How to Present Game Difficulty Choices? Exploring the Impact on Player Experience

Jan D. Smeddinck[1], Regan L. Mandryk[2], Max V. Birk[2], Kathrin M. Gerling[3], Dietrich Barsilowski[1], Rainer Malaka[1]

[1]Digital Media Lab, TZI, University of Bremen, Germany, lastname@tzi.de

[2]University of Saskatchewan, Saskatoon, SK, CA, firstname.lastname@usask.ca

[3]University of Lincoln, Lincoln, UK, kgerling@lincoln.ac.uk

ABSTRACT

Matching game difficulty to player ability is a crucial step toward a rewarding player experience, yet making difficulty adjustments that are effective yet unobtrusive can be challenging. This paper examines the impact of automatic and player-initiated difficulty adjustment on player experience through two studies. In the first study, 40 participants played the casual game THYFTHYF either in motion-based or sedentary mode, using menu-based, embedded, or automatic difficulty adjustment. In the second study, we created an adapted version of the commercially available game *fl0w* to allow us to carry out a more focused study of sedentary casual play. Results from both studies demonstrate that the type of difficulty adjustment has an impact on perceived autonomy, but other player experience measures were not affected as expected. Our findings suggest that most players express a preference for manual difficulty choices, but that overall game experience was not notably impacted by automated difficulty adjustments.

Author Keywords

Game difficulty; player experience; game-user research; flow; dynamic difficulty adjustments; feedback.

INTRODUCTION

Research has demonstrated a breadth of benefits of games, for example, on player cognition [22], physical health [33], and general well-being [44]. Therefore, games are now targeting broad audiences with heterogeneous expectations and abilities. Particularly in the area of serious games, researchers and designers are often addressing audiences with special needs, for example, young children [21], people with disabilities [25,28], or older adults [23]. A crucial step in this process is ensuring that games meet the needs of players to provide a positive, empowering experience. To this end, it is important to provide balanced gameplay that does not overwhelm individual players by being too challenging and that enables players of different abilities to play together [26]. Balancing game settings to achieve captivating experiences that can harness the full motivational potential of games is challenging, and previous work has only begun to explore this area. Manual difficulty choices, typically presented as menu settings, have long been an important element of games [11]. In addition to predefined difficulty level choices that change the base level for usually ongoing difficulty increases as a game progresses, dynamic difficulty adjustments (DDA) in games can improve gameplay [30]. A growing body of work is concerned with automated difficulty adjustments, which not only promise to reduce the burden placed on players and avoid breaking the magic circle – the special place in time and space created by a game [29,43] – but also bear the potential to influence a large amount of fine-grained variables. However, selecting matching settings for individual players is a complex problem, involving the unpredictability of human agents and the variety and interplay of game settings.

Thus, recent work suggests improvements in manual difficulty choices that can also interact with dynamic difficulty adjustment [11,12]. Based on flow theory [16], the work highlights the challenge of difficulty balancing and adjustments in games with an emphasis on the importance of enabling *personal control* whilst retaining *autonomy* and *avoiding interrupting the flow* of an activity, leading to the concept of *player-oriented difficulty choices* that are embedded within the game world [11]. While related work presents a theoretical basis for the interaction with game difficulty choices, we found no reported empirical evaluations on this topic.

Stemming from our research in the area of full-body motion-based games for health, we were interested in the question: *"Do different modes of presenting difficulty choices impact the player experience?"* The concept of player-oriented difficulty choices suggests the method of embedding the choices in a way that blends with the game world, while the most common solution employed in games on the market are difficulty settings in classic menus that may adhere roughly with the visual style of the game, but do not blend with the regular game interaction or mechanics. Games with DDA frequently offer no additional difficulty choices.

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/05…$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858574.

We test the hypothesis that *embedded* difficulty choices lead to a better overall player experience by retaining autonomy and control, and by avoiding interruptions of the gaming that endanger immersion or might force players out of the magic circle when compared to traditional *menu* difficulty choices or *auto*matic difficulty adjustments without interaction options. Our work contributes empirical insights relating to the impact of different modalities for interacting with game difficulty choices on player experience and preferences.

BACKGROUND

Getting the level of challenge to match the capabilities and needs of a player is a core element of good player experience. This has long been discussed in game user research [14], and the most cited psychological foundation on the balance of challenges and skills is flow theory [16].

Flow in Games

Csikszentmihalyi [17] describes flow as state of being fully "in the zone" when engaging with an activity. This can have positive effects, such as increased motivation to perform an activity again, which may be explained through a feeling of enjoyment related to evolutionary benefits of performing certain activities [17]. The most prominent precondition that is required if flow is to be achieved is an optimal balance of risk of failure (in games e.g., losing a move, a life, a level, or the entire game) and the chance to attain the goals (in games e.g., winning some points, a bonus, a level, or the entire game) when performing an activity [41]. The entire set of conditions that are prerequisites for flow experiences are in short: clear goals, immediate feedback, match of challenges and skills, action and awareness merge, concentration on the task at hand, sense of potential and control, loss of self-consciousness, sense of time altered; resulting in an experience becoming autotelic [17]. Seeing these conditions, it is not surprising that Csikszentmihalyi frequently uses games as examples of activities that can induce flow. Related work lists explicit examples of how the facilitating factors for flow are present in games [13,31,48]. Because the balance of challenge and skills is an important precondition, and skills differ between people, it becomes clear why balancing challenges is such an important aspect of game design [41]. It also becomes clear why most games still offer manual difficulty choices in setting menus: even after iterative testing and balancing, average solutions are likely suboptimal.

Difficulty Choices in Games

Game difficulty choices that are presented in menus with typical labels such as "easy, medium, hard" can be found even in very early and simple games. The "classic way to present difficulty choices" has arguably evolved largely as a matter of technical circumstance and the prominent use of difficulty selection menus today is arguably the result of established customs. Although manual – i.e., explicit – feedback has been used for difficulty adjustments with unusual input modalities (such as biofeedback), and has shown increased immersion and affect [34], menus with multiple levels of one monolithic and unspecific parameter remain the most common form of difficulty choice UI.

Dynamic Difficulty Adjustments and User-guidance

Because the difficulty of a game is often the product of many different game variables, and because it can change from one moment to the next, manual difficulty choices are not always compatible with seeking the optimal game experience. *Dynamic difficulty adjustments* (DDA) that automatically adjust difficulty based on threshold heuristics [30], or on machine learning models [18], have been explored as a reasonable alternative. Such systems have even been used in regular consumer games, such as *Half-Life 2* and *Max Payne* [1,3] and in serious games [2,47].

There are great potential benefits to DDA, such as high frequency and detailed adjustments [3,30], and in theory they merely require an adequate *performance evaluation* and an *adjustment mechanism* [1]. However, DDA adoption faces challenges, because decisions that contradict the will of the players are potentially harmful to player experience, and in the details of the implementation, assessing affective states is hard [32], especially when the tools should be unintrusive. Balancing adjustment mechanisms is extremely challenging due to the very personalized nature of the outcomes. Especially when first confronted with a new player, be it for balancing between players, or for adjusting the difficulty for an individual player, such DDA or adaptive systems suffer from cold-start problems. Even if the cold-start problems can be overcome and players are provided with a well-matched player experience that results in continued play, adaptive systems suffer the risk of getting stuck in local extrema, or alternatively of enacting strongly fluctuating difficulty settings (rubber banding), which may result in unacceptably balanced play sessions.

Asking players to provide explicit feedback can provide a direction and extent of settings that supports a DDA system to function well. Existing work on user-guided adaptive systems focuses on which information the guidance provides and how it can improve the system [9,50]. User modeling and adaptive systems work also looks at explicit feedback for user-guidance under the term (advisory) dialogue / communication [37] and some approaches are explicitly built around flow theory, with boredom [8] and frustration [27] as critical measures. However, while in traditional software having an explicit dialogue about the adaptive system may often be acceptable, in games, this might endanger immersion [11]. It is therefore important to consider the effects that the presence and presentation of manual difficulty choices have on game user experience.

Difficulty Choices and Flow in Games

Flow in games has been discussed by a number of researchers in the field [11,13,48]. In this work, we focus on the work by Chen, as the concept of player-oriented (embedded) difficulty choices was introduced in this work [11,12]. He builds on arguments for DDA, suggesting that frequent difficulty adjustments can support flow in games, and that allowing players to exert user-guidance by providing explicit feedback can provide a feeling of "being in control" and can also inform more adequate adaptations. Chen is cautious that frequent interactions with difficulty setting menus might be disruptive to being immersed in a game, which would in-turn be detrimental to experiencing flow, and so he develops the concept of embedded difficulty choices, which are implemented in such a manner that they blend with the actual game, so that the feeling of being in control as an autonomous actor is provided, while players remain in the magic circle [29] of the game world.

In his line of arguments, Chen relies heavily on flow theory and the theoretical advantages of the embedded difficulty choice approach intuitively appear coherent in this light. He also presents an informal study with two games (called *Traffic Light* and *fl0w*) to underline the arguments, with the latter one being specifically designed to implement embedded difficulty choices. However, despite the large number of references to the work in related literature (> 500 references), we could not find an empirical investigation of the effects of embedding player difficulty choices into the core of the interactive experience (or *player-oriented DDA*).

Self-Determination Theory and Flow

Considering the prerequisites for flow, embedded difficulty choices appear prone to be supportive of flow experiences. However, the benefits can also be reasoned based on other motivational theories, such as self-determination theory, which may be beneficial due to the following reasons:

Flow as an indicator for player experiences is highly debated and *"it simplifies the dynamics of intrinsic motivation"* [17] (p. 83), whereas the self-determination theory (SDT) approach has seen growing adoption in player experience research [6,7,46]. SDT in games is assessed using the *Player Experience of Needs Satisfaction* (PENS) questionnaire [42], and results in subscales that can directly inform game design decisions as opposed to the common flow scales [36,49]; flow is a multi-dimensional construct and a matter of present experience – a process variable – that is difficult to measure [5]. For example, the levels of satisfaction of autonomy or competence have been shown to be good indicators of the motivational power of games [40], and they are linked to "feeling in control". Links between flow and SDT and its measures for intrinsic motivation have been discussed in related work [42] making connections via the aspect of presence / immersion. Csikszentmihalyi also acknowledges similarities between flow and SDT, highlighting the aspect of autonomy. In his view, flow theory arose from an interest in what propels people to initiate or continue actions because they enjoy the performance in the present [17], while other theories (such as needs satisfaction) are more outcome oriented. If SDT and flow theory set a different emphasis (in which SDT is concerned with the preconditions and building factors for intrinsic motivation, and flow theory is concerned with the current and sustained experience of intrinsically motivated activities), then arguably the major SDT dimensions of competence and autonomy needs satisfaction can be interpreted as provisions for flow experiences, mapping in particular to balance of challenge and skills and the sense of potential and control.

EXPECTED IMPACT OF DIFFICULTY CHOICE MODES

Based on this theoretical background, we explore approaches to game difficulty adjustment in a comparative study with three conditions and the following expectations:

Menu – Players select one option from multiple levels of a single difficulty parameter that is presented with through a classic WIMP (windows, icons, menus, pointer) interface.
H_{mA1}: Players will experience higher levels of *autonomy* relative to conditions without explicit choices (here: *auto*).
H_{mB1}: Players will experience reduced *presence / immersion* relative to conditions where the gaming experience is not interrupted by elements that do not blend seamlessly with the game world (here: *embedded* and *auto*).

Auto – An implementation of dynamic difficulty adjustments that are performed automatically and where the players are not able to make explicit difficulty choices.
H_{aA1}: Players will experience reduced levels of *autonomy* relative to conditions with explicit difficulty choices.
H_{aB1}: Players will experience increased *presence / immersion* relative to conditions where the gaming experience is interrupted by elements that do not blend seamlessly with the game world (here: *menu*).

Embedded – Building on the approach of player-oriented difficulty choices [11], players make explicit difficulty choices by interacting with the game world, integrating closely with the visual design and game mechanics.
H_{eA1}: Players will experience increased levels of *autonomy* relative to conditions where explicit difficulty choices are not possible (here: *auto*).
H_{aB1}: Players will experience increased *presence / immersion* relative to conditions where the gaming experience is interrupted by elements that do not blend seamlessly with the game world.

These assumptions were considered most reasonable based on related work, while other effects were also deemed possible. In the *auto* condition, for example, with a DDA system in place, control over difficulty is executed by the system. This may be appreciated by the players, if they are not interested in executing control over this aspect of their interactive experience and the system does not make obvious mistakes. Yet, players might also (to an extent) appreciate full control over those aspects of the system [45]. Conditions with manual (explicit) control might also have an impact on perceived competence need satisfaction; although, in this case, the direction does not appear clear. The impact might be positive (due to perceived control of the system), or negative (e.g. due to becoming aware of requiring "easier" settings). We therefore opted to include competence in our measures but did not attach a directed hypothesis during our investigation.

We were motivated to this research by our ongoing efforts in the area of motion-based games for health. With applications in the context of therapy (where extrinsic motivation and heteronomy play an important role) re-gaining competence and autonomy (thus boosting intrinsic motivation) can be valuable. On the other hand, matching the individual capabilities and needs of different members of very heterogeneous target groups is especially important in such use cases and could benefit from the timely adjustment of multiple parameters, calling for automated support with difficulty choices. In this light, the concept of embedded feedback bears the promise that expression takes place in an unobtrusive manner, as a meaningful part of the interactions with the game world.

STUDY 1: THE IMPACT OF DIFFERENT INTERACTION STYLES FOR GAME DIFFICULTY CHOICES

In order to explore the impact of different modalities for game difficulty choices on the player experience, we conducted a study based on a casual skill game called *The Higher You Fly, The Harder You Fall* (THYFTHYF). In order to attain connectivity of the findings to our ongoing research on full-body motion-based games, the study was implemented as a mixed design with a two level between-groups independent variable of *controller type* being either *motion-based input*, or *gamepad input* and a three level within-subjects independent variable *game difficulty choice interaction modality*. For the purpose of this paper we focus on the within-subjects analysis of the *gamepad* group and only remark on selected between group comparisons.

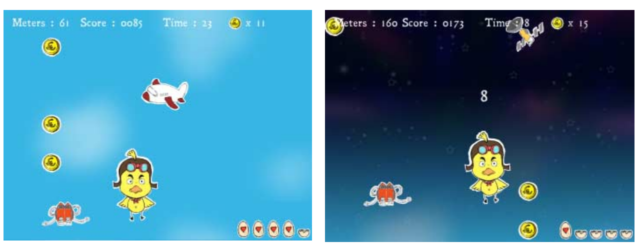

Figure 1. Two screenshots of the game THYFTHYF taken at different height (progress) levels show common game objects.

Design & Implementation

The game was selected to fulfill a number of requirements: It should transparently and immediately reflect changes in player performance, in order to assure that players would experience effects of their difficulty choices. This includes offering dense in-game and post-game audio-visual feedback. The game was also balanced following pilot study runs to include the possibilities of feeling overstrained, or of losing, even within a short time.

In the game THYFTHYF, the player controls a bird player character (PC) with the goal to fly as high as possible whilst collecting points on the way (cf. Figure 1). It can be played with two types of controls: *motion-based control*, where the player is tracked with a Kinect (v1) and has to move her/his arms, mimicking the motion with which a bird flaps its wings and *gamepad control* (here: XBOX360 controller), where the player repeatedly pushes and releases the continuous trigger buttons, also mimicking the motion with which a bird flaps its wings. The game world is composed of tiles. Each tile belongs to one of three classes: ground, sky and space. At runtime, the tiles are procedurally placed and populated with "good items", which increase the player's score if they are collected by hitting them with the player character and "enemies", which hurt the player character when contacted and give a sideways impact to the player character that reduces the current balance. The PC can be moved upward by "flapping" both wings at equal speed, where the frequency controls the speed. The direction of flight can be controlled by executing wing-flapping movements with a relatively faster speed on one side (causing the bird to "lean" to the opposite side). If the PC leans too far to either side or stops flapping the wings for too long, it falls down a bit and the player loses a life (starting with a supply of five per round). If all lives are lost, the bird falls all the way to the ground and the player has to start flying up again from ground level with newly refreshed lives. These design decisions were made to assure comparable stimuli; the duration was set to 60 seconds per round. The difficulty, realized by changing speed, balance support and of bad/good objects, increases with increasing height in the level as offsets of base parameters that were influenced by the manual or automatic difficulty choices. After each round, a summary screen showed the final score.

Difficulty Choice Interaction Modalities

A player-oriented embedded difficulty choice mode was implemented in the form of five different start boosts that reflected the five levels present in the alternative difficulty selection menu (cf. Figure 2). Selections in the difficulty menu would be reflected in a different duration of a jetpack starting boost. Higher boosts would result in increased difficulty settings, whereas lower boosts result in lower difficulty settings. Both modalities were invoked at the beginning of each round in order to assure comparable exposure. The player would also receive a boost start in the DDA only condition without any explicit difficulty choices.

Dynamic Difficulty Adjustments

Performance based DDA was implemented to allow for a condition without difficulty choices. The mechanism was designed to be limited to a controllable number of effectors, while allowing for distinctly notable adjustments. In the context of the three challenge mechanics that were present in the game (height, collection/avoidance, and balance), the *performance metrics* were defined as follows:

Over the last 5 rounds – weighted by recentness –, the height reached, the ratio of bad objects hit, the ratio of good objects collected and the times out of balance were each evaluated by threshold-based heuristics. The balance was set in a way that required the players to perform at a challenging level in order to reach the space level, which always started at a fixed height relative to the final starting position (after a starting boost), and to assure that players don't struggle with "out-of-balance" events too much, enabling them to collect coins, while avoiding enemies. Before each session, the according *difficulty parameters* of max speed / fall speed, number of bad / good objects, and balance support (via damping) were adjusted.

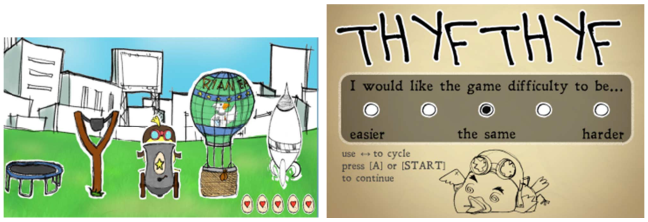

Figure 2: Left/right: Embedded difficulty choice / menu difficulty choice in THYFTHYF.

THYFTHYF was implemented in C# based on .NET 3.5, the XNA 4.0 framework and the Kinect 1.8 SDK.

Setup & Procedure

Trials of the laboratory study lasted about 60 minutes. We recruited from the local student body, and each participant received $10 CAD compensation for their time. Figure 3 illustrates the technical setup. The study language was English. The procedure was performed with 40 participants (19f, 21m, age: M = 26.88, SD = 5.79, min/max = 18/45), who were randomly assigned to the sedentary (gamepad) or the motion-based (Kinect) group. Four participants said they never play video games in a typical week, 16 said they play up to three hours per week and two participants said they play more than 35 hours/week, with the rest spending between 3 hours and 35 hours a week gaming. They have been gaming since 10.76 years on average (SD = 6.7).

Following the greeting and gathering of informed consent, participants were asked to complete a pre-study questionnaire before engaging in a warmup round with a short introduction by demonstration and explanation of the core game mechanics. Each subject then participated in four trials of gameplay consisting of three one-minute rounds interspersed with a difficulty choice in the conditions *embedded* and *menu*, followed by a fixed set of post-trial questionnaires. The fourth (and last) trial always featured a zero effect menu choice placebo condition that was omitted from analysis and report for brevity. The difficulty choice modality conditions *embedded*, *menu,* and *none* were presented in Latin square randomized order to counter-balance potential learning-, customization- and fatigue-effects. Lastly, the participants completed a short personality trait index and responded to a semi-structured interview about experiences and preferences.

The independent variable of *controller type* (motion-based vs. sedentary) was added to the design in order to facilitate carrying over results, because potential interaction effects (e.g. of feeling more / less in control or autonomous when exposed in regard to one's physical appearance and skills while playing motion-based games) might exist. We found no evidence for meaningful interaction effects and for reasons of brevity, the analysis of that group was largely omitted from this report. Notable fixed variables include design choices around interacting with explicit user feedback, which was always collected between one round of gameplay and the next (fixed intervals) and not based on the players' volition to assure comparable outcomes. Players were informed that they could influence the difficulty of the game and would be asked to do so between rounds. It was explained to them, what each difficulty choice in both the *menu* and the *embedded* condition would mean before the respective conditions.

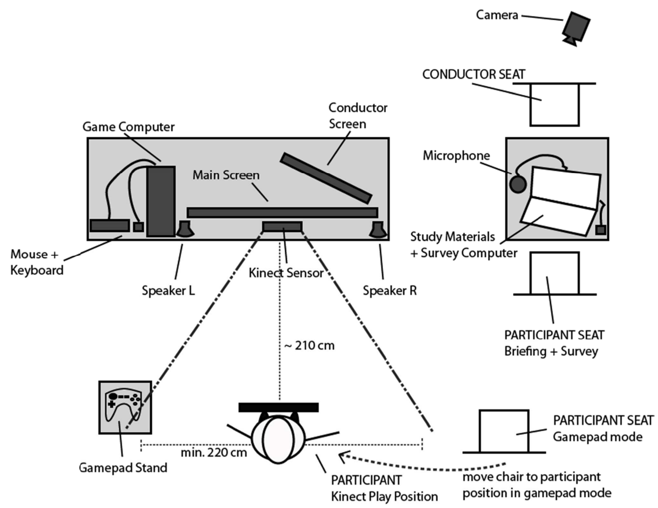

Figure 3: The technical setup of the study.

Measures

In addition to demographic background information, gaming preferences, and a brief personality questionnaire (TIPI), we measured *affect* with the *Positive and Negative Affect Schedule* (PANAS) [15], *player experience* with the *Player Experience of Needs Satisfaction* (PENS) [39] questionnaire (the dimensions *interest-enjoyment* and *effort-importance* of the *Intrinsic Motivation Inventory* (IMI) [35] were added as additional dimensions rooted in SDT to augment the PENS results), and a *Task Load Index* (TLX) was collected as a measure of *frustration* rooted in usability research, as the interaction with difficulty choices was realized either "inside" a game, or with regular GUI components. The questionnaires were presented in the order of introduction in this text and with randomized item order. Each questionnaire was completed multiple times by each participant (once after each trial) and the PANAS was also completed once before the first trial in order to facilitate relative offset calculations that compensate for individual differences in initial affect. The participants were instructed to consider their entire interaction sessions with the game (including interface components). Before the final interview, players also reported a ranking of their preference for the difficulty choice modalities together with responses to a number of question items regarding their usual perception of – and interaction with – difficulty choices in games. Video recordings and observational protocols completed the data collection.

Results & Analysis

The quantitative experiential measures were analyzed with a general linear model facilitating a mixed design repeated-measures analysis of variance (ANOVA) at a significance level of $\alpha = .05$ and Mauchly's sphericity tests, as well as Bonferroni adjustments prior to post-hoc pairwise t-tests for multiple comparisons. The analysis was performed in R with the *ezStats* package for the ANOVA operations, and variances were winsorized [19] (leveling outliers in the top/bottom .2 quantiles to the trim edge values). The results were confirmed using SPSS (version 20) and are summarized in Table 1 and Table 2.

STUDY 1 (PANAS, PENS, IMI)	**embedded [M (SD)]:**	**menu [M (SD)]:**	**auto [M (SD)]:**
positive affect	2.95 (.77)	2.96 (.68)	2.94 (.82)
positive affect (MB)	3.37 (.74)	3.44 (.75)	3.29 (.74)
negative affect	1.59 (.53)	1.58 (.64)	1.54 (.46)
competence	2.72 (.9)	2.8 (.7)	2.83 (.81)
presence / immersion	2.83 (.74)	2.84 (.73)	2.73 (.71)
autonomy	3.07 (.75)**	2.9 (.74)	2.68 (.9)**
relatedness	2.73 (.68)	2.77 (.75)	2.65 (.97)
intuitive control*	2.93 (.72)	3.03 (.74)	3.11 (.88)
intuitive cont. (MB)*	3.52 (.76)	3.57 (.55)	3.57 (.77)
interest enjoyment	3.34 (.53)	3.38 (.47)	3.3 (.61)
effort-importance	3.71 (.61)	3.71 (.44)	3.7 (.62)

Table 1: Mean (M) and standard deviation (SD) results of study 1. All items were recorded on 5 pt. Likert scales. Group effects comparing between sedentary and motion-based (MB) are indicated with (*). Within-subjects effects between the embedded, menu, and auto conditions are indicated with (**).

There were no significant effects on the *positive affect* or on the *negative affect* scale. Notably higher *positive affect* than *negative affect* and also higher *positive affect* in the *motion-based* (MB) game group are in line with the following measures and suggest that the scale is sensitive to changes in affect caused by the experience of playing different versions of the game. While the lack of significances does not prove the absence of effects, we assume that there are no large effects on *affect* between the three conditions, because the means between all conditions are very close.

The PENS results on the *competence* and *presence / immersion* dimensions show a similar picture. While the result on *competence* appeared difficult to predict due to the complex interaction of self-perceived, practical skill and interacting with difficulty choices, *presence / immersion* could be expected to be lowered in the *menu* condition, which was not the case in our sample and the similarity in means suggests an absence of strong effects. There was a significant difference on the PENS *autonomy* dimension ($F(2,76) = 5.01$, $p = .009$, gen. $\eta^2 = .02$ [4], Mauchly not sig.) with post-hoc pairwise comparisons confirming a sig. diff. between *embedded* and *auto* ($p = .02$). This finding can be seen as evidence to support an increased sense of autonomy needs satisfaction in the *embedded* condition that was predicted based on Chen's arguments. However, there is no discernable difference between the *embedded* and the *menu* condition. The PENS dimensions *relatedness* and *intuitive control* showed no significant differences on the indep. variable *difficulty selection*, as expected, while *intuitive control* was sig. increased under the *motion-based* control condition ($F(1,38) = 6.33$, $p = .016$, gen. $\eta^2 = .12$ [4], Mauchly not sig.), which could also be expected, as the game was originally designed to be motion-based.

The IMI dimensions *interest-enjoyment* and *effort-importance* showed remarkable similarity in means and no significant differences, suggesting that both player enjoyment and motivation were not notably different under the different difficulty choice modalities.

STUDY 1 (TLX)	**embedded** [M (SD)]:	**menu** [M (SD)]:	**auto** [M (SD)]:
physical demand*	.8 (10.33)	.65 (10.93)	1.15 (10.68)
phys. dem. (MB)*	11.05 (4.76)	8.35 (6.18)	9.5 (6.41)
mental demand	2.9 (7.02)	3.15 (7.34)	3.1 (8.5)
temporal demand*	7.35 (5.37)	7.1 (3.87)	6.75 (3.35)
temp. dem. (MB)*	2.3 (8.02)	2.35 (8.7)	2.05 (10.45)
performance*	-1.9 (7.72)	-.3 (8.3)	-1 (9.46)
perf. (MB)*	3.65 (8.42)	4.15 (7.21)	3.45 (7.07)
effort	8.15 (3.82)	8.25 (3.6)	7.4 (5.06)
frustration	-.75 (8.4)	-1.65 (8.93)	-1.95 (9.78)

Table 2: The mean (M) and standard deviation (SD) results of the TLX in study 1. All items were recorded on 40 pt. Likert scales (range -20 to 20). Group effects comparing between sedentary and motion-based (MB) are indicated with (*).

The TLX dimension *physical demand* showed no differences between the within-subject conditions, although the *motion-based* game group recorded significantly higher *physical demand* ($F(1,38) = 11.6$, $p = .002$, gen. $\eta^2 = .22$ [4], Mauchly not sig.), providing further evidence for the sensibility of the chosen measures. The TLX *mental demand* and *temporal demand* dimensions showed no significant differences, although the mean of the *temp. dem. auto* condition appears lowered, which seems reasonable given the lack of a manual selection process. *Temporal demand* appears sig. decreased in the *motion-based* game group ($F(1,38) = 5.66$, $p = .023$, gen. $\eta^2 = .11$ [4], Mauchly not sig.), which cannot be explained by observable differences in actual time spent and seems contradictory to the physical effort measure, hinting at a potential interaction with overall motivational effects. There were no sig. diffs. on *difficulty choice* in the TLX *performance* dimension, although there was a sig. diff. between *sedentary* and *motion-based* ($F(1,38) = 4.43$, $p = .042$, gen. $\eta^2 = .09$ [4], Mauchly not sig.), which may be a secondary effect to the observed difference in *physical demand*. While THYFTHYF appears to be a rather high-*effort* game, the final TLX dimensions of *effort* and *frustration* showed no further sig. diffs., hinting at a further lack of notable negative or positive effects of the *difficulty choice modality*.

Interviews

In the interviews, three out of four participants said that they *like being able to change settings*, as opposed to being *more happy just playing the game*. Regarding the difficulty selection in THYFTHYF, one participant expressed it this way *"I did like being able to choose the difficulty settings because it felt like I could cater the game to how well I played, but I also liked playing the game itself"*, while some others expressed notions such as: *"I'm just more happy playing the game"*. Some participants added that they *appreciate a broad range of (difficulty) settings*.

When asked *how they decided which difficulty to select*, participants mentioned various strategies they followed to make difficulty choices ranging from *depending on last time*, over *first easier then harder* and *just stay in the middle*, to *just hard* and *first easiest, then hardest*, and in three cases even just *random* (which was only mentioned with embedded difficulty selection but not regarding the menu). A common notion was *"I just did it based on how it felt the previous time I played it"*.

Players expressed a broad range of opinions regarding how difficult they found it to play the game, ranging from *too easy* (2) over *okay / all right* to *too hard* (4 mentions, 3 of those relating to difficult controls; occurring in both controller type groups). 20 participants added that they generally *prefer games to be challenging (or hard)*.

Regarding whether they felt that their difficulty selections made a difference, 22 participants explicitly said they *noted that their difficulty selections made a difference*. Three did not clearly reply to this question, while 15 did *not clearly say that they noted differences*. However, some added positive unasked remarks on the DDA, such as *"I liked the idea that the game difficulty was automatically adjusted to my performance"*. In those parts of the open-ended discussion elements that led to players making a statement whether they like the idea of DDA in general, 9 were positive about DDA, while 6 expressed rather negative connotations. During the post-study interviews, the participants were eventually informed about the difficulty selection interaction modality manipulation and later question items named the conditions, which may explain why some participants expressed preferences for automatic adjustments or manual adjustments.

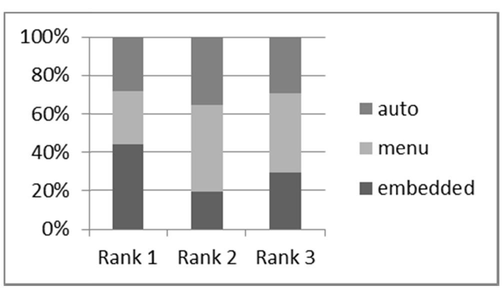

Figure 4: Ranking by participants regarding preference in study 1 (with the game THYFTHYF).

The ranking of all conditions (cf. Figure 4) in comparison after completion of the trials does not show significant results on the Friedman rank sum test. While *embedded* received slightly more first rank preferences, Exact Wilcoxon-Pratt Signed-Rank Tests also did not show any notable contrasts in pairwise comparisons.

While the findings seem to support Chen's cautioning of an impact on the player *autonomy*, *presence/immersion* does not appear to be impacted and the difficulty choice modality similarly has *no notable impacts* on a large variety of enjoyment related measures. We were intrigued by the lack of significant effects on presence / immersion and any secondary or related experiential measures (especially interest-enjoyment, effort-importance, and affect).

STUDY 2: THE IMPACT OF DIFFERENT LEVELS OF FEEDBACK EXPLICITNESS IN CHEN'S GAME FL0W

Results from study 1 showed a surprising homogeneity, especially of the resulting directly enjoyment-related measures, which were thought to react to the different game difficulty choice modes. This partially contradicted predictions that we drew from Chen's considerations. However, Chen's arguments relate directly to his own game implementations. Our game THYFTHYF – while also being a casual game – bears quite a number of differences, especially with regard to how difficulty choices were triggered and how the embedded difficulty choice mode integrated into the game world and the game mechanics. We thus repeated the study in a comparable setup with Chen's well-known game *fl0w* in order to further investigate the unexpected absence of effects and to seek confirmation for the effect on autonomy.

Design & Implementation

In the game *fl0w*, which can be described as a casual atmospheric game, the player controls a microorganism that can chase and "eat" (collide with the mouth section) available food items that float around in multiple fluid-like layers of a game world that may have been imagined as a petri dish. The player organism moves towards the position of the mouse cursor on screen, when the left button is clicked. With increasing "depth", other microorganisms are present in the layers and some will attack the player's organism by attempting to "eat" some of its components. The player organism can similarly "fight" other organisms by "eating them" piece-by-piece. The embedded difficulty choice mode originally implemented in the game allows players to move up and down through different difficulty layers of their own volition by eating either a special red piece (one layer down; harder) or a special blue piece (one layer up; easier) as shown in the instructions screenshot in Figure 5. The game *fl0w* was converted to *Haxe* based on the *AS2* source code made available by J. Chen [10].

Difficulty Choice Interaction Modalities

In addition to the original embedded difficulty choices method, we implemented a traditional menu, which could be opened at any time during play by clicking the right mouse button (see Figure 6), as well as a fully automatic DDA method, relating to the methods *embedded*, *menu*, and *auto* from study 1. The most notable difference in difficulty selection, aside from the selection upon the players' volition at arbitrary times during game play in *fl0w* was that THYFTHYF uses a five point difficulty choice scale, whereas the difficulty choice in *fl0w* is binary.

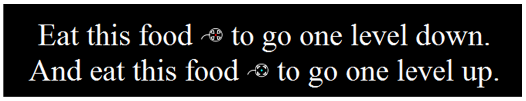

Figure 5: Instructions for the condition *embedded*.

Instructions provided for the condition *menu* were: *"Click the right mouse button at any time to change the level."*

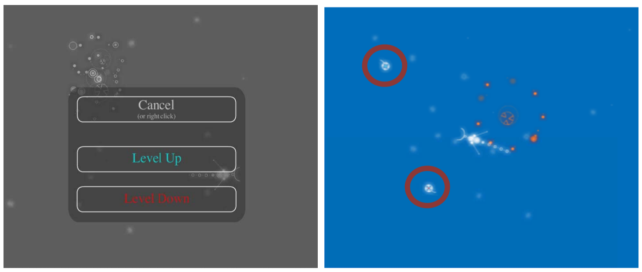

Figure 6: Difficulty choice interaction styles for the condition *menu* (left) and *embedded* (right; red circles are clarification markers not present in the game). Left background elements and right game assets: © Jenova Chen / thatgamecompany.

Dynamic Difficulty Adjustments

In order to automatically increase difficulty, a timer-based heuristic check was performed regularly to determine whether any other organisms or food items were still visible to the player. If there were no further other organisms alive in the layer, as well as no more than two food items available, the difficulty level was increased by moving one layer down. The mechanism for automatically decreasing difficulty was kept as originally implemented in *fl0w*, meaning that the player's creature would automatically be moved up one layer if its health status (as expressed by intact body segments) fell one level due to an attack by another creature.

Setup & Procedure

The study setup duplicated the setup from study 1 with the following adjustments for brevity: The BPNS and TIPI personality trait questionnaires were skipped, as they were not of interest to this comparison study. Study 2 also did not feature interviews or a motion-based game condition (safe to omit due to the between groups mixed design of study 1). Latin square randomization was maintained for the order of within-subjects conditions.

The study was completed by 18 participants (7f, 11m, age: M = 26.78, SD = 5.71, min/max = 20/45) to test whether effects with notable effect sizes would occur. Four participants said they never play video games in a typical week, seven said they play up to 3 hours per week and none said that they played more than 21 hours per week. They have been gaming since 13.28 years on average (SD = 6.64). Participants did not receive monetary compensation and were conveniently sampled mostly from a student population in Germany. The study language was English.

Results & Analysis

There were no significant effects on the *positive affect* or on the *negative affect* scale (cf. Table 3), mirroring results of study 1 and suggesting again the absence of large effects between the three conditions, because the means are very similar. Also in agreement with study 1, the PENS results on the *competence* and *presence / immersion* dimensions show a similar picture with the lack of an effect on *presence / immersion* in contrast to the menu condition, contradicting the expectations. The significant difference on the PENS *autonomy* dimension from study 1 was confirmed ($F(2,34) = 4.17$, $p = .024$, gen. $\eta^2 = .05$ [4], Mauchly not sig.) with post-hoc pairwise comparisons confirming a sig. diff. between *menu* and *auto* ($p = .047$). This finding can be seen as evidence to support an increased sense of *autonomy* needs satisfaction in the *menu* condition that was predicted based on the Chen's arguments, providing a helpful complement to the significant difference found in post-hoc pairwise comparison between the *embedded* and *auto* condition in study 1, while again finding very similar means in the *embedded* and the *menu* conditions. The PENS dimensions *relatedness* and *intuitive control* showed no significant differences, as expected, and in agreement with study 1. Findings were also repeated for the IMI dimensions *interest-enjoyment* and *effort-importance*, which show remarkable similarity in means, including the *embedded* condition, and no significant differences.

STUDY 2 (PANAS, PENS, IMI)	**embedded** [M (SD)]:	**menu** [M (SD)]:	**auto** [M (SD)]:
positive affect	26.79 (3.57)	26.97 (5.65)	26.02 (5.72)
negative affect	12.57 (2.65)	12.13 (1.83)	12.94 (1.99)
competence	3.37 (.6)	3.47 (.49)	3.61 (.47)
presence / immersion	2.67 (.62)	2.6 (.77)	2.63 (.7)
autonomy	3.06 (.7)	3.17 (.67)*	2.84 (.47)*
relatedness	1.69 (.66)	1.72 (.6)	1.65 (.6)
intuitive control	4.42 (.51)	4.35 (.39)	4.56 (.45)
interest enjoyment	3.67 (.5)	3.82 (.36)	3.64 (.39)
effort-importance	2.89 (.67)	2.69 (.58)	2.71 (.39)

Table 3: Mean (M) and standard deviation (SD) results of study 2. All items were recorded on 5 pt. Likert scales (pos. and neg. affect use mean sums of item scores; range 10 - 50). Within-subjects effects are indicated with (*).

The TLX dimensions *physical demand* and *performance* (cf. Table 4) showed no significant differences, also repeating the findings from study 1. The TLX *effort* ($F(2,34) = 6.47$, $p = .004$, gen. $\eta^2 = .11$, Mauchly not sig.) and *frustration* ($F(2,34) = 3.6$, $p = .038$, gen. $\eta^2 = .07$, Mauchly not sig.) dimensions did, however, show sig. diffs., with post-hoc pairwise comparisons showing a sig. diff. between *embedded* and *manual* on the *effort* dimension ($p = .011$) and between *embedded* and *auto* on the *frustration* dimension ($p = .047$). This finding is not in line with the results from study 1, yet it appears reasonable considering the game design of *fl0w*, as players are required to make an effort to chase the special food item of their liking in order to move one layer up or down (which affects the difficulty level accordingly). That argument can be supported with the results from the *mental demand*, and the *temporal demand* dimensions, which show higher mean scores for the *embedded* condition and for which the ANOVA returns a trend ($F(2,34) = 3.24$, $p = .051$, gen. $\eta^2 = .04$, Mauchly not sig.) in the case of *temporal demand*. Pairwise comparison hints at the contrast between *embedded* and *auto* ($p = .069$) as the main cause.

STUDY 2 (TLX)	**embedded** [M (SD)]:	**menu** [M (SD)]:	**auto** [M (SD)]:
physical demand	2.36 (1.03)	2.7 (1.07)	2.72 (.89)
performance	13.87 (3.45)	13.81 (3.15)	14.03 (2.77)
effort	8.28 (3.23)*	5.97 (2.36)*	6.76 (2.69)
frustration	4.52 (2.25)*	3.8 (1.86)	3.18 (1.94)*
mental demand	8.52 (3.82)	7.91 (4.55)	7.56 (4.72)
temporal demand	8.67 (3.36)	7.37 (3.58)	7.17 (3.34)

Table 4: The mean (M) and standard deviation (SD) results of the TLX in study 2. All items were recorded on 20 pt. Likert scales (range 1 to 20). Sig. effects are indicated with (*).

The ranking of all conditions (cf. Figure 7) in comparison after completion of the trials does not show significant results on the Friedman Rank Sum Test ($\chi^2(2) = 3.44$, $p = .18$), although *embedded* received the highest total score and an Exact Wilcoxon-Pratt Signed-Rank Test comparing *embedded* with *auto* shows a trend ($Z = 1.82$, $p = .08$, $r = .3$; suggesting a medium effect size).

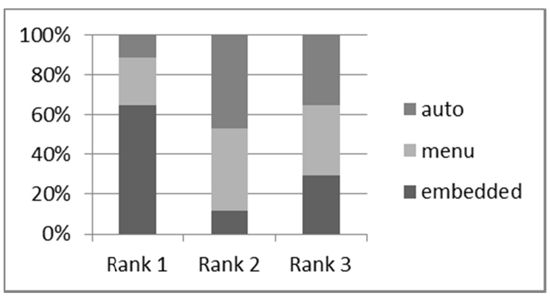

Figure 7: Ranking by participants regarding preference in study 2 (with the game *fl0w*).

Study 2 repeated the specific player experience items from study 1 almost exactly. We confirmed an effect on *autonomy*. There are significant differences between study 1 and 2 in the TLX dimensions *effort* and *frustration*, which are supported by differences in means on the *mental demand* and *temporal demand* dimensions. If interpreted as a classical "task load index", these measures put *embedded* in a worst position regarding usability. For game design choices, high task loads arguably need not necessarily be seen as adverse, because – given the right circumstances – overcoming challenges in a game generates positive player experiences (the similarity in most PENS and IMI player experience scales may be interpreted to support this argument). At the same time, frustrating controls can hinder game experiences. Either way, game designers can benefit from an awareness of possible frustrations triggered by their game difficulty choice implementations.

DISCUSSION

In this paper, we presented two studies on the impact of game difficulty choice interaction modes on player experience, focusing on potential effects on autonomy and presence/immersion for which outcome expectations were discussed based on related work.

We found evidence for a reduced autonomy needs satisfaction if no explicit difficulty choice is present, which appears in line with the expectations given the presence of manual difficulty choices in the other conditions. However, we could neither find a significant impact on presence / immersion in the case of non-embedded difficulty choices nor on any other player experience measure (with the reasonable exception of TLX measures in study 2), indicating that the differences in perceived autonomy did not have any strong impact on the overall enjoyment, motivation and affect; or in short on most of the game experience. The significant differences and trends measured on the TLX dimensions mentioned in the analysis of study 2 point towards the embedded difficulty choice being a potential source of frustration, due to the considerable effort that it required when seeking out the special food items that are not always in the same position and not always in close reach, thereby breaking interaction best practices when considered from a usability point of view. While further investigations about how players perceive difficulty choices in the forms of game mechanical elements and settings interfaces could help clarify such assumptions, the notable closeness of all modalities on most player experience measures indicates that the impact of the game difficulty selection modality on the overall game experience is likely to be small enough to allow for other considerations to guide game designers' choices of modality. With complex motion-based games for health, for example, where large numbers of sensible difficulty-related variables can be influenced by difficulty adjustments, this suggests that a strong role of DDA / adaptivity could lead to player experiences that are at least comparable to those that can be achieved when notable manual difficulty choices are present, even if they were well embedded.

While elements of choice have been linked to increased cognitive and affective engagement in learning scenarios [20], the suggested link to game enjoyment may not have results that are as large as an intuitive theoretical understanding suggests. The strong homogeneity in results between studies 1 and 2 suggests that these negative results were not accidental for the genre of casual games, and the similarities in means across conditions in the measures without significant findings indicate that strong effects on the overall player experience are unlikely to exist. The comparison with the motion-based group in study 1 serves as an indicator that the employed measures did pick up on relevant player experience factors, adding to our confidence in interpreting the large number of homogenous results.

The TLX as a classic demand measure was included because it has been used in the context of games by other researchers [24,38]. Our study provides further evidence for the sensitivity of the separate dimensions in game user research. While it requires careful interpretation in games, as desirable challenges may be part of the game design, it can be seen as an asset in the young area of game user research, where reliable psychometrics are still rare.

Limitations & Future Work

A number of limitations should be noted. Explicit feedback to an adaptive system is an unusual game design element. A further study that investigates the framing of "pre-play settings" (e.g. settings menu) vs. "post-play feedback" might thus deliver interesting additional insights. However, from our observation and the user responses during the interviews, we gather that the menu-feedback condition was seen as being similar to typical difficulty settings that are usually accessible through a main menu. It is also not clear in how far experiences related to flow are triggered in short episodes of casual game-play, although other researchers suggest that flow can emerge from a broad range of durations of captivating activities [11,17]. More highly immersive, longer-term gaming sessions might lead to different reactions to the offered difficulty choice modes.

A number of specific game design choices that were made when preparing the games and setup offer opportunities for further investigations in alternative design choices. The point of intervention for the DDA, for example, was set to be between rounds for study 1, whereas continuous or sparser adjustments are also feasible. Likewise, if an adaptive mechanism is present, feedback can strongly influence the DDA settings, adjusting not only the specific game mechanical variables, but also the estimated optimal performance thresholds for an individual player (moving from fixed threshold DDA to an individual solution). Other feedback mechanisms, such as (implicit) general affective feedback, more specific technical feedback, or feedback on multiple dimensions, may be taken into account, and other intervals for providing feedback (or other modalities to express volition) could be evaluated. Because both games were "casual", many other game types offer opportunities for further investigations on game difficulty choices. Due to the differences between the embedded difficulty choice modalities in study 1 and 2 (i.e., it was always available yet required effort to attain in study 2, whereas it was only available once per game round in study 1), further studies might expand on our results by focusing on frequency and ease of access as independent variables.

CONCLUSION

We presented an investigation of the impact of difficulty adjustments that were performed either through manual control via a classic selection menu, through an embedded difficulty choice, or with an automatic heuristics-based DDA system, on game user experience. With prior considerations regarding the impact based largely on flow theory, we presented a transfer of the expectations to the dimensions measured by validated game user research tools rooted in self-determination theory. The expected effect on autonomy was observed in the data based on a study with a game of our own design, however, no impact on presence / immersion could be found, which contradicted our outcome expectations. In the interviews, many players expressed that they prefer the presence of manual difficulty choices in games, yet our participants were more likely to be positive about DDA than to be negative about it and did not remark negatively on the presence of DDA. This ambiguity, together with a lack of notable differences on any resulting game experience measure besides the PENS dimension of autonomy, prompted us to repeat the study design with the game *fl0w*. Because Chen employed the latter to argue for the benefits of his concept of embedded (player-oriented) difficulty choices, we aimed to double-check for potential biases introduced with specific game design decisions of our game. The findings were largely repeated, including the surprisingly similar means across conditions on almost all game experience dimensions. Due to the repeated absence of effects in both studies, we find the evidence to be worth reporting. Significant differences in the TLX dimensions of effort and frustration in study 2 were found; these can be explained by game design aspects of the *embedded* mode in *fl0w*, which requires players to manually seek out special game objects in order to influence the game difficulty. Since no game motivation related measure besides autonomy showed an impact between conditions in either study, we conclude that in practical (casual) game design, all versions could lead to a very similar game experience.

Our findings suggest that the game difficulty choice interaction mode in casual games might play a minor role compared to other game design choices. These could therefore be prioritized by game designers. Letting other considerations influence the choice of game difficulty interaction modality, or opting for common and simple difficulty selection menus, appear to be reasonable choices. In terms of game user research and arguments on game design based in motivation theory, we could only partially confirm our expectations, which appeared firmly rooted in theory and intuitively logical. Hence, we can only support the call for empirical investigations of game design theory so that designers are enabled to make more certain, well-informed, and detailed decisions.

ACKNOWLEDGMENTS

We thank the *GRAND NCE*, the *Klaus Tschira Stiftung*, as well as the *Federal Ministry of Education and Research, Germany* (BMBF) for funding, and members of the *Interaction Lab* at the University of Saskatchewan, the *Digital Media Lab* at the University of Bremen, the original THYFTHYF developer team, as well as all study participants, Rodrigo Vicencio-Moreira, Jenny Cramer, Jenova Chen, and *thatgamecompany* for their kind support.

REFERENCES

1. Ernest Adams. 2010. *Fundamentals of Game Design*. New Riders.
2. Gazihan Alankus, Amanda Lazar, Matt May, and Caitlin Kelleher. 2010. Towards customizable games

for stroke rehabilitation. *Proceedings of the 28th international conference on Human factors in computing systems*, ACM, 2113–2122. http://doi.org/10.1145/1753326.1753649

3. Christine Bailey and Michael Katchabaw. 2005. An experimental testbed to enable auto-dynamic difficulty in modern video games. *Proceedings of the 2005 GameOn North America Conference*, 18–22. Retrieved September 21, 2015 from http://mmm.csd.uwo.ca/faculty/katchab/pubs/gameonna2005_add.pdf
4. Roger Bakeman. 2005. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods* 37, 3: 379–384. http://doi.org/10.3758/BF03192707
5. Gary Bente. 2009. Making the implicit explicit. Embedded measurement in serious games.
6. Max Birk and Regan L. Mandryk. 2013. Control your game-self: effects of controller type on enjoyment, motivation, and personality in game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 685–694. http://doi.org/10.1145/2470654.2470752
7. Max V. Birk, Regan L. Mandryk, Matthew K. Miller, and Kathrin M. Gerling. 2015. How Self-Esteem Shapes Our Interactions with Play Technologies. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, ACM, 35–45. http://doi.org/10.1145/2793107.2793111
8. Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2008. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, ACM, 13–17. http://doi.org/10.1145/1457199.1457203
9. Darryl Charles and Michaela Black. 2004. Dynamic player modeling: A framework for player-centered digital games. *Proc. of the International Conference on Computer Games: Artificial Intelligence, Design and Education*, 29–35. Retrieved November 26, 2012 from http://research.rmutp.ac.th/paper/cu/DynamicPlayerModelling.pdf
10. Jenova Chen. 2006. *fl0w - source code*. Retrieved September 25, 2015 from http://www.thatgamecompany.com/forum/viewtopic.php?p=4255&sid=5af8dfad64831dfb41b7bc4e66c1a165#p4255
11. Jenova Chen. 2007. Flow in games (and everything else). *Commun. ACM* 50, 4: 31–34. http://doi.org/10.1145/1232743.1232769
12. Jenova Chen. Flow in Games. Retrieved September 19, 2011 from http://jenovachen.com/flowingames/Flow_in_games_final.pdf
13. Ben Cowley, Darryl Charles, Michaela Black, and Ray Hickey. 2008. Toward an Understanding of Flow in Video Games. *Comput. Entertain.* 6, 2: 20:1–20:27. http://doi.org/10.1145/1371216.1371223
14. Chris Crawford. 1984. *The Art of Computer Game Design*. Osborne/McGraw-Hill.
15. John R. Crawford and Julie D. Henry. 2004. The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology* 43, 3: 245–265.
16. Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York.
17. Mihaly Csikszentmihalyi and K. Rathunde. 1992. The measurement of flow in everyday life: toward a theory of emergent motivation. Nebraska Symposium on Motivation. Nebraska Symposium on Motivation 40: 57–97.
18. Anders Drachen, Alessandro Canossa, and Georgios N. Yannakakis. 2009. Player modeling using self-organization in tomb raider: underworld. *Computational Intelligence and Games*, 2009. CIG 2009. IEEE Symposium on, 1–8.
19. David M. Erceg-Hurn and Vikki M. Mirosevich. 2008. Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist* 63, 7: 591–601. http://doi.org/10.1037/0003-066X.63.7.591
20. Terri Flowerday and Gregory Schraw. 2003. Effect of Choice on Cognitive and Affective Engagement. *The Journal of Educational Research* 96, 4: 207–215. http://doi.org/10.1080/00220670309598810
21. Sandro Franceschini, Simone Gori, Milena Ruffino, Simona Viola, Massimo Molteni, and Andrea Facoetti. 2013. Action Video Games Make Dyslexic Children Read Better. *Current Biology*. http://doi.org/10.1016/j.cub.2013.01.044
22. Yue Gao and Regan L. Mandryk. 2012. The Acute Cognitive Benefits of Casual Exergame Play. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, 1863–1872.
23. Kathrin Maria Gerling, Frank Paul Schulte, Jan Smeddinck, and Maic Masuch. 2012. Game Design for Older Adults: Effects of Age-Related Changes on Structural Elements of Digital Games. In *Entertainment Computing - ICEC 2012*, Marc Herrlich, Rainer Malaka and Maic Masuch (eds.).

Springer Berlin Heidelberg, 235–242. Retrieved January 18, 2013 from http://link.springer.com/chapter/10.1007/978-3-642-33542-6_20

24. Kathrin M. Gerling, Kristen K. Dergousoff, and Regan L. Mandryk. 2013. Is Movement Better?: Comparing Sedentary and Motion-based Game Controls for Older Adults. *Proceedings of Graphics Interface 2013*, Canadian Information Processing Society, 133–140. Retrieved September 21, 2015 from http://dl.acm.org/citation.cfm?id=2532129.2532153
25. Kathrin M. Gerling, Regan L. Mandryk, Matthew Miller, Michael R. Kalyn, Max Birk, and Jan D. Smeddinck. 2015. Designing Wheelchair-Based Movement Games. *ACM Trans. Access*. Comput. 6, 2: 6:1–6:23. http://doi.org/10.1145/2724729
26. Kathrin M. Gerling, Matthew Miller, Regan L. Mandryk, Max Birk, and Jan Smeddinck. 2014. Effects of Balancing for Physical Abilities on Player Performance, Experience and Self-Esteem in Exergames. *CHI'14: Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems*. Retrieved February 11, 2014 from https://hci.usask.ca/uploads/331-paper122.pdf
27. Kiel M. Gilleade and Alan Dix. 2004. Using frustration in the design of adaptive videogames. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology, ACM*, 228–232. http://doi.org/10.1145/1067343.1067372
28. Hamilton A. Hernandez, T. C. Graham, Darcy Fehlings, et al. 2012. Design of an exergaming station for children with cerebral palsy. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, 2619–2628. Retrieved May 24, 2013 from http://dl.acm.org/citation.cfm?id=2208652
29. Johan Huizinga. 2008. *Homo ludens: proeve eener bepaling van het spel-element der cultuur*. Amsterdam University Press.
30. Robin Hunicke. 2005. The case for dynamic difficulty adjustment in games. *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 429–433.
31. Marshall G. Jones. 1998. Creating Electronic Learning Environments: Games, Flow, and the User Interface. Retrieved September 17, 2015 from http://eric.ed.gov/?id=ED423842
32. Arvid Kappas. 2010. Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. *IEEE Transactions on Affective Computing*, 1, 1: 38–41.
33. Pamela M. Kato. 2012. Evaluating Efficacy and Validating Games for Health. *Games for Health Journal* 1, 1: 74–76. http://doi.org/10.1089/g4h.2012.1017
34. K. Kuikkaniemi, T. Laitinen, M. Turpeinen, T. Saari, I. Kosunen, and N. Ravaja. 2010. The influence of implicit and explicit biofeedback in first-person shooter games. *Proceedings of the 28th international conference on Human factors in computing systems*, 859–868. Retrieved August 17, 2012 from http://dl.acm.org/citation.cfm?id=1753453
35. E McAuley, T Duncan, and V V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1: 48–58.
36. Giovanni B. Moneta. 2012. On the Measurement and Conceptualization of Flow. In *Advances in Flow Research*, Stefan Engeser (ed.). Springer New York, 23–50. Retrieved September 17, 2015 from http://link.springer.com/chapter/10.1007/978-1-4614-2359-1_2
37. A. Quilici. 1994. Forming user models by understanding user feedback. *User Modeling and User-Adapted Interaction* 3, 4: 321–358.
38. Leonard Reinecke, Ron Tamborini, Matthew Grizzard, Robert Lewis, Allison Eden, and Nicholas David Bowman. 2012. Characterizing Mood Management as Need Satisfaction: The Effects of Intrinsic Needs on Selective Exposure and Mood Repair. *Journal of Communication* 62, 3: 437–453. http://doi.org/10.1111/j.1460-2466.2012.01649.x
39. Scott Rigby and Richard Ryan. 2007. *The Player Experience of Need Satisfaction(PENS): an applied model and methodology for understanding key components of the player experience*.
40. Scott Rigby and Richard Ryan. 2011. *Glued to Games: How Video Games Draw Us In and Hold Us Spellbound*. Praeger.
41. Ute Ritterfeld, Michael Cody, and Peter Vorderer. 2009. *Serious Games: Mechanisms and Effects*. Routledge.
42. Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4: 344–360. http://doi.org/10.1007/s11031-006-9051-8
43. Katie Salen and Eric Zimmerman. 2003. *Rules of play: game design fundamentals*. MIT Press, Cambridge, Mass.
44. Rosa García Sánchez, Alasdair G. Thin, Jannicke Baalsrud Hauge, et al. 2012. Value Propositions for Serious Games in Health and Well-Being. In *Serious*

Games Development and Applications, Minhua Ma, Manuel Fradinho Oliveira, Jannicke Baalsrud Hauge, Heiko Duin and Klaus-Dieter Thoben (eds.). Springer Berlin Heidelberg, 150–157. Retrieved July 5, 2013 from http://link.springer.com/chapter/10.1007/978-3-642-33687-4_12

45. Thomas B. Sheridan. 2001. Rumination on automation, 1998. *Annual Reviews in Control* 25: 89–97. http://doi.org/10.1016/S1367-5788(01)00009-8

46. Jan David Smeddinck, Marc Herrlich, and Rainer Malaka. 2015. Exergames for Physiotherapy and Rehabilitation: A Medium-term Situated Study of Motivational Aspects and Impact on Functional Reach. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 4143–4146. http://doi.org/10.1145/2702123.2702598

47. Jan Smeddinck, Sandra Siegel, and Marc Herrlich. 2013. Adaptive Difficulty in Exergames for Parkinson's disease Patients. *Proceedings of Graphics Interface 2013*.

48. Penelope Sweetser and Peta Wyeth. 2005. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.* 3, 3: 3–3. http://doi.org/10.1145/1077246.1077253

49. Jane Webster and And Others. 1993. The Dimensionality and Correlates of Flow in Human-Computer Interactions. *Computers in Human Behavior* 9, 4: 411–26.

50. Philip Zigoris and Yi Zhang. 2006. Bayesian adaptive user profiling with explicit & implicit feedback. *Proceedings of the 15th ACM international conference on Information and knowledge management*, ACM, 397–404. http://doi.org/10.1145/1183614.1183672