further analysis.

There exist two primary challenges that must be addressed in this data collection strategy. First, data must be collected from each review site that metacritic provides access too. However, each site provides different information formatted in different ways. The web crawler will need to be customized to ensure that it is capable of sorting data correctly when presented in different formats. This challenge is exacerbated because metacritic aggregates across more than one hundred game reviews. Second, metacritic uses some publications that are only available in print.
mediums, restricted to subscribers, or in a language other than English. The web crawler cannot mine these publications; therefore we will have to omit them from our evaluation database. Fortunately, the number of restricted or print publications is small in comparison to those that remain, so a substantial number of reviews are still accessible.

Analysis Strategy
Perhaps the most challenging aspect, and one of the primary goals, for this project is the development of an automated analysis tool that will mine information from the text of game reviews. The analysis strategy will validate game specific heuristics by fulfilling the five requirements that we have outlined above. Since game reviews are essentially untrained user experience reports, we can use the information from them to validate or reject game heuristics.

Analysis of the review data will be performed in a number of different ways. Initially the textual data will be analyzed using text analysis tools such as LIWC (Pennebaker, Chung, Ireland, Gonzales, & Booth, n.d.) and ANEW (Bradley & Lang, 1999). The results from LIWC can be examined to determine how words are used within a specific text. The use of heuristic-specific words could be used to determine the frequency of a specific heuristic problem, for example artificial intelligence issues. ANEW is a similar tool to LIWC except focusing specifically on the emotional content of words.

Discussion
Casual games are known to have smaller development budgets and quicker turnaround times. Thus, a low-cost and simple game usability evaluation approach is needed. Many of the expert reviews and heuristic techniques in use for evaluating the playability of games provide such a quick turnaround, however not always at small budgets. In addition, the major challenge of current playability evaluation approaches is that they are usually only based on evaluator perception and lab-constrained user studies. Many of these approaches have no quantitative backing by structured scientific studies using large numbers of evaluators. However, to make
these approaches widely accepted in the games industry, a wide range of globally situated evaluators can provide the credibility needed for the acceptance and deployment of these techniques in the game industry. Since Metacritic is already an accepted source of criticism in the game industry, using the reviews that build the foundation of Metacritic to validate the CP game evaluation technique will provide an unique opportunity to establish this technique for quick and easy game interface and usability evaluation. Having such a technique available that is half-automated or assisting in evaluating games, the game quality of casual games will be easy to improve.

Conclusion

In this paper, we have presented preliminary work into a novel game evaluation technique we have dubbed critic-proofing and a possible evaluation technique to improve its validity and foster application in the game industry. Previous work (Livingston et al., 2010; Pinelle et al., 2008a; Pinelle et al., 2008b) in this area has relied on a small sample of game reviews to validate this method. If CP and game heuristics in general are to be applied in a game industry context, they need to have established validity, ideally based on a source that is trusted in the industry. Thus, they are likely to receive large-scale adoption by the game industry. In summary, to validate previous game-specific heuristics and make them relevant to the game industry, we are developing an automated game review data collection and analysis tool that can be used to validate existing game-specific heuristics and will be used to further strengthen existing CP techniques.

References


Let's Start Playing Games!
How games can be less about complying and more about playing

Menno Deen* & Ben A.M. Schouten
Fontys University of Applied Sciences

Abstract
The input of players becomes increasingly valuable to game designers. In the past, players mainly negotiated with game designers in sale-numbers, coin-drops and through critical acclaimed game magazines. Today’s game designers tend to communicate more directly with gamers through internet forums, beta-testing and data mining of gameplay. The increased negotiations between game players and game designers may account for the increased focus on user-generated games, making games more playful.

Introduction
Today’s playful activities in games mainly concern the restructuring of game objects, goals and environments. Hereby game designers remain the sole authors of the game rules. Even more, gamerules remain enclosed in difficult to breach systems. Due to a lack of programming skills or the inability of modders and hackers to change these game mechanics, games are historically related to formal and rigid rule sets. While play concerns rich social negotiations, elaborate restructuring and creative reconstruction of mechanics, games seldom offer these playful activities. Traditionally game designers are perceived as the sole author of game mechanics. We urge game designers to become less an author, and more a fellow gamer, so we can start playing games.

*menno.deen@fontys.nl
‘Bang! You’re dead!’
‘No, you missed me.’
‘I did not!’
‘You sure did!’
‘Yes I did... kill you! Now play dead for 20 seconds!’

Two children are playing soldier, they pretend to fight a war. They change the rules every minute, constructing new rules, new goals, new weapons, or new worlds. This is play in one of its purest forms. A simple wooden branch becomes a deadly Desert-5 pistol. Seconds later the branch transforms into a devastating M79 grenade launcher by holstering it on one’s shoulder. By breaking the branch in two, the children run around yielding two pistols: Lara Croft\(^\text{VII}\) style. One of the children gets bored, and changes the game. The branches become two razor-sharp rai. The child, still bearing the long stick turns his branch into a bö staff.

‘Now we are Rafael and Donatello!’
‘No better: Kilik and Taki from Soulcalibur!’
‘I don’t wanna be girl! I wanna be a Turtle.’

The children continuously negotiate about the objects, rules and goals of their game. They change the meaning of objects, the rules, and by breaking the branches, they actually reconstruct the objects themselves. This way, new activities and new play experiences emerge from continuously restructuring the relations between rules, goals, objects and their environment.

This is why we like to play: we enjoy experiencing the new, and/or unexpected. What’s more, we enjoy the continuous play of social negotiations. As videogames are considered to be played, we would expect the same dynamically social play, and enjoyment as we find in children’s play. However it feels like most video gamers are complying to, instead of playing with the game rules, goals, objects and environments.

For example, in Uncharted 2: Amongst Thieves (Naughty Dog, 2009), gamers can choose between a Desert-5 and M79, similar to children imagining the wooden

\(^\text{VII}\) Lara Croft is the main character of the Tomb Raider (Core Design, 1995) series.
branch to be a gun. However, gamers cannot wield Drake’s grenade launcher as a Donatello’s bō staff. Nor will the game allow for close-combat similar to the swordplay of *Soulcalibur* (Namco, 2008). In contrast to the children’s play, the gamer cannot restructure the game’s environment, objects, goals and rules to construct new game mechanics. The construction of something new by restructuring the existing seems inherent to play. Games, however, seldom offer players these possibilities. Gamers merely comply with, or deviate from preconfigured game mechanics. Arguably, the game designers are the only ones whom are actually playing. Nevertheless, recent innovations in game design seem to turn the tides.

**Game designers as authors**

Traditionally, the work of game designers includes a wide area of creative work. They code, design and shape visualizations of the videogame. Today’s game industry brings forth a fragmentized workspace in which programmers, asset developers, graphical artists, audio designers, writers, managers and game designers work collaboratively on one game. All team members contribute to the gameplay from their respective specialism. But frankly, it is the game designer whom remains sole author of the game’s rules and goals. Designers take the final decision about the toughness of an end-boss, whether or not a game contains a warp-zone, or if a special attack requires a three- or four-button combination.

A game designer is an artist: iteratively creating the game’s rules, goals, objects and environments, to construct the optimal game-flow. As a direct consequence of the authoring behavior, gamers can either comply with the game mechanics or deviate (cheat) from them. Within strongly authored games, gamers cannot restructure essential game mechanics, let alone construct something new or personal. The gamers’ restructuring activities mainly concern the dynamics and aesthetics of the game.

---

*MDA model* describes Mechanics, Dynamics and Aesthetics consequently as: 

- **Mechanics** describes the particular components of the game, at the level of data representation and algorithms. 
- **Dynamics** describes the run-time behavior of the mechanics acting on player inputs and each others outputs over time. 
- **Aesthetics** describes the desirable emotional responses evoked in the player, when she interacts with the game system. (Hunicke, LeBlanc, & Zubek, 2004)
This becomes particularly clear in online role-play. Role players change the meaning of game objects through negotiations within the online community or by altering (player constructed) narratives. In World of Warcraft (Blizzard Entertainment, 2004), an Aurora Robe signifies +54 points on armor, +5 stamina, and +15 spirit\textsuperscript{IX}; however, a role-player could use it as a fancy white dress for occasional role-playing parties. Gamers can construct new dynamics and aesthetics, but they can seldom restructure the game’s mechanics like the playing children as mentioned above. Authored games, are games as Salen & Zimmerman (2003) define them: artificial conflicts, based on rules, with quantifiable outcomes. In their perspective, the designers wield the scepter. They decide what can, and what cannot be done.

Nevertheless, the game mechanics of MMORPGs\textsuperscript{X}, Social Games\textsuperscript{XI} and User-Generated-Content Games seem less authored than the traditional retail titles are. In these games, gamers and designers increasingly negotiate about the game mechanics. These negotiations enable the restructuring of formal rules and regulations, constructing new mechanics and alternative playing styles. However, most re(con)structing practices are still authored by the game designer. Gamers have a say, but are not empowered: we argue that games can become more playful if gamers are presented with a higher degree of re(con)structing activities.

We will discuss a taxonomy from the most authoring practices (i.e. game creation) to the most re(con)structing (i.e. playing) practices of today’s videogamescape, to clarify the aforementioned trend. Firstly, this is done from a game designer perspective by discussing patching, beta-testing, tweaking, and editing. Secondly modding, adding, hacking and cheating from a player’s point of view will be discussed. In the latter practices, gamers have a greater influence on the game mechanics, which can be considered to be more playful.

\textsuperscript{IX} See Thottbot for more information on the Aurora Robe: http://thottbot.com/i6415
\textsuperscript{X} MMORPGs are Massively Multiplayer Online Role Playing Games. These games are played online with (or against) other players around the world. EverQuest, WarhammerOnline, Lord of the Rings Online, and Star Trek Online are such games.
\textsuperscript{XI} Social games are typically found on online social network sites like FaceBook, or Hyves. These games often transform the quantity of friends connections into an ingame asset / commodity. FarmVille, MafiaWars, RestaurantWorld, HappyIsland are such games.
Authoring Practices

When we think of authoring practices in games the initial creation of a game comes to mind. In this view, patching can be described as an authoring practice of a lesser degree. It describes the process of making alterations to existing games. For its greater part, patching is made possible by the connectivity of the internet. Blizzard Entertainment releases monthly patches for the MMORPG *World of Warcraft*, in which the differences between classes are changed, bugs are fixed and some significant rules are altered. An example of a significant change in game mechanics is the drop-rate alteration in *World of Warcraft*. In the MMORPG, gamers loot items from defeated foes. Which item is looted depends on the foe and on loot-probability\textsuperscript{xii}. Players, however, expected the

\textsuperscript{xii} Common items have a 15% change to be looted.
items would drop more easily after repetitive killings. Clearly in practice this does not happen. Some players kill numerous foes without looting a particular item, while other players loot the precious item after their first fight. In response to players’ looting-expectations, Blizzard created a system in which the drop-rate increased every time a foe was killed. So, the more foes killed, the higher the drop-rate, the more likelihood of looting an item. By restructuring game mechanics in monthly patches, mechanics become less rigid.

Social Games, like *FarmVille* (Zynga, 2009-2010) or *Happy Island* (CrowdStar, 2010), build on this iterative design process. Apparently, these Social Games never leave beta phase (i.e. the development stage of a production cycle). Instead, Zynga changes the mechanics almost real-time to satisfy immediate needs and actions of the player’s community. Blizzard, Zynga and CrowdStar do not only change the rules, they add new objects, administer environmental changes, and even present gamers with new goals to accomplish. Social Game designers carefully analyze the players’ actions and participate on game-forums to accommodate to the player’s expectations. While the time span is shortened in comparison to patching, it is still the game designer who authors the changes instead of the players themselves.

The same is true for tweaking. Tweaking is the possibility to make small changes to game mechanics, which may significantly influence the gameplay. An example of tweaking can be found in *Mass Effect* (BioWare, 2007). Gamers can choose to improve the abilities of the game character, which in turn will change the way that gamers comply to the game mechanics. Also players can choose to act polite or ruthless in conversations and circumstances. This changes the narrative of the game and its sequel. Tweaking parameters are common to many RPG-like games. *Super Smash Brothers Brawl* (Sora, 2008), however, offers other alterable parameters. Amongst others, gamers can play with the amount of usable power-ups, add multiple characters, limit play opportunities per character or change the time limit. Although gamers are given more freedom to tweak specific rules in the game, the game designer authors the parameters.

The game *WarioWare D.I.Y.* (Nintend SPD Group No.1 & Intelligent Systems, 2009) offers players the opportunity to create their own *WarioWare*-minigames, without any programming knowledge whatsoever. These games are based on tapping
or flicking the *Nintendo DS stylus* on a specific point, or in a particular way, within a certain time limit. Players can change an extensive number of parameters (graphics, time, number of tabs, flick gestures, sounds, etc). Although Nintendo’s advertisements recruit gamers by saying ‘I want YOU to create games’, gamers merely edit typical *WarioWare* games. Gamers are not actually constructing new game mechanics, the gameplay remains essentially the same as other *WarioWare* titles. *WarioWare D.I.Y.* seems to equal the restructuring qualities of ‘level editors’.

**Level editors** are available for an extensive number of games. They present players with the opportunity to create their own game levels\(^{\text{XIII}}\), racing tracks\(^{\text{XIV}}\), battling stages\(^{\text{XV}}\), cars\(^{\text{XVI}}\) or characters\(^{\text{XVII}}\). Still, the addition of a level editor may not offer players the possibility to restructure the actual game mechanics. Most of the time, editors concern environments and objects, not the rules and goals of a game. Moreover, editors are less about restructuring, but more about constructing instead. A level editor is to a game, as a text editor is to an ebook. By writing fanfiction, readers are not playing with the initial text, but readers are playing with the theme, characters, timeline or environment of the original story to create something new. By reading a book, we cannot rearrange sections or sentences to change its meaning. That is because a printed (Epub) text in itself is unchangeable. Game rules do not have to. Games, especially videogames, can do so much more if we give players the opportunity to reconstruct both the sociocultural values associated with the gamrules (like role players\(^{\text{XVIII}}\) and cheaters do), as the game mechanic itself.

**The game players’ perspective**

Without an editor, gamers may only comply or deviate from pre-configured mechanics. Nevertheless, some gamers change more than the aesthetics and dynamics alone.

---

\(^{\text{XIII}}\) *LodeRunner* (Smith, 1983), StarCraft (Blizzard Entertainment, 1998)

\(^{\text{XIV}}\) *Stunts* (Distinctive Software Inc., 1990)

\(^{\text{XV}}\) *SuperSmashBrothers Brawl* (Sora, 2008)

\(^{\text{XVI}}\) *LEGO Racers* (High Voltage Software, 1998)

\(^{\text{XVII}}\) *World of Warcraft* (Blizzard Entertainment, 2004)

\(^{\text{XVIII}}\) For an extensive account of the restructuring of sociocultural rules by role-players, read Taylor (Taylor, 2006). In turn, Consalvo (Consalvo, 2009) presents a extensive description of cheater-culture.
Modders and hackers actually restructure existing games and reconstruct them to their personal liking.

‘The word ‘hacker’ describes in computer culture a person with sophisticated programming skills developing creative solutions to complex and challenging problems’ (Scheäfer, 2008, p. 29). Many hackers develop addons or cheat programs for games, which are a debated issue within the gamers’ community. For example in World of Warcraft, The AVR (Augmented Virtual Reality) mod allows players to visualize strategies by drawing in 3D space. Blizzard did not allow this mod, and broke the ability for the AVR to continue functioning in their latest patch (3.3.5). The game developer reported that addons are never intended to interact with the game world itself. AVR and the act of visualizing strategy within the game world simply goes beyond what Blizzard is willing to allow (Boubouille, 2010). Blizzard upholds a rather strict policy when it comes to addons and other mods. According to Blizzard and many other game developers, addons or mod may significantly destroy gameplay. In other words, addons or mods may liberate gamers too much from the authoring gameplay that was preconfigured by the game designer. We wonder if this anxiety is justified.

Not all game developers police the modding community as Blizzard does. Valve Corporation, for example, seemed more willing to have modders change their game by opening Half-Life’s (Valve Corporation, 1998) Source-Engine to the public. In response, Le & Cliffe created a new game, based on its predecessor Half-Life, but with new characters and objects, new environments, and above all new rules and goals. Half-Life: Counter Strike (Le & Cliffe, 1999) may be the most stunning example of gamers utilizing their programming or artistic skills to change the game’s environment, objects, rules and goals, to actually reconstruct the game mechanics and enrich the playful experience in accordance to their personal likings. Modders do not only play the games, they play with the games (Sihvonen, 2009).

Facilitating play

Creative outlets, similar to the modding community, are recognized in the game Little Big Planet (Media Molecule, 2008). In contrast to WarioWario D.I.Y., Little Big
Planet players are not confined to tap- and flick minigames. Instead, players are encouraged to use the objects in Little Big Planet to create their own rules, goals and environments. Players can even upload personal imagery into the game, and alter visualization and characteristics of objects and environments. Little Big Planet offers players the most re(con)structing possibilities to date without the need to learn a programming language. Its flourishing community, with more than 2.3 million uploaded levels, high praise from both gamers and game critics, and the sale of three million copies so far, prove that the game has a huge appeal to the gaming community.

Still, Little Big Planet remains a level-editor for the majority of gamers. It is extremely difficult to create other games than platformers\textsuperscript{XIX}. Its successor, Little Big Planet 2 (Media Molecule, 2010), may change this. By adding building blocks with simplified artificial intelligence, Media Molecule makes it easier to create new games by using the Little Big Planet 2 tools. Interestingly, the game designers of Little Big Planet 2 claim that they themselves have no other tools at their disposal than those available in the game editor. Even more, Media Molecule promised to strengthen social negotiations in the game, by improving interface design, the visualization of user-generated games, and enhancing communication services.

This may be one of the most significant changes in game design today. Media Molecule is promising an environment in which the play experience is as rich, diverse and communicative as in children’s play. Game designers and game players may become equals in their ‘programming abilities’, much alike children are equals in their ability to change play mechanics. In Little Big Planet 2 the game designer ceases to be the sole author of gameplay. Thanks to negotiation tools like online leader boards, popularity voting systems, game-objects sharing, ingame chat and out-of-game forums, game designers no longer wield the scepter, deciding what is possible and what is not. Players and designers become peers as they are presented with the same toolset and digital mechanics.

\textsuperscript{XIX} Platformers are games like Super Mario Bros. in which gamers need to make their character jump from one platform to the other.
In theory, games like Little Big Planet 2 can be re(con)structured in an even richer way than a children’s playground and two branches, described in the introduction. One could say, a wooden branch has many functionalities; an interactive virtual object, however, can do so much more, thanks to unending combination of potential functionalities.

Through restructuring practices, game mechanics are changed and new games emerge. No longer are game mechanics rigid and definite. Like play, game rules can be in a constant state of flux and subjective to social negotiations. MMORPGs and Social Gaming already showed that game rules are not unchangeable after all, and that these changes and social negotiations are key to an enjoyable playful experience. Little Big Planet 2 promises to go even further, not only by changing the aesthetics and dynamics, but also by offering players to restructure the mechanics and construct new playful experiences.

Considering the definitions of gaming

Salen & Zimmerman’s definition of games as an artificial conflict, based on rules, with a quantifiable outcome, suggests that the games are all about rigid (unchangeable) rules, challenge and artificial (i.e. not real) experiences. However, in the future starting today, every action, every rule, goal, object, and environment can be debated and changed. Games can no longer be considered ‘formal systems that provide informal experiences’ (Juul, 2005), as the rules are no longer fixed. Therefore, Caillois’ (2001) distinction between ‘ludus’ (explicit and rigid regulations) and ‘paidia’ (implicit and spontaneous regulations) becomes debated, which as such asks for a redefinition of gaming.

First and foremost, playing is not something artificial as Salen & Zimmerman (Salen & Zimmerman, 2003) suggest. Abt incorrectly stated that ‘games do not have direct real-life effects’ (1971, p. 127). Conversely, Huizinga (1951) already pointed out that play is of profound value to our culture and that games exist on the border of culturally defined boundaries. Games, in Huizinga’s opinion, have a tremendous influence on our real-life and vice versa. Sutton-Smith (2001) suggests that play’s definition must be broad, including human and inhuman participants: ‘play is like language: a system of communication and expression’ (2001, p. 129).
In turn, Marinka Copier (2007) describes this system of communication as continuous negotiation of (role) players with socio-cultural network of human and inhuman actors. Underlining the importance of a player community. Inspired by the Actor Network Theory of Bruno Latour, Copier formulates a comprehensive description of (role) play, which does not focus on actors like rules, goals, objects, or environments, but instead describes / investigates the relations between all actors. Earlier Copier (2005) emphasized that a game only exists when it is played. Role-players actively negotiate with the game mechanics, socio-cultural mechanics, and individual-personal ones. From these negotiations a play experience emerges. The play experience and the activities related to these experiences are in a constant state of flux. How players act differs every negotiation, because they themselves are part of the actor network as well. It is in this continuous change that the characteristic of play can be found.

The developmental psychologists Jarvis, Brock & Brown (2008) describe that play ‘emphasizes the restructuring, enrichment and discovery – building on personal experiences and knowledge to create new concepts and experiences’ (Jarvis et.al, 2008, p. 25). Through this restructuring, new forms of activities and meaning are constructed. The restructuring process described by Jarvis, Brock & Brock concerns cognitive psychological changes, and Copier focuses on the cultural changes that emerge from social-negotiations. Both do not specifically address the restructuring of game mechanics themselves. In other words, the iterative design of rules, goals and activities to construct new experiences. This is not surprising, as rules are easily changed in the aforementioned children’s play, while it is increasingly difficult to change rules in a videogame. However, as Little Big Planet, the modding community, MMORPGs, Social Games, and the advance of user generated contend show, the game mechanics of today’s games do change and by doing so, gaming can become more playful.

Conclusion

We advocate to bridge the gap between game designers and gamers. By doing so, gamers can become game designers, and vice versa. We discussed the open construction of rules, objects and contexts in play. People tend to search for ways
to restructure the game mechanics and construct new playful experiences as well. Traditionally this was not fully facilitated, due to the inability of many gamers to code, and due to the authoring attitude of game designers. However, we witness a change in gamers’ abilities and designers’ attitude toward authorship, which make games more playful than they were before. Instead of being an author, game designers become the sidekick of gamers, presenting gamers with alternate points of view, knowledge, skills, or anything else gamers can play with.

By presenting game designers with the same tools as the game players. Gamers and designer can become equals in their restructuring and constructive possibilities. Instead of rewarding gamers for compliant behavior, or punishing cheaters for deviating from preconfigured regulations, game designers could inspire, enrich, and empower gamers, by creating games that are less about complying and more about playing.

Moreover, gaming can no longer be described as a social negotiation with fixed, and formal rules. Gaming is a continuous restructuring of relations between human and inhuman actors, which constructs new experiences, and in turn, new games. This change in paradigm may open new avenues for game design, game engine development, and most importantly, the way we play games.

References


Notes on Contributors

Licia Calvi is a senior lecturer at the NHTV Breda University of Applied Sciences (the Netherlands) and a Senior Researcher at the Centre for Experimental Media Effects Research in the same university. She has been lecturing HCI and Hypertext and Hypermedia in several universities in Europe, such as TCDublin, TU/Eindhoven and University of Parma. Her research interests include interaction design, usability and evaluation of (especially) e-learning systems, rhetoric and argumentation in digital media. She is a member of the editorial board of the International Journal of E-learning, of Social and Humanistic Computing and of Technology Enhanced Learning.

Cédric Courtois is junior researcher and PhD candidate at the research group for Media and ICT (IBBT-MICT) at Ghent University. His current research currently focuses on changing audience practices in a convergent media landscape.

Frederik De Grove is junior researcher and PhD candidate at the research group for Media and ICT at Ghent University (IBBT-MICT). Next to a passion for video games in general, his research focuses on video game experience and serious games.

Yvonne de Kort is assistant professor environmental psychology and co-director of the Game Experience Lab at the EUT. As environmental psychologist, she specializes in the interaction between humans and their socio-physical environment, studying the molar, reciprocal relationships between behavior and experience, and the social, built and natural environment.

Anouk de Regt graduated as a Bachelor student at the Academy for Digital Entertainment of NHTV Breda University of Applied Sciences. Her research interest is in persuasive communication in non-linear media. Her contribution to this issue reflects research she performed within the context of her graduation thesis.
Menno Deen graduated as a (BA) designer at the Utrecht School of Arts and subsequently as (MA) cultural researcher at Utrecht University. His design and research is on virtual citizenship, homosexuality, and the educational potential of video games. Deen wrote his MA thesis on the correspondence between learning styles and playing styles. Subsequently he works as Ranj’s game researcher, validating the learning outcome of a second language learning game called CheckOut!

Since 2009 Deen is PhD candidate at Fontys University of Applied Sciences. He published about serious games and education (OSG), co-authored a report about the attractiveness of casual games (My Child Online Foundation), and co-authored a chapter on online casual games. Deen’s PhD research is on motivations for games and learning. He works on a serious game design method for educational games that may change students’ motivation towards learning for the better.

Brian Gajadhar received his bachelor’s degree in Applied Physics, and he earned his M.Sc. degree in Human Technology Interaction (HTI). Since November 2006, he is a PhD candidate at the HTI group of the Eindhoven University of Technology (EUT). His doctoral research examines the social context of playing digital games, focusing on measurable differences between experiences in solitary and social play.

Wijnand IJsselsteijn has a background in psychology and artificial intelligence with a M.Sc. in cognitive neuropsychology from Utrecht University, and a PhD in media psychology/HCI from the EUT. He currently holds the position of associate professor in the area of human factors and advanced media environments, specializing in social digital media, immersive media technology, and digital gaming. His focus is on conceptualizing and measuring human experiences in relation to these advanced media.

Ian Livingston is a graduate student in Computer Science at the University of Saskatchewan, under the supervision of Dr. Regan Mandryk. In 2009, He received his B.Sc. (Hon) degree in Computer Science. His areas of research are diverse including cognitive psychology, collaborative play in massively multiplayer online games,
exercise games, and video game usability evaluation techniques. His current research focus is in usability and playability heuristic evaluation in video games, where he continues to work in the validation and improvement of usability evaluation techniques specifically for video games.

Steven Malliet (Ph.D.) performs research on the social/psychological impact and educational potential of digital game play. He has published internationally on subjects such as digital game realism, game theory, educational games, video game history, and digital game effects. He is a co-founder of the DiGRA Flanders research organisation. He currently works as a visiting professor at the Play&Game research group (University College of Limburg) and as an assistant professor at the centre for Media, Policy a Culture (University of Antwerp).

Regan Mandryk is an Assistant Professor in Computer Science at the University of Saskatchewan. Her educational background includes a B.Sc. degree in Mathematics and Physics, an M.Sc. in Kinesiology, and a Ph.D. in Computing Science. She explores the use of physiological data to continuously model a user’s emotional experience when interacting with technology. Her current research interests are on sensing and modeling user state, designing interaction techniques and games for emerging devices, and designing persuasive games to encourage healthy living.

Lennart Nacke is a gameplay and user experience researcher, whose scientific interests are game usability evaluation techniques, affective player testing and physiological player-game interaction using brainwaves, muscle contractions and eye tracking. He holds a Ph.D. degree in Digital Game Development from Blekinge Institute of Technology, Sweden, and currently works on affective computing, gaming for fitness, and evaluating virtual worlds as a postdoctoral research fellow at the University of Saskatchewan, Canada.

Henk Herman Nap has a background in cognitive ergonomics, with a M.Sc. degree in psychology. His main interest is on reducing the complexity of current technology interfaces by cognitive support systems, with a special interest in user groups that face
the most difficulties interacting with current media (e.g. seniors). His PhD focused on stress in senior computer interaction.

Joerg Niesenhaus is a research associate at the Interactive Systems and Interaction Design research group at the University of Duisburg-Essen. He is engaged in several research activities within the context of interactive digital games and entertainment computing such as user-generated content, interactive storytelling, health & serious games and game usability. Before he joined the chair in 2003 he participated in the development of more than a dozen games for companies like Ubisoft and Blue Byte (e.g. The Settlers, Battle Isle, Incubation). Joerg holds a Master of Science in Applied Communication and Media Science and is a PhD candidate since 2007.

Koos C.M. Nuijten has an MA in remedial education and a PhD in communication science. He is senior lecturer and head of research in the International Center for Experimental Media Effects Research at the Academy for Digital Entertainment at NHTV Breda University of Applied Sciences. His expertise covers media processing theories and research linking media form and content to media effects (e.g., learning, persuasion). His research interests include media and learning, effects of in-game product placement, subliminal advertising, and behavioral research in virtual settings. He taught courses in communication science, general research methodology, content analysis and media ethics at several universities in the Netherlands and he is member of the advisory board for a PSB sponsored project on media literacy.

Niels Quinten graduated in 2010 with great distinction as a Master in Communication and Multimedia Design, at the Media & Design Academy, Limburg. In his Master thesis he investigated how persons can teach each other physical activity through digital gaming. A central issue in his research is the role of narration, both in-game and out-of-game, as a guide in the process of digital physical education.

Jonas Schild is, since April 2009, researcher at the Entertainment Computing Group of Prof. Maic Masuch at the University of Duisburg-Essen in Duisburg, Germany. Currently working on his PhD thesis, he is passionate about developing new forms
of game experience, in particular related to stereoscopic vision and 3D interaction technologies. He further investigates formalizations of game development processes that help create innovative gameplay mechanics. Jonas previously worked for the Fraunhofer Gesellschaft on input devices in the Virtual Reality domain. Before that, he was a product manager at Wacom Europe for pen-based input devices and created the kids game Colorelli. The game was internationally awarded for its innovative combination of a drawing input device with a creativity-fostering game world, e.g. with the German kids game award Tommi 2007 for “The Best Tool.” Jonas’ educational background is computer science and media technologies. He graduated on diploma-level with distinction in 2006 as media systems scientist at the Bauhaus-University Weimar, Germany under supervisors Prof. Charles Wüthrich and Prof. Bernd Fröhlich.

Ben Schouten graduated from the Rietveld Art Academy and worked as a professional artist for more than twenty years. He founded Desk.nl, an Application Software Provider (ASP), providing innovative internet related solutions. Together with the Dutch Design Institute, Desk was internationally awarded a webby award in gaming.

In 2001 he received his PhD on content based image retrieval interfaces that express in an adaptive and intuitive way image similarities according to human perception. In the same year he resumed teaching at the Utrecht School of Art & Technology (HKU) in interaction design, supervising masters and bachelors in interaction design and gaming. He is an advisor for the Dutch Media Fund, supporting E-culture for national broadcasting companies. Since 2010 he is Professor Playful Interaction at the Faculty of Industrial Design at Eindhoven University of Technology as well as Lector Serious Game Design at Fontys University of Applied Sciences.

Jan van Looy is senior researcher at the research group for Media and ICT (IBBT-MICT) at Ghent University, Belgium, where he does research into various aspects of video gaming. His PhD (2006) dealt with the socio-cultural shaping of the video game medium, published as “Understanding Video Game Culture” (book, 2010). In 2007 he participated in a study commissioned by the Flemish Parliament into the social impact of gaming in Flanders, published as “Jongeren en gaming“ (Youngsters and Gaming, book), and in 2007 and 2008 he worked on a research project on self-imagineing in
games as a postdoctoral fellow at HUMlab in Umeå, Sweden. Since 2008 he works as a senior researcher at MICT where he is currently doing research into video game experience, particularly identification, serious games for raising social awareness and serious games for learning foreign languages.

**Veerle van der Sluys** has a master degree in nuclear physics. She obtained a PhD degree at the University of Gent in 1995. She has worked for several ICT companies and learned to tackle software projects of different sizes. She has a passion for software development and new technologies. She was involved in several research projects: Knosos (Knowledge Sharing over Social Networks), ISCRAM (International Community on information systems for crisis response and management). In June 2009 she started at the Media and Design Academy, KHLIM, Hasselt as research coordinator for the Play&Game research group.
NHTV Expertise Series

In the context of contributing to the development of knowledge in the subject areas of digital entertainment, hotel and facility, city planning, logistics and mobility, tourism and leisure, NHTV Breda University of Applied Sciences has launched the NHTV Expertise Series. These publications have a logical link with NHTV’s strategy and lines of research and they will contribute to stressing NHTV’s distinct features as a knowledge institute.

The publications in the NHTV Expertise Series are:
1. ‘Coast tourism: a tour d’horizon’
   Research report by Martijn Smeenge
2. ‘Education, research and the art of creative thinking’
   PhD thesis by Paul Delnooz
3. ‘The global and the local: thinking inclusively about cultures in Breda and the rest of the world’.
   Speech Associate Professorship Cross-cultural Management by Vincent Platenkamp
4. ‘Systemic Constellations Work in Organizations’
   Dissertation by Joseph Roevens
5. ‘Understanding the behaviour of cultural tourists; towards a classification of Dutch cultural tourists.’
   Dissertation by Rami Isaac
   A pilot-project in cooperation with NBTC-NIPO by Kim de Brujin, Rob Dirven, Eke Eijgelaar and Paul Peeters
7. Het talent van Brabant. De rol van diverse actoren in talentontwikkeling binnen de sport
   Research report by Theo Hutten
8. Voices in Tourism Development | Creating Spaces for Tacit Knowledge and Innovation
   Conference proceedings edited by R. Isaac, V. Platenkamp and A. Portegies
9. Customer Relationship Management in Hospitality | A theoretical introduction & guidelines for applying the CRM-7-18 model | MSc. Olaf Hermans and Dr. Dan Mount

Colophon

Playability and player experience: Fun and Games 2010 Workshop proceedings

Copyright © 2010, NHTV Breda University of Applied Sciences
All rights reserved. Nothing may be published, or reproduced without prior consent of the authors.

NHTV Breda University of Applied Sciences
Academy for Digital Entertainment
P.O. Box 3917
4800 DX Breda
The Netherlands
T +31 (0) 76 533 2203
F +31 (0) 76 533 2205
www.nhtv.nl
The development of systems for establishing playability and player experience, as well as implementing such measures earlier in the game design cycle can be expected to optimize the game development process as well as the final game design.

This book contains the 7 papers that were selected for the Playability and Player Experience workshop that was held in conjunction with the Fun and Games Conference 2010, hosted by the Centre for User Experience Research of the Katholieke Universiteit Leuven in Belgium. All these papers were positively refereed by an International Programme Committee consisting of 12 qualified experts with a wide variety of backgrounds and expertise domains.

The focus of the Playability and Player Experience workshop was on the development and use of all kinds of playability and player experience measures (like quantitative metrics and data harvesting) to evaluate digital games. Because playtesting metrics are standard practice in larger game development studios, we decided to predominantly focus the workshop on casual games playability measures: playtesting metrics are very much in demand yet not often available within the casual gaming industry.