# Balancing Group Exergames in Virtual Reality: Fairness through Dynamic Difficulty Adaptation

Tamara Mantz[1], Laura M. König[2], Anke V. Reinschluessel[3], Tiare Feuchtner[3], Martina Kanning[3] and Hanna Hauptmann[1]

[1]*Utrecht University, Utrecht, The Netherlands*
[2]*University of Vienna, Vienna, Austria*
[3]*University of Konstanz, Konstanz, Germany*

**Abstract**

Competitive exergaming is a promising way to encourage physical activity. However, in group exergames, differences in fitness can lead to anxiety, fear of failure, or frustration. Dynamic difficulty adaptation (DDA) might mitigate these effects. However, DDA might lead to overbalancing or diminishing self-efficacy. To investigate the effects of DDA, we designed an exergame based on focus group input. Our within-subject study with 14 participant pairs compares an adaptive and a non-adaptive version of our game. We show that perceived fairness was significantly higher in the adaptive game while perceived competence and motivation showed no differences. Finally, pair interviews revealed that the adaptive game fostered motivation by creating a close game.

**Keywords**

exergame, personalization, physical activity, balancing, fairness

## 1. Introduction

Physical activity is an essential component of a healthy lifestyle, serving as a preventive measure against obesity [1] and cardiovascular disease [2]. However, it has been observed that up to 65% of people who start exercising are prone to dropping out [3, 4, 5]. Common reasons for discontinuation include time constraints [6] and loss of motivation. A possible solution could be exercise games (exergames). Due to their simplicity, exergames are easy to learn, suitable for a wide audience, and thus, increase motivation [7]. Especially group exergames have shown the potential to increase motivation, as players often significantly increase their physical effort to win competitive exergames [8, 9, 10, 11, 12].

However, competing with friends of different physical skill levels can lead to unbalanced games, which in turn leads to frustration of less skilled players or boredom of skilled players. Previous research has also shown that competitive computer games are more enjoyable for players if the game outcome was close, i.e., players performed similarly [13, 14]. A possible strategy for achieving such ”close games” is to individually adjust the difficulty of the game, such as by using players' heart rates as a measure for effort [15], or changing target number, distance, and size [16, 17, 18]. By including the physical effort of players in this dynamic adaptation, this effort can be rewarded by adaptive changes in game difficulty. One disadvantage of DDA is that it can decrease the self-efficacy of less skilled players and can be perceived as unfair by skilled players [17, 18]. In our study, we investigated whether dynamic difficulty adaptation (DDA) based on physical effort influences perceived fairness and game experience.

The way in which competitiveness is experienced as motivating or frustrating may also depend on personal characteristics such as gender or social comparison traits [19]. We thus measure these covariates and explore possible correlations with the perception of the game. Further, exergames have been shown to reduce the barrier for physical activities in groups, such as fear of being observed [20], as

*Joint Proceedings of the ACM UMAP Workshops 2026, UMAP 2026, June 8–11, 2026, Gothenburg, Sweden*
tamara.mantz@gmail.com (T. Mantz); laura.koenig@univie.ac.at (L. M. König); anke.reinschluessel@uni-konstanz.de (A. V. Reinschluessel); tiare.feuchtner@uni-konstanz.de (T. Feuchtner); martina.kanning@uni-konstanz.de (M. Kanning); h.j.hauptmann@uu.nl (H. Hauptmann)
0009-0007-6971-5580 (T. Mantz); 0000-0003-3655-8842 (L. M. König); 0000-0002-6389-6436 (A. V. Reinschluessel); 0000-0002-9922-5538 (T. Feuchtner); 0000-0001-6471-6268 (M. Kanning); 0000-0002-6840-5341 (H. Hauptmann)

individuals are more focused on the screen [21, 22, 23]. However, this effect is yet unclear when playing group exergames in virtual reality (VR). Our study thus also investigated whether this transparency may diminish players' self-efficacy during gameplay.

In a pre-study, we derived requirements for the design of the group exergame. We chose rhythm games, as these are easy to learn due to simple game mechanics. They require general fitness and coordination and allow for comparability between players, as they need to perform the same moves when competing with each other. Consequently, a design for DDA was derived from the identified requirements. To evaluate whether the implemented solution has the desired effects, a user study was conducted. We investigate the impact of DDA on how players perceive their competence, their motivation, and the game's fairness. Our study contributes to current research by defining design requirements for group rhythm exergames in VR and investigating the effects of the resulting solution. The findings presented in this work support the positive effects of DDA on players, and our derived requirements may inform the future design of fair and motivating exergames.

## 2. Related Work

This section reviews work on (VR) exergames, balancing skill mismatches, and on games with DDA.

### 2.1. Exergames and Virtual Reality

Serious games [24, 25] can be defined as games that serve a purpose other than entertainment and offer the possibility to persuade, educate, or motivate players for a certain topic [26]. Exergames are a domain within serious games that focuses on encouraging players to be physically active. These games can often be played in the home environment and have various benefits for players: they can improve coordination and spatial awareness, motor skills, cognitive skills, and social behavior in the case of group exergames [27, 28, 27]. To keep players engaged, it is important to strike the right balance between challenges and rewards. As physical exhaustion can occur, exergames should be kept simple while avoiding physical exhaustion by adapting a game to players' skills and fatigue during gameplay [7]. Additionally, motor control in exergames should feel as natural as possible to increase perceived presence in the game [29]. Group modes, leader boards, and badges can keep players engaged over the long term [30]. Immediate feedback in exergames and the possibility to achieve smaller, incremental goals can furthermore facilitate a 'flow state' in which players are fully engaged in an activity [31]. In addition, VR exergames offer an immersive experience [32], further increasing motivation to play exergames and being physically active [33]. In a study comparing group and solo exergame play, participants who chose to play in groups played more often than participants who preferred solo play [34]. Thus, group exergames increase motivation [35] and engagement [36, 37]. However, in group games, challenges are dependent on the opponents' skills, making it hard to keep all players in a flow state. Realistic body movement during gameplay was identified as another of the aspects for engaging VR exergames [38]. Nowadays, there are multiple VR exergames commercially available, including group support, movement according to music, and gradual acquisition of skills, which have been identified as engaging VR game features [38].

### 2.2. Balancing Skill Mismatches

While commercial games match player groups by their chosen level to avoid big skill differences [39], DDA provides a way to balance games among groups with differing skills. Group exergames can be competitive or collaborative. The latter leads to higher self-efficacy, encourages pro-social behavior, and keeps players motivated to play long-term [10]. Competition, on the other hand, increases the amount of effort players put into playing exergames, making it a useful tool for increasing physical activity [8, 9, 10, 11, 12]. However, competition can lead to a decrease in motivation [9] and decline of self-efficacy [40, 17] in less skilled players while creating positive feelings for strong players [9]. In contrast, providing assistance for less skilled players creates a closer game [41, 42], which is perceived as more fun [13, 14]. Players will face appropriate challenges leading to an experience of Flow [43].

According to Mueller et al.'s framework for balancing exertion [44], implicit, in contrast to explicit, assistance can create the illusion that no players receive assistance. This can mitigate the decrease in self-esteem [17, 18] and enhance the value of a win. According to Baldwin's framework for designing group DDA for games, players want to be aware of DDA and prefer balancing methods that still require skill or effort and provide a way to catch up in the game [45, 46]. Effort can be measured either by performance or exertion [44], with exertion being represented, e.g., as heart-rate [15, 47]. However, overbalancing can occur in less fit players as they reach higher heart rates faster than fit players. Furthermore, static and dynamic balancing can be differentiated. Prior studies show that a static balancing approach in (virtual) table tennis can lead to frustration for skilled players as the added challenge made it hardly possible to win [48]. Jensen and Grønbæk investigated DDA in an exergame that incorporates physical balancing, implicit-digital and explicit-digital balancing [18]. While some participants preferred implicit assistance as they felt less exposed as a weak player, explicit balancing was perceived positively by most players. To accommodate these different preferences, Jensen and Grønbæk propose a combination of both explicit and implicit balancing, or giving the user the choice of their preferred approach. They also stated that players must be aware of the DDA for higher acceptance. Lastly, the game should balance difficulty for all players to minimize overbalancing [18].

### 2.3. Prior Exergame Design with DDA

There are several existing exergames that use DDA in their design. We want to show how our approach of including physical effort differs. Our design is largely in line with existing work in VR exergames. For example, [49] shows that elements of uncertainty and randomness can increase invested physical effort. Similarly, [50] shows that VR exergames should not only include upper body movement but can also integrate lower and full body movement for a fulfilling experience. In our own design, we stimulate full-body movement by placing action elements outside the regular action radius of the player. Other work indicates that VR exergames should provide cues on the user's movement behavior [51]. In our study, pose was not as relevant to the chosen gameplay, but auditory cues were provided for (un)successful striking actions in the game. In terms of algorithmically supporting exercise in VR games, [52] provides a detailed approach to prolong and motivate exercise in VR. Similarly, some exergames [53, 54, 55] provide algorithms for DDA based on heart rate. While we see these as potential extensions of our own, relatively limited, exertion model, this work does not take into account the social experience of our DDA approach. Similarly, some prior work has more detailed approaches to tailoring self-competition [56], but does not model the social considerations, such as fairness and social comparison, due to the competitors being ghosts of prior performances. Other works do not include any social interaction but only skill-based DDA in their game [57]. In addition to VR exergames, other technology, such as the ExerCube [58] can facilitate 3D full body exercise. They further provide adaptive workouts to a player's skill and exertion. While such technology could in the future be used to replicate our ideas, this work also does not include the social aspect of competing during a workout. In summary, there are relevant prior studies on both effective and motivating VR exergame design and adaptation algorithms based on individual traits, but the nature of competitive VR exergames with two players, including social perceptions, is missing in the literature to date.

## 3. Requirement Analysis

This study followed the design science methodology by Wieringa [59]. Two requirement focus groups discussed requirements for DDA that create a fair and motivating game. Subsequently, two design focus groups generated ideas for integrating DDA into a rhythm exergame based on the derived requirements. The analysis was done using thematic analysis [60] with a mix of inductive and deductive coding. All identified codes and derived categories, as well as the number of mentions, can be found in [61]. The collected ideas were refined and implemented as a playable exergame incorporating DDA.

### 3.1. Focus Groups Methods

In the following, we briefly describe the focus group materials, participants, and analysis methods.

**Materials:** All focus groups were audio recorded. After collecting informed consent, participants were shown a video of Beat Saber (https://www.beatsaber.com/). The design focus groups received an additional sheet with short descriptions of each requirement derived from the preliminary focus groups.

**Participants:** We conducted four focus groups, three with 4 participants each and one with 5 participants, whom we recruited through convenience sampling. The participants were either students or working (age 22-40, 10 male, 7 female). Previous research found that group exergames can induce negative feelings in less competitive players [9, 40]. We thus divided participants of the requirement focus groups into either competitive (4 participants) or non-competitive (5 participants) players according to self-judgment. The third focus group was conducted with 4 students studying Human-Computer Interaction and Game and Media Technology. The fourth focus group was held with two participants working as user experience designers, one as a software engineer, and one in artificial intelligence.

**Qualitative Analysis:** The results of the four focus groups were analyzed using a thematic analysis [60] approach to identify key themes and derive requirements for designing DDA. Preliminary codes, based on the discussion topics, included preference for DDA, static difficulty adaptation, bi-directional adaptation, handicapping skilled players (uni-directional), and decreasing difficulty for less skilled players (uni-directional). Although these codes were predefined, the analysis remained open to new patterns emerging from the data. All focus group recordings were transcribed and systematically coded by one researcher. Segments of the data were either matched to the initial codes or led to the creation of new codes where relevant insights were observed. These codes were then reviewed and grouped into broader themes representing critical considerations for designing DDA in VR exergames. The prominence of each theme was assessed on the basis of the frequency and consistency of mentions across the focus groups. Based on these themes, a set of design requirements was formulated.

### 3.2. Requirement Focus Groups

We ran two focus groups to discuss the requirements for DDA in group exergames.

**Procedure:** The participants were able to try out a VR exergame prior to the focus group in order to get a feeling for the particular environment of VR games. The chosen VR game was Beat Saber, as this study is focused on rhythm games. Participants were encouraged to openly discuss questions in a semi-structured way. Two focus groups started with a general question about how participants would solve the presented problem of mismatched skills in group mode. Afterward, they were asked to discuss the advantages and disadvantages of DDA as a solution. Participants then had to discuss the advantages and drawbacks of static vs. dynamic balancing, increasing vs. decreasing difficulty, and targeting only the strongest/weakest players vs. all players. The two focus groups discussed concrete ideas for implementing DDA in a rhythm exergame. They were asked to focus on designing the game in a way that ensures fairness for all players while also encouraging more movement during gameplay. No specific game was presented as a use case in these focus groups. The participants were divided into pairs to create a low-fidelity paper prototype in two iterations.

In the first iteration, participants were asked to design a rhythm exergame with DDA that is fair for all players. During the conduction of the focus group, it became clear that the pairs needed time to understand the requirements and brainstorm various ideas. Therefore, they asked clarifying questions and held discussions. After the iteration, the pairs were asked to present all the ideas that they had brainstormed and discuss the advantages and drawbacks. Participants were also asked which idea they preferred and for what reasons.

In the second iteration, participants were asked to create an idea that takes into account the previous discussion and the ideas of the other pair. The focus of this iteration was on encouraging players to move more during gameplay. This should mitigate the problem that players might play exergames with minimal physical effort. This particular task required more technical knowledge about what movements are possible to track with a VR headset. Therefore, the participants struggled with creating initial ideas. It was decided that the participants were allowed to discuss ideas that were presented to them by the facilitator in the entire group. These ideas were: rewarding players with higher scores if targets were hit with a larger arm swing, rewarding players if they generally made more movements, and providing players who are lagging behind the possibility to catch up by increasing the amount of movements. The participants also shared their own ideas, which were then discussed in the group. Lastly, the participants were asked how to personalize difficulty adaptation.

**Resulting Requirements for DDA:** Firstly, participants feared that the system would overbalance and, therefore, make it too easy for the weaker player and too difficult for the stronger player. Additionally, DDA could facilitate cheating. For example, players could purposefully play badly in the beginning to receive an advantage. However, increasing physical effort should be incentivized. Another concern was mentioned regarding less skilled players who may feel demotivated if they become aware of a decrease in difficulty, as this means that they are performing badly [17, 18]. Certain advantages of DDA were mentioned that can lower the probability of such disadvantages occurring. Most importantly, participants would enjoy close games. According to participant P3, "everybody gets more dragged into the middle", which creates a game where players can still catch up. Participant P4 states that by adjusting the difficulty for all players, the game can become more fun, given that players primarily play for fun rather than competing in a serious context, such as in athletic competitions. The competitive group was more open to accepting DDA in games compared to the non-competitive group, and they discussed more examples of how to implement DDA. Their biggest issue with this approach is the potential for creating an unfair game for skilled players, as these players expect that their higher skill will lead to higher rewards. However, they also mentioned that this would likely only be the case for very competitive players. The non-competitive participants preferred static difficulty adaptation over a dynamic approach, where they get matched to other players based on skill level. However, this does not accommodate the scenario where players compete against friends with higher or lower skill levels. Participants agreed that they would like *transparency* about DDA during gameplay to prevent feeling cheated by the game afterward. The participants rejected the idea of difficulty decreasing automatically for players and emphasized *player control*. Any game advantages that a player receives should be activated by the player, thus leaving the decision to decrease difficulty in the hands of the player. Furthermore, the participants would prefer subtle increases or decreases in difficulty to avoid a player getting overbalanced. Further, participants would like to see their opponent's movements as an avatar, as this can create a game that feels more fair: When players see that their opponents put effort into playing even if they are less skilled, a loss might be more accepted. Furthermore, they would like to see their friend as an avatar simply to feel more connected to them. Another requirement is that all players are able to receive bonuses when falling behind, creating a fair game, as all players could benefit from DDA.

**Final Requirements:**

**R1: Create close games:** In order to create a fair game without frustrations, it is important to balance scores between players during gameplay.

**R2: Avoid demotivation of the less skilled player by player control:** By offering opportunities that players have to actively take, they are put in control. This means that they can choose not to use the help that they receive, and therefore, they may feel less "outed" as a weak player.

**R3: Add the right amount of randomness:** Randomness can make games less predictable and more interesting. However, this must be carefully balanced, as with too much randomization, players may feel the outcome is unpredictable and detached from their effort and performance.

**R4: Make adaptations transparent:** Players should be aware that they play an adapted game and should be able to notice when they receive assistance due to their weaker performance.

**R5: Subtle adaptations:** It was mentioned that "attacks" on opponents should not be too extreme, as they should not feel disadvantaged over the player who receives assistance.

**R6: Seeing opponents as avatars:** Participants would like to see their opponents in a group game to feel like they are playing with another person rather than alone.

**R7: Both players should be able to get assistance:** To create a fair game, both players should be able to receive assistance when they are lagging behind, even if they were previously leading.

### 3.3. Design Focus Groups

Following the two focus groups on the requirements, we ran two design focus groups.

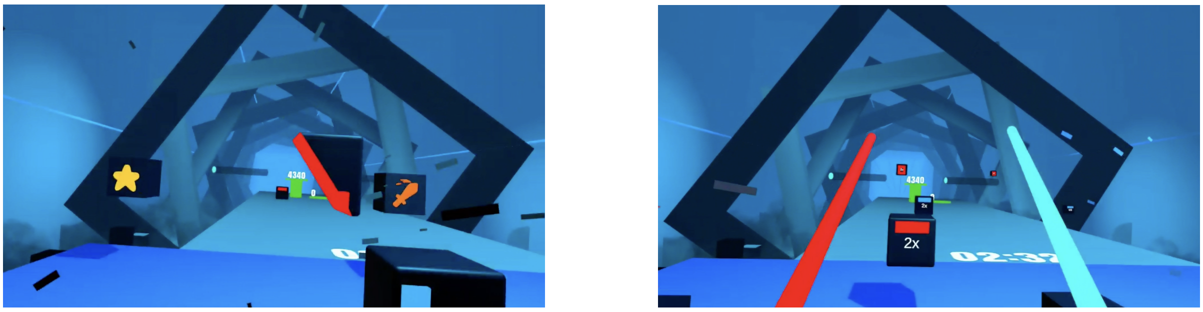
**Procedure:** In the first design, focus group participants were given a sheet listing the requirements, with the most important one highlighted. In pairs, the participants discussed ideas for a rhythm exergame, with DDA taking into account fairness. Results were presented and discussed in the group for their preference and to further develop ideas. Afterward, the group discussed how to encourage players to move more during the game. As this task required more technical knowledge about VR games, the facilitator also presented potential options for discussion, such as rewards for more movement and catching up through more movement. After conducting the first design focus group, the ideas had to be evaluated. Therefore, a second design focus group was conducted, during which participants discussed the previous ideas and narrowed them down to a single concept. Furthermore, it was discussed how to combine this idea with ways to encourage movement during gameplay.

**Proposed Design Ideas for DDA:** Participants of the design focus groups favored a combination of power-ups and attacks as an implementation for DDA in a rhythm exergame. Power-ups are ways to decrease the game's difficulty, while attacks aim to increase the game's difficulty for the opponent. The idea of being able to disrupt an opponent's game resonated strongly as it encourages interactions between players. This idea was perceived as both fun and original, e.g., P1: *And that's again a lot more interactivity, right. That's because then you're trying to prevent the other [player to score].* Moreover, participants proposed a mechanism where bonuses (power-ups and attacks) get unlocked through increased movement during gameplay, visually represented by a loading bar. This loading bar advances more rapidly for players falling behind. Thus, less skilled players unlock bonuses faster and receive more opportunities to catch up. Attacking opponents and increasing their difficulty adds unpredictability to the gaming experience. Combining power-ups with attacks seemed fair to participants because it provided an alternative for weaker players to succeed, requiring less skill but incentivizing physical activity. Furthermore, the idea was seen as a means to challenge stronger players, adding an extra feature for competition, e.g., P3: *I think the issue is [...] being worse than another person, missing the beats or something because it's all rhythm-based. Right? So maybe [...] if you're messing with someone's game, it's more fun and less [...] about the rhythm part.*

## 4. Exergame Design

The prototype, a VR exergame, is written in Unity version 2021.3.16f1, utilizing its XR (extended reality) framework along with the XR Toolkit to facilitate the integration of VR interactions. The game is based on open-source code simulating the Beat Saber gameplay. A link to the full game materials is available at https://github.com/mantzta/danceandattack. A video of the game in action is available at https://drive.google.com/file/d/1l-mWnaYaTghejBbEfge9N7CYu78KjDc9/view. Game mechanics are tailored in accordance with the requirements and findings from the above-mentioned focus groups. The prototype features a group mode realized through the Photon Pun library. The game is designed to be playable by two players, whereby players can see each other as avatars within the game, to align with *R6: Seeing opponents' effort through avatar movements* as they can look at each other during the game. The prototype was iteratively play tested by four players (friends, work colleagues). Based on these tests, we improved the game in several ways: A threshold for adaptation to prevent excessive bonuses, clearer indication of who was leading the game, and clearer scoring for correct actions. To facilitate catching up, we included three mechanisms for power-ups and attacks that adapt to score differences.

The goal of the adaptation is two-fold. The outcome of the game should be closer by making actions more difficult for the opponent or easier for the player. At the same time, the acquisition of these supportive mechanisms requires the players to perform with more physical activity, thus supporting our primary goal of a balanced competitive exercise game. The game is a rhythm exergame, where players have to hit flying targets in various directions with sabers of different colors. For example, red blocks have to be hit by the red saber on the side indicated by the block. The blocks spawn such that they need to be hit in the rhythm of the playing song.




(a) Power-up and attack blocks (b) Power-up activated indicated by a 2x multiplier

**Figure 1:** Two screenshots showcase the game's bonus design. The first screen (left) shows two bonus blocks after being spawned and moving toward the player, but before being activated. The second screen (right) shows how factors by which the score is multiplied or divided once a bonus is activated are made visible to players.

The new DDA feature is introduced as bonuses that can be caught after certain amounts of activity. The game shows a loading bar that slowly depletes when the player is inactive. Movement is encouraged, as it also contributes to filling up the player's bonus loading bar (see Fig. 1b). Once the loading bar is full, two bonus blocks are spawned, as can be seen in Figure 1a. Furthermore, the game adds a factor to the loading bar depending on a combination of the score difference between players and the mistakes a player has made. Mistakes are counted when blocks are missed or hit incorrectly (e.g., with a non-matching color or in the wrong direction). The loading bar of the player lagging behind fills up faster and thus, spawns more bonuses within a shorter time frame. After playtesting the game, an upper threshold was added for the loading factor of the bar to avoid too many bonus spawns.

There are two bonus blocks: a power-up (star block) and an attack block (sword block). These bonus blocks are spawned simultaneously on the left and right sides, such that players have to decide which block to hit - hitting both at the same time is hardly possible. Once a bonus block is hit, it activates a power-up or attack for 10 seconds. Bonus blocks are also spawned further away from normal blocks, making it harder to hit them simultaneously with normal blocks. Therefore, players need to put extra effort into activating bonuses, which is aligned with the requirement *R2: Avoid demotivation of the less skilled player by player control* by increasing physical effort while reducing the gap in performance.

Movement is further incentivized by multiplying scores by a higher factor when blocks are sliced with higher acceleration. For example, if a block is hit correctly with the least amount of acceleration possible, the player receives 10 points. With more acceleration, the player can receive up to 100 points. If players miss blocks or slice them incorrectly, they receive minus points after slicing a block.

The game includes a limited amount of randomness, which aligns with *R3: Add the right amount of randomness*: The vertical position of the bonus blocks is spawned randomly. Furthermore, bonus blocks can be spawned simultaneously with normal blocks, of which there may be more or fewer. Therefore, hitting bonus blocks can be harder depending on the timing of their appearance. Players cannot choose when bonus blocks are spawned, as these depend on prior accumulated movement.

The effectiveness of the bonus blocks depends on the difference in score between the players. Thus, three levels of bonuses can be activated, creating a low or high advantage for the player. The individual effects of each bonus type can be found in [61]. This allows for more *Subtle adaptations (R5)*. However, it is necessary to increase the effect of bonuses to balance large score differences between players.

The factors by which the score is multiplied or divided are made visible to players by indicating them on each block (see Figure 1b). The type of adaptation is explicit, as players are able to see a higher amount of bonus blocks spawned when they receive assistance in the game. They can also see when their opponents' loading bar will soon be filled and whether they scored bonuses.

Additionally, a player correctly hitting 10 blocks in a row receives a multiplier added to their score, doubling with each successive set of 10 blocks. This multiplier is written as text above each player's loading bar. When a player misses a block, the multiplier becomes the default value of 1. By showing loading bars and factors, which change the scoring, players have transparency on how their scoring changes due to DDA. This aligns with requirement *R4: Make adaptations transparent.* The dynamic aspect of the adaptation and targeting for both players fulfills requirement *R1: Create close games*, as it should continuously balance the scores of both players. Furthermore, as both players are able to receive bonuses if they invest the required physical effort, the requirement *R7: Both players need to get positive adaptations* is fulfilled. If players were to actively lag behind, they would need to conduct more physically active tasks to receive these bonuses, encouraging movement again.

For the purpose of this study, two versions of the game were implemented. One incorporates DDA with the different effects dependent on the score difference between players. The loading bar fills up faster as well for the weaker player. The game without DDA does not include changing factors for the loading bar. Thus, for each player, it is solely dependent on how much they move during gameplay. Furthermore, as there is no DDA, only the low effects can be activated when hitting bonus blocks.

## 5. User Study

To assess the effect of the DDA, we ran a within-subject user study, using Latin-square counterbalancing. We are measuring (1) perceived competence using the Player Experience Inventory (PXI) [62], (2) motivation using the *Interest/Enjoyment* and *Effort/Importance* subscales of the Intrinsic Motivation Inventory (IMI) [63], and (3) fairness using nine self-designed statements on a 7-point Likert-scale, as there are no standardized fairness questionnaires regarding gaming. Furthermore, we log the player's scores to determine how close the game was and the heart rate to assess physical effort. To see whether DDA was a better solution for players of various degrees of social comparison (c.f. [64]), we used the shortened social comparison Scale INCOM [65]. All participant duos were asked questions in a short semi-structured interview to understand how participants perceived DDA and how it can be improved.

### 5.1. Participants & Procedure

Using convenience sampling at a local university, we recruited 14 pairs (i.e., 28 participants) for our study. Participants were required to play with someone they knew and, as such, were given the autonomy to select their gaming partner. This approach was chosen because DDA is expected to be most useful when playing against friends with potential skill variations. The participants (9 female, 19 male) were between 23 and 42 years old, had a background in Information Technology, and had limited or no prior experience with rhythm games (11 never played, 8 played once or twice, 5 a few times a year, 2 a few times a month, 2 weekly). Based on the International Physical Activity Questionnaire (IPAQ), most participants (20) were highly physically active, 6 moderately active, and 2 reported low physical activity.

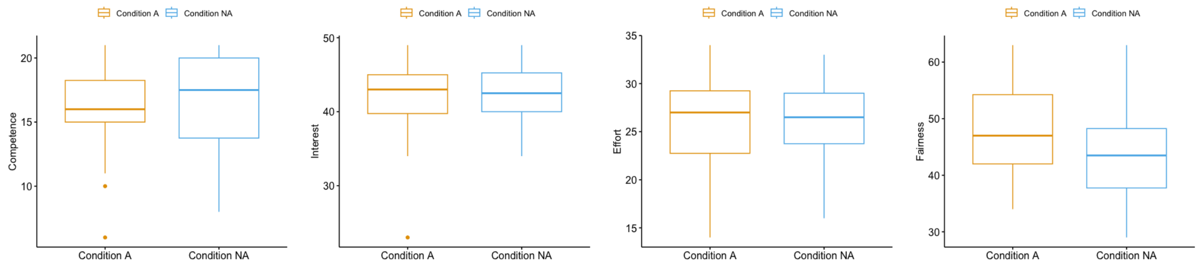
After giving informed consent, the participants put on a smartwatch heart-rate monitor, and a baseline was recorded. Next, they were introduced to the exergame and the DDA. The participants played both versions of the game (with DDA and without DDA) – whereby the players were not informed which condition it was. After each game session, they answered the questionnaires (PXI, IMI, fairness), and their heart rate was measured and recorded. Because of potential motion-induced noise, there was no continuous heart-rate recording during the gameplay. Each game lasted about 4 minutes. After completing both game sessions, they received the demographic questionnaire and the INCOM and were interviewed in pairs about their experience and perception of DDA.

### 5.2. Quantitative Findings

Following the preprocessing of the survey data, the subsequent analysis was conducted using R version 4.3.2[66]. To assess the three hypotheses corresponding to competence, motivation, and fairness, a significance level of 0.05 was adopted, and the cumulative scores of each scale for the game with DDA were compared with the game without DDA. In this analysis, the sum of all values of a Likert scale was computed; hence, it was treated as an interval. For all scales (*competence*, *interest*, *effort*, and *fairness*), the Shapiro-Wilk test was conducted to see if the distribution of data was significantly different from a normal distribution. For all scales, the p-value was above 0.05. Examining Q-Q-Plots and p-values of the Shapiro-Wilk test, it can be assumed that the data points are normally distributed for the differences between the scores of the adapted and non-adapted game in all four attributes. Thus, the parametric paired t-test was conducted. It was run as a one-tailed paired test to examine whether a game with DDA would yield significantly higher values in any attribute.

**Competence:** The one-tailed, paired t-test showed no significant difference between adapted ($M = 16.04, SD = 3.38$) and non-adapted games ($M = 16.43, SD = 3.8$), $t(27) = -0.5, p = 0.69$. The effect size was small (Cohen's $d = -0.10$), indicating a negligible difference between conditions.

**Motivation:** A one-tailed, paired t-test showed no significant difference between adapted ($M = 41.86, SD = 5.54$) and non-adapted games ($M = 42.54, SD = 4.07$) for *interest* ($t(27) = -0.84, p = 0.8$), with a small effect size, Cohen's $d = -0.16$. Similarly, there was no significant difference between adapted (M = 26, SD = 5.24) and non-adapted games ($M = 26.04, SD = 4.53$) for *effort*, $t(27) = -0.01, p = 0.51$, with a negligible effect size (Cohen's $d = -0.01$).




(a) Competence (b) Interest (c) Effort (d) Fairness

**Figure 2:** Four boxplots show the difference in perceived competence (a), interest (b), effort (c), and fairness (d) between the adapted and non-adapted game conditions.

**Fairness:** The box plot in Figure 2d shows that the adapted game has a higher median, while higher *fairness* scores were reached compared to the non-adapted game. For *fairness*, a one-tailed paired t-test showed that *fairness* was perceived significantly higher for adaped games ($M = 47.89, SD = 7.86$) compared to non-adapted games ($M = 43.82, SD = 8.03$), $t(27) = 2.33, p = 0.01$. The effect size was moderate (Cohen's $d = 0.44$).

## 5.3. Exploratory Findings

Effects based on social comparison levels, heart rate, gender, and score difference were analyzed. Due to the small sample size per compared sub-group, the results should be interpreted with caution. There were not enough participants to examine the effects of physical activity, as most participants were highly active. Similarly, the effects of rhythm game experience have not been examined, as most participants were new to such games. Some variables (e.g., heart rate, competence, interest) did not meet the conditions of a parametric test and were thus conducted as a non-parametric analysis.

The survey entailed one question with 6 statements about social comparison using a 7 Likert scale and thus, the maximum score was 6*7=42. We first conducted an ANCOVA to examine the effects of the game condition (adapted vs. non-adapted) and social comparison on *competence*, *motivation* (*effort* and *interest*), and *fairness*.

**Competence**: The ANCOVA result for *competence* showed that there was no significant effect of the game condition (adapted vs. non-adapted) on *competence* scores, $F(1, 52) = 0.178, p = 0.6746$. However, there was a marginally significant effect of *social comparison* on *competence*, $F(1, 52) = 4.018, p = 0.0502$, suggesting that higher social comparison scores may be associated with slightly higher *competence* scores. The interaction between the game condition and social comparison was not significant, $F(1, 52) = 1.557, p = 0.2177$.

**Motivation (Effort and Interest)**: The ANCOVA results for *effort* showed no significant effect of the game condition (adapted vs. non-adapted) on *effort* scores, $F(1, 52) = 0.001, p = 0.979$, nor an effect of *social comparison*, $F(1, 52) = 0.032, p = 0.859$. The interaction between the game condition and social comparison was also not significant, $F(1,52) = 0.093, p = 0.761$. For *interest*, there was no significant effect of the game condition, $F(1,52) = 0.288, p = 0.5941$, but *social comparison* had a significant effect on *interest*, $F(1,52) = 4.843, p = 0.0322$, suggesting that higher social comparison scores were associated with higher *interest* scores. This is in line with people who have a higher need for social comparison, caring more about the impact of skill differences in competition. The interaction between the game condition and social comparison was not significant, $F(1,52) = 0.010, p = 0.9200$.

**Fairness**: The ANCOVA results for *fairness* showed a significant effect of the game condition, $F(1,52) = 4.590, p = 0.0369$, indicating that *fairness* was perceived higher in the adapted game. *social comparison* also had a significant effect on *fairness*, $F(1,52) = 12.497, p = 0.000866$, suggesting that higher social comparison scores were associated with higher *fairness* ratings. This is in line with people who have a higher need for social comparison feeling safer in balanced performance comparisons. The interaction term was not significant, $F(1,52) = 2.877, p = 0.0958$.

We further extended our exploration by looking into the heart rates, as a proxy for effort, gender differences, and player scores, as a proxy for close games.

**Heart-rate:** To determine if DDA had a significant effect on heart-rate, a Kruskal-Wallis test was conducted. A one-way ANOVA was excluded as a Levene's test for testing homogeneity of variance resulted in a p-value below 0.05. The Kruskal-Wallis test showed a significant difference between heart rates ($H(2) = 19.8, p < 0.001$). Furthermore, the effect size was large with 0.22. A pairwise comparison showed that heart rate was significantly different before playing the exergame compared to after playing the adapted game ($p < 0.001$), as well as after playing the non-adapted game ($p < 0.001$). However, no significant difference was detected between adapted and non-adapted games.

**Gender:** Mixed ANOVA tests were performed for *effort* and *fairness* and the Kruskal-Wallis test was used for *competence* and *interest*. *interest* was significantly higher for female participants compared to male participants according to a Kruskal-Wallis test (H(1) = 5.70, p = 0.02) with a moderate effect size of 0.09. There was no significant difference in *interest* of male and female participants between the two conditions. No significant results could be found for *competence*, *effort*, or *fairness*. These tests may be impacted by the gender imbalance of the sample (32.1% female participants).

**Scores:** A paired one-tailed t-test showed a significant difference between adapted ($M = 3787, SD = 8746$) and non-adapted games ($M = 3946, SD = 4750$), $t(27) = 4.95, p < 0.001$. The effect size was large with $-0.917$. The median score difference is much smaller with DDA, indicating a closer game.

## 5.4. Qualitative Findings

To analyze the interview results, a thematic analysis approach was applied [60] by one researcher, similar to the focus groups. The process began with initial coding, where meaningful segments of the data were labeled with descriptive codes—either newly generated or based on recurring ideas. These codes were then reviewed and grouped into broader themes that captured key patterns across the dataset. Each code includes the number of associated statements. Participants could make multiple statements belonging to the same code. Additionally, individual statements could be assigned to more than one code. The numbers stated indicate the individual statements. As interviews were conducted in duos, we did not mark individuals separately during the coding of the interviews.

**Favorite game:** Pairs were able to discuss and compare the two games they were exposed to without knowing the explicit labels of these two versions. The statements were then attributed to the correct label by the researcher. There were 20 participants stating support for DDA, while 8 participants stated that they preferred the game without DDA. However, 6 participants of these 8 stated that they see value in DDA and would prefer it in certain situations, such as for a beginner player.

**Advantages:** It was mentioned 28 times that the game, including DDA, was fairly balanced, creating the perception that every player had a fair chance of winning. In this case, the 28 statements came from 12 duo interviews in which a total of 21 people participated, e.g., P2B: *"For me, it was nice because it motivated me more to catch up because the defense was too high in the first game [without DDA]."* Ten statements were found about the characteristic of DDA to increase challenges for skilled players. This has been seen positively, as skilled players had to put effort into playing. In the game without DDA, score differences were becoming large enough that skilled players were sure to win, which discouraged them from increasing effort. Furthermore, participants mentioned that winning the adapted game while the opponent received more bonuses was rewarding, e.g., P11A: *"So, at some point it's like, I'm gonna win. You're gonna lose. [...] Because it was really kind of neck-on-neck. You're gonna be like, oh, let me put in just a bit more effort."* This was also part of the reason why 9 statements were made that DDA increased motivation. Another reason stated was that the adapted game was more dynamic, fast-paced, and thus, more interesting to play. It was also perceived as more motivating as participants received more points through bonuses. Furthermore, participants stated 5 times that they liked earning bonuses. Even though participants may have received a lot of bonuses, the fact that they were hard to reach made them feel as if they were not assisted by the game.

**Disadvantages:** Participants experienced that the constant flow of bonus blocks created a more challenging gameplay in the adapted game. Furthermore, participants had to consciously choose which bonus block to hit to increase their own score or increase their opponent's game difficulty. This made the adapted game mentally more challenging during a fast-paced game, e.g., P7B: *"I didn't feel like it got easier. I found I was behind, and it got harder because there were more things to hit, and I was too busy trying to hit the bonuses [than] to hit the actual blocks."* Five times it was mentioned that either participants are more interested in comparing their true skills with other players or that it might become unfair to adjust games differently for each player.

## 6. Discussion

Contrary to our expectations, no significant positive effect of DDA on perceived competence was found. Some participants reported feeling less competent in the adapted game due to the perception of increased speed and difficulty. The types of attacks were perceived as too extreme, particularly in the adapted game, where attacks intensified when scores differed greatly. Therefore, difficulty must be carefully adapted to avoid overwhelming or overexerting players, which aligns with Flow theory [43] and the findings of Baldwin et al. [46]. In particular, high-performing players expressed that it should be possible to nullify the negative effects of DDA [46]. Thus, it should be possible to mitigate attacks using skill, and we derived as a design recommendation: **DDA Should Add Manageable Challenges.**

Analyzing the questionnaires revealed that participants did not feel significantly more or less motivated in the game with DDA compared to the game without DDA. It is important to note that the majority of participants were new to rhythm games in VR. Likely, both games induced a novelty effect and created high motivation. It can also be observed that there is a significant increase in heart rate after playing both exergames compared to before the experiment started. Thus, participants felt motivated to put physical effort into playing, but no significant difference in heart rate between the two games emerged. Note that participants received instructions on how to play the game, where movement was emphasized, leading to higher heart rates. During interviews, participants stated multiple times that the adapted game created a more motivating experience by creating a closer game. Participants felt that they had opportunities to win until the end and were challenged more by their opponent. It remains unclear whether DDA may lead to higher player motivation in the long term when the novelty effect has worn off. Importantly, our exploratory analysis indicates that interest might be dependent on player characteristics, with interest being generally higher for women and players with stronger social comparison orientation, which is in line with prior research [19]. **Group exergames need personalization in addition to DDA.**

The quantitative analysis results reveal a significantly lower score difference in games with DDA, indicating the successful balancing effect of the implemented DDA approach in the exergame while avoiding overbalancing. Similarly, Jensen and Grønbæk could mitigate overbalancing by incorporating a dynamic balancing method for both players [18]. Additionally, fairness was perceived significantly higher in the adapted game, possibly due to the successful creation of a close game. Due to our non-validated fairness scale, this result should be seen with caution. In our exploratory analysis, the higher fairness perception was especially the case for players with high social comparison orientation, similar to [19]. It is crucial to emphasize that both players maintain the belief that their success is a result of individual effort. Therefore, even when a game provides assistance, it should require the user's actions and be skill-dependent, similar to Baldwin et al.'s findings [46]. As a recommendation, we suggest: **DDA Should Provide Opportunities Rather Than Free Bonuses.**

The extent to which difficulty adaptation is noticeable in gameplay may also play a role in its impact on the player's competence. For instance, Jensen and Grønbæk explored a physical balancing approach which made participants feel like they were put on the spot, their sense of self-efficacy was negatively affected [18]. In contrast, in our study, participants were focused on their blocks and did not have the time to monitor their opponent due to the fast pace of the game. This may have reverted our expected impact of having full visual access to the opponent in VR. Thus, DDA may be more acceptable in games where opponents are not directly in the player's visual focus, such as in rhythm games, compared to exergames with direct player interaction, such as tennis. The fact that players are immersed in VR and represented as virtual avatars may have contributed to this as well. Therefore, we derived: **DDA Acceptance Dependent on Game Type.**

Specific participants expressed concerns about fairness, as competitive games are expected to assess skills. Interestingly, the quantitative analysis indicates that fairness was perceived to be higher in the adapted game. One factor contributing to the acceptance of DDA may be a person's level of competitiveness, which was not explicitly measured but mentioned in interviews and focus groups as a potential factor. Thus, DDA seems to be best suited for casual games where competition does not play a big role. Similarly, participants in Gonçalves et al.'s study expressed that DDA is more appropriate in casual contexts [39]. Thus, we derived that **DDA is more accepted in casual games**.

**Limitations and Future Research:** Multiple variables, such as gender, the level of weekly physical activity, and the degree of competitiveness or social comparison, could influence the dependent variable. To adequately address these factors, more participants would be necessary, and participants would need to be specifically recruited for this purpose. Additionally, future research could focus on recruiting (non-) competitive participants to understand if competitiveness influences acceptance of DDA. Also, despite their predominant IT background, VR games were fairly new to most participants, and therefore, a novelty effect was likely to occur, which may have led to high values in the motivation questionnaire and the physical effort represented by heart rate for both game modes. Furthermore, a long-term study in a realistic context could yield more realistic results. In particular, this would allow us to understand whether players are still motivated to play long-term after the novelty effect has worn off. Lastly, our presented results should be seen in the context of rhythm exergames in VR. Especially the fact that no full vision of the embodied opponent was provided and attacks/bonuses were only perceived indirectly might limit some effects, such as the fear of being observed [20]. Further research is needed to investigate whether the found requirements can be applied to other exergames and yield a positive impact on fairness. Finally, our fairness metric was not a validated scale and might thus not accurately represent the fairness perception of DDA in exergames. In future research, long-term effects of exertion should be studied to determine effectiveness, but also to avoid overexertion or physical harm.

## 7. Conclusion

This study explored dynamically adapting the difficulty of a group VR exergame for individual users as a potential solution to "level the playing field" and examined the positive effects of DDA on players' competence, motivation, and perceived fairness. Our findings support that DDA based on players' performances and score differences led to a significantly higher perception of fairness. Participants noted that the ability to catch up during the game encourages them to put effort into playing. Skilled players also feel more challenged by their opponents due to DDA. Even though some participants stated that the adapted game motivated them, there was neither a significant difference regarding motivation nor perceived competence between the adapted and non-adapted game. Based on our results, we offer recommendations for designing DDA exergames according to players' needs and expectations, such as providing opportunities, managing challenges, a casual gaming context, and different game types.

## Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References


[1] W. H. Organization, Obesity: preventing and managing the global epidemic (2000).

[2] U. Ekelund, J. Tarp, M. W. Fagerland, J. S. Johannessen, B. H. Hansen, B. J. Jefferis, P. H. Whincup, K. M. Diaz, S. Hooker, V. J. Howard, A. Chernofsky, M. G. Larson, N. Spartano, R. S. Vasan, I.-M. Dohrn, M. Hagströmer, C. Edwardson, T. Yates, E. J. Shiroma, P. Dempsey, K. Wijndaele, S. A. Anderssen, I.-M. Lee, Joint associations of accelerometer-measured physical activity and sedentary time with all-cause mortality: a harmonised meta-analysis in more than 44 000 middle-aged and older individuals, British Journal of Sports Medicine 54 (2020) 1499–1506. doi:https://doi.org/10.1136/bjsports-2020-103270.

[3] J. Annesi, Effects of a cognitive behavioral treatment package on exercise attendance and drop out in fitness centers, European Journal of Sport Science 3 (2003) 1–16. doi:https://doi.org/10.1080/17461390300073206.

[4] J. J. Annesi, Effects of computer feedback on adherence to exercise, Perceptual and motor skills 87 (1998) 723–730.

[5] R. K. Dishman, Exercise adherence: Its impact on public health, Human Kinetics, 1988.

[6] C. Fitness, L. R. Institute, 2005 Physical Activity Monitor, 2005.

[7] J. Sinclair, P. Hingston, M. Masek, Considerations for the design of exergames, in: Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia, 2007, pp. 289–295. doi:https://doi.org/10.1145/1321261.1321313.

[8] M. Nunes, L. Nedel, V. Roesler, Motivating people to perform better in exergames: Competition in virtual environments, in: Proceedings of the 29th annual ACM symposium on applied computing, 2014, pp. 970–975. doi:https://doi.org/10.1145/2554850.2555009.

[9] H. Song, J. Kim, K. E. Tenzek, K. M. Lee, The effects of competition and competitiveness upon intrinsic motivation in exergames, Computers in Human Behavior 29 (2013) 1702–1708. doi:https://doi.org/10.1016/j.chb.2013.01.042.

[10] A. M. Marker, A. E. Staiano, Better together: outcomes of cooperation versus competition in social exergaming, Games for health journal 4 (2015) 25–30. doi:https://doi.org/10.1089/g4h.2014.0066.

[11] C. Anderson-Hanley, A. L. Snyder, J. P. Nimon, P. J. Arciero, Social facilitation in virtual reality-enhanced exercise: competitiveness moderates exercise effort of older adults, Clinical interventions in aging (2011) 275–280. doi:https://doi.org/10.2147/CIA.S25337.

[12] A. L. Snyder, Virtual and live social facilitation while exergaming: Competitiveness moderates (2011).

[13] A. E. Depping, R. L. Mandryk, C. Li, C. Gutwin, R. Vicencio-Moreira, How disclosing skill assistance affects play experience in a multiplayer first-person shooter game, in: Proceedings of the 2016 CHI conference on human factors in computing systems, 2016, pp. 3462–3472. doi:https://doi.org/10.1145/2858036.2858156.

[14] R. Vicencio-Moreira, R. L. Mandryk, C. Gutwin, Now you can compete with anyone: Balancing players of different skill levels in a first-person shooter game, in: Proceedings of the 33rd annual ACM conference on human factors in computing systems, 2015, pp. 2255–2264. doi:https://doi.org/10.1145/2702123.2702242.

[15] T. Stach, T. N. Graham, J. Yim, R. E. Rhodes, Heart rate control of exercise video games, in: Proceedings of Graphics interface 2009, 2009, pp. 125–132.

[16] D. Altimira, M. Billinghurst, F. Mueller, Understanding handicapping for balancing exertion games,

in: CHI'13 Extended Abstracts on Human Factors in Computing Systems, 2013, pp. 1125–1130. doi:https://doi.org/10.1145/2468356.2468557.
[17] K. M. Gerling, M. Miller, R. L. Mandryk, M. V. Birk, J. D. Smeddinck, Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, pp. 2201–2210. doi:https://doi.org/10.1145/2556288.2556963.
[18] M. M. Jensen, K. Grønbæk, Design strategies for balancing exertion games: A study of three approaches, in: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, 2016, pp. 936–946. doi:https://doi.org/10.1145/2901790.2901843.
[19] L. M. König, M. Kanning, H. Hauptmann, T. Feuchtner, D. Arigo, Who is willing to play skill-adapted exergames? influences of sociodemographic factors and social comparison processes, Computers in Human Behavior 165 (2025) 108562. doi:https://doi.org/10.1016/j.chb.2025.108562.
[20] J. A. O'dea, Why do kids eat healthful food? perceived benefits of and barriers to healthful eating and physical activity among children and adolescents, Journal of the American Dietetic Association 103 (2003) 497–501. doi:https://doi.org/10.1053/jada.2003.50064.
[21] A. E. Staiano, S. L. Calvert, Exergames for physical education courses: Physical, social, and cognitive benefits, Child development perspectives 5 (2011) 93–98. doi:https://doi.org/10.1111/j.1750-8606.2011.00162.x.
[22] A. Quintas-Hijós, C. Peñarrubia-Lozano, J. C. Bustamante, Analysis of the applicability and utility of a gamified didactics with exergames at primary schools: Qualitative findings from a natural experiment, PloS one 15 (2020) e0231269. doi:https://doi.org/10.1371/journal.pone.0231269.
[23] B. Barron, D. Chorney, Perceptions of multiplayer exergames within pe: A mixed method study, Think Like a Mountain Comprehensive School Health in Six Priority Areas (2016) 41.
[24] C. C. Abt, Serious games, University press of America, 1987.
[25] B. Sawyer, Serious games: Improving public policy through game-based learning and simulation (2002).
[26] D. R. Michael, S. L. Chen, Serious games: Games that educate, train, and inform, Muska & Lipman/Premier-Trade, 2005.
[27] S. L. Serrano, A. Ruiz-Ariza, M. De La Torre-Cruz, E. J. M. López, Improving cognition in school children and adolescents through exergames. a systematic review and practical guide, South African Journal of Education 41 (2021). doi:https://doi.org/10.15700/saje.v41n1a1838.
[28] J. Höysniemi, Design and evaluation of physically interactive games, Tampere University Press, 2006.
[29] E. J. Lyons, Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards, Games for health journal 4 (2015) 12–18. doi:https://doi.org/10.1089/g4h.2014.0072.
[30] A. Matallaoui, J. Koivisto, J. Hamari, R. Zarnekow, How effective is "exergamification"? a systematic review on the effectiveness of gamification features in exergames (2017). doi:https://doi.org/10.24251/HICSS.2017.402.
[31] J. Nakamura, M. Csikszentmihalyi, Flow theory and research, Handbook of positive psychology 195 (2009) 206. doi:https://doi.org/10.1007/s10055-017-0320-5.
[32] Á. Tóth, E. Lógó, The effect of gamification in sport applications, in: 2018 9th IEEE international conference on cognitive infocommunications (coginfocom), IEEE, 2018, pp. 000069–000074. doi:https://doi.org/10.1109/CogInfoCom.2018.8639934.
[33] N. N. Nor, M. S. Sunar, A. Y. Kapi, User experience of gamified virtual reality (vr) in sport: A review, Science and Technologies for Smart Cities: 5th EAI International Summit, SmartCity360, Braga, Portugal, December 4-6, 2019, Proceedings (2020) 440–449. doi:https://doi.org/10.1007/978-3-030-51005-3_36.
[34] M. D. Kaos, R. E. Rhodes, P. Hämäläinen, T. N. Graham, Social play in an exergame: how the need to belong predicts adherence, in: Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–13. doi:https://doi.org/10.1145/3290605.3300660.

[35] A. E. Staiano, A. A. Abraham, S. L. Calvert, Adolescent exergame play for weight loss and psychosocial improvement: a controlled physical activity intervention, Obesity 21 (2013) 598–601. doi:https://doi.org/10.1002/oby.20282.

[36] A. L. Martin-Niedecken, Designing for bodily interplay: engaging with the adaptive social exertion game" plunder planet", in: Proceedings of the 17th ACM conference on interaction design and children, 2018, pp. 19–30. doi:https://doi.org/10.1145/3202185.3202740.

[37] N. Bianchi-Berthouze, W. W. Kim, D. Patel, Does body movement engage you more in digital game play? and why?, in: Affective Computing and Intelligent Interaction: Second International Conference, ACII 2007 Lisbon, Portugal, September 12-14, 2007 Proceedings 2, Springer, 2007, pp. 102–113. doi:https://doi.org/10.1007/978-3-540-74889-2_10.

[38] N. Farič, L. Smith, A. Hon, H. W. Potts, K. Newby, A. Steptoe, A. Fisher, et al., A virtual reality exergame to engage adolescents in physical activity: Mixed methods study describing the formative intervention development process, Journal of medical Internet research 23 (2021) e18161. doi:https://doi.org/10.2196/18161.

[39] D. Gonçalves, D. Barros, P. Pais, J. a. Guerreiro, T. Guerreiro, A. Rodrigues, The trick is to stay behind?: Defining and exploring the design space of player balancing mechanics, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024. URL: https://doi.org/10.1145/3613904.3642441. doi:10.1145/3613904.3642441.

[40] A. Macvean, J. Robertson, Understanding exergame users' physical activity, motivation and behavior over time, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013, pp. 1251–1260. doi:https://doi.org/10.1145/2470654.2466163.

[41] S. Hwang, A. L. J. Schneider, D. Clarke, A. Macintosh, L. Switzer, D. Fehlings, T. N. Graham, How game balancing affects play: Player adaptation in an exergame for children with cerebral palsy, in: Proceedings of the 2017 Conference on Designing Interactive Systems, 2017, pp. 699–710. doi:https://doi.org/10.1145/3064663.3064664.

[42] A. MacIntosh, L. Switzer, S. Hwang, A. L. J. Schneider, D. Clarke, T. N. Graham, D. L. Fehlings, Ability-based balancing using the gross motor function measure in exergaming for youth with cerebral palsy, Games for Health Journal 6 (2017) 379–385.

[43] M. Csikszentmihalyi, Flow: The Psychology of Optimal Experience, Harper Perennial, 1990.

[44] F. Mueller, F. Vetere, M. Gibbs, D. Edge, S. Agamanolis, J. Sheridan, J. Heer, Balancing exertion experiences, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, pp. 1853–1862. doi:https://doi.org/10.1145/2207676.2208322.

[45] A. Baldwin, D. Johnson, P. Wyeth, P. Sweetser, A framework of dynamic difficulty adjustment in competitive multiplayer video games, in: 2013 IEEE International Games Innovation Conference (IGIC), 2013, pp. 16–19. doi:https://doi.org/10.1109/IGIC.2013.6659150.

[46] A. Baldwin, D. Johnson, P. Wyeth, Crowd-pleaser: Player perspectives of multiplayer dynamic difficulty adjustment in video games, in: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 326–337. doi:https://doi.org/10.1145/2967934.2968100.

[47] A. T. Bayrak, R. Kumar, J. Tan, D. AhMu, J. Hohepa, L. A. Shaw, C. Lutteroth, B. C. Wünsche, Balancing different fitness levels in competitive exergames based on heart rate and performance, in: Proceedings of the 29th Australian Conference on Computer-Human Interaction, 2017, pp. 210–217. doi:https://doi.org/10.1145/3152771.3152794.

[48] D. Altimira, F. Mueller, G. Lee, J. Clarke, M. Billinghurst, Towards understanding balancing in exertion games, in: Proceedings of the 11th conference on advances in computer entertainment technology, 2014, pp. 1–8. doi:https://doi.org/10.1145/2663806.2663838.

[49] W. Xu, H.-N. Liang, K. Yu, N. Baghaei, Effect of gameplay uncertainty, display type, and age on virtual reality exergames, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–14. doi:https://doi.org/10.1145/3411764.3445801.

[50] S. Cmentowski, S. Karaosmanoglu, L. E. Nacke, F. Steinicke, J. H. Krüger, Never skip leg day again: Training the lower body with vertical jumps in a virtual reality exergame, in: Proceedings of the

2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–18. doi:10.1145/3544548.3580973.

[51] S. Karaosmanoglu, M. Wegner, S. Cmentowski, F. Steinicke, Move, react, repeat! the role of continuous cues in immersive exergames, Proceedings of the ACM on Human-Computer Interaction 8 (2024) 1–23. doi:10.1145/3677091.

[52] X. Li, Y. Chen, Auto-hierarchical data algorithm: focus on increasing users' motivation and duration in virtual reality, in: 2020 5th IEEE International Conference on Big Data Analytics (ICBDA), IEEE, 2020, pp. 150–153. doi:https://doi.org/10.1109/ICBDA49040.2020.9101254.

[53] R. Lima, M. A. Ahmad, H. Sousa, S. B. i Badia, Adaptive control of cardio-respiratory training in a virtual reality hiking simulation: A feasibility study., in: BIOSIGNALS, 2022, pp. 91–99. doi:https://doi.org/10.5220/0011004400003123.

[54] A. L. Martin-Niedecken, U. Götz, Go with the dual flow: evaluating the psychophysiological adaptive fitness game environment "plunder planet", in: Joint International Conference on serious games, Springer, 2017, pp. 32–43. doi:https://doi.org/10.1007/978-3-319-70111-0_4.

[55] S. A. Pérez, A. M. Díaz, D. M. López, Personalized tracking of physical activity in children using a wearable heart rate monitor, International Journal of Environmental Research and Public Health 17 (2020) 5895. doi:https://doi.org/10.3390/ijerph17165895.

[56] A. Michael, C. Lutteroth, Race yourselves: A longitudinal exploration of self-competition between past, present, and future performances in a vr exergame, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–17. doi:https://doi.org/10.1145/3313831.3376256.

[57] W. Xu, H.-N. Liang, X. Ma, X. Li, Virusboxing: A hiit-based vr boxing game, in: Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play, 2020, pp. 98–102. doi:https://doi.org/10.1145/3383668.3419958.

[58] A. L. Martin-Niedecken, K. Rogers, L. Turmo Vidal, E. D. Mekler, E. Márquez Segura, Exercube vs. personal trainer: evaluating a holistic, immersive, and adaptive fitness game setup, in: Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–15. doi:https://doi.org/10.1145/3290605.3300318.

[59] R. J. Wieringa, Design science methodology for information systems and software engineering, Springer, 2014. doi:https://doi.org/10.1007/978-3-662-43839-8.

[60] V. Clarke, V. Braun, Thematic analysis, in: Encyclopedia of critical psychology, Springer, 2014, pp. 1947–1952.

[61] T. Mantz, Exploring the effects of difficulty adaptation in Virtual Reality Multiplayer Exergames to create a fair game between players with mismatching skills, Master's thesis, Utrecht University, 2024. URL: https://studenttheses.uu.nl/handle/20.500.12932/46318.

[62] V. V. Abeele, K. Spiel, L. Nacke, D. Johnson, K. Gerling, Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences, International Journal of Human-Computer Studies 135 (2020) 102370. doi:https://doi.org/10.1016/j.ijhcs.2019.102370.

[63] E. McAuley, T. Duncan, V. V. Tammen, Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis, Research quarterly for exercise and sport 60 (1989) 48–58. doi:https://doi.org/10.1080/02701367.1989.10607413.

[64] T. Kojić, J.-N. Voigt-Antons, S. Schmidt, L. Tetzlaff, B. Kortowski, U. Sirotina, S. Möller, Influence of virtual environments and conversations on user engagement during multiplayer exergames, in: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2018, pp. 1–3. doi:https://doi.org/10.1109/QoMEX.2018.8463370.

[65] S. M. Schneider, J. Schupp, Individual differences in social comparison and its consequences for life satisfaction: introducing a short scale of the iowa–netherlands comparison orientation measure, Social Indicators Research 115 (2014) 767–789. doi:https://doi.org/10.1007/s11205-012-0227-1.

[66] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. URL: https://www.R-project.org/.