Reframing Game Difficulty in Player-Game Interaction: Concept, Measurement, and Design

by

Zhixing Guo

Student ID Number: 1258002

A dissertation submitted to the
Engineering, Department of Engineering,
Graduate School of Engineering,
Kochi University of Technology,
Kochi, Japan

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Assessment Committee:
Supervisor: Xiangshi Ren
Co-Supervisor: Hiroki Nishino
Co-Supervisor: Yukinobu Hoshino
Shinichi Yoshida
Zixue Cheng, University of Aizu

September 2024

Abstract

Reframing Game Difficulty in Player-Game Interaction: Concept, Measurement, and Design

Human-Engaged Computing (HEC) aims to realize the synergism between humans and computers and enhance humans' capability. One of the approaches is to develop engaging computers. Video games are a kind of potentially engaging computer that helps realize the vision of HEC. Game difficulty is a critical issue in video game design and is highly concerning for designers and players. It has been reported that game difficulty impacts gameplay, player experience, and engagement. Therefore, games that are towards HEC goals should be designed based on adequate consideration of the difficulty factor.

Despite various empirical and design studies, current research on game difficulty faces three challenges. First, there is no broad consensus or clarification research on the concept of game difficulty. Therefore, different definitions and understandings are mixed in use, which compromises the clarity of research findings. Second, due to the first issue, there is no standard measuring method for the game difficulty measurement. The current quantification or measurement of game difficulty is insufficient in terms of comprehensiveness and thus lacks effectiveness. Third, these two issues further restrict the design practice of game difficulty. Especially as a promising difficulty mechanism proposed in recent years, Dynamic Difficulty Adjustment (DDA) has not yet fully achieved the effectiveness and potential it is expected to in game design.

We found that player-game interaction is a promising perspective for providing solutions to these challenges. Therefore, this dissertation reframed one player's game difficulty in both single and multiplayer games in concept, measurement, and design. The theoretical, exploratory, quantifying, and empirical studies were accordingly conducted. More specifically, through a systematic literature review, we first sorted out various concepts of game difficulty, its current measuring methods, its impacts on players, and design issues about the DDA mechanism. The interactive perspective was subsequently introduced to clarify the connotation and cause of the game difficulty. We then defined subjective game difficulty

(SGD) and objective game difficulty (OGD) separately and built an interaction model to illustrate how they occurred in the player-game interaction process.

After these theoretical efforts, we explored the relationship between SGD and OGD by experiment. We found that although OGD and SGD are two parts of the original difficulty concept, they matched partially. This finding called for an improvement in the measuring methods of both OGD and SGD. Therefore, we developed a new OGD measuring method that quantified OGD by the input time and incorrectness factors and validated the method by an experiment. For SGD, we developed a scale with six dimensions to measure SGD. The developed scale was verified for its reliability and validity in measuring SGD.

Based on these works, we focused on the design of game difficulty, especially the DDA mechanism. We found the current DDA design lacks solid theoretical fundamentals but narrowly relied on Flow theory. Therefore, based on our proposed difficulty definitions and framework, we redefined DDA and proposed a new DDA design methodology. A case study follows, which implemented our DDA design methodology to design a cognitive training game for the elderly. The results showed the effectiveness of our DDA design in enhancing participants' cognitive abilities and player experience.

The main contributions of this dissertation are three-fold: (1) Enhancing the theoretical foundations of game difficulty by clarifying its conceptual connotations. (2) Clarifying the link between concepts and measurement to propose effective methods for measuring game difficulty. (3) Exploring the game difficulty's impacts on players and rethinking the DDA mechanism to provide practical design methodology and implications for game difficulty to support the HEC game design.

The other key contributions include (i) Proposing new definitions of SGD and OGD, and an interpretive interaction model. (ii) Determining the partial matching relationship between SGD and OGD. (iii) Proposing and validating a new DDA definition and design methodology.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the background, issues, and scope of this research and provides the structure of the dissertation.

1.1 Human-Engaged Computing and Video Games

The human-computer interaction (HCI) community focuses on how humans and computers interact and how to improve that interaction through purposeful design. Four waves in the history of HCI reveal the development of the computing philosophy in this field (Ren et al., 2019). The focus of HCI changes from human factors and psychology to better human experience and well-being achieved by HCI (Harrison et al., 2007; Card, 2018; Kling and Star, 1998; Calvo and Peters, 2014). However, current computing technologies are still compromising human potential to develop their capabilities. The usage of smarter computers has been reported to diminish concentration span (Stothart et al., 2015), weaken cognitive capability (Javadi et al., 2017; Sparrow et al., 2011), and cause fatigue, stress, and depression (Thomée et al., 2011). Recently, the large language model (LLM) has become popular since it is intelligent in supporting communication, writing, and design in various contexts (Yenduri et al., 2024; Megawati et al., 2023). This development in artificial intelligence (AI) enhances efficacy and productivity but leads to issues, such as layoffs in companies and academic fraud. It is urgent to reflect on the relationship between humans and computers.

To address the antibiosis issue between humans and computers, Ren (2016) proposed and developed (Ren et al., 2019) a new conceptual framework, Human-Engaged Computing (HEC), see Fig. 1.1. HEC is the philosophical approach that aims to promote the synergism between humans and computers. Compared with past thoughts and proposals, HEC is focusing on how to realize engaged humans whose capacities are fully recognized, activated, and appropriately enhanced through synergized interaction. Therefore, developing engaging computers to enhance and complement human capacities is a critical topic under the HEC framework. Video games (games for short) have been widely recognized for their ability to engage humans. This nature of games makes them an important kind of engaging computer that helps realize the vision of HEC.

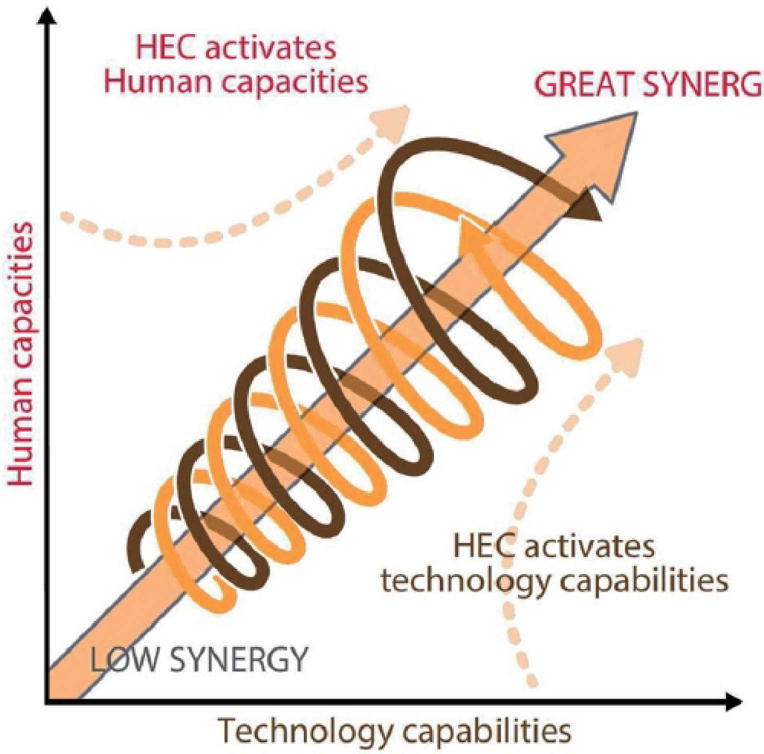

Fig. 1.1 The concept of Human-Engaged Computing (HEC) calls for the synergism between humans and computers to realize the enhancement of engaged human capacities through engaging computer development (Ren et al., 2019).

Games can be roughly divided into entertainment games and serious games based on their functions. Entertainment games gained a global market commercial value of 347 billion U.S. dollars in 2022 (Clement, 2024), while serious games for education, medicine, advertisement, and other non-entertainment goals have also received extensive attention from the academic community (Laamarti et al., 2014; Dörner et al., 2016). As a kind of interactive system, research on games is also important in the HCI community community (Tekinbas and Zimmerman, 2003; Carter et al., 2014). Corresponding to the three elements of interaction: humans, computers, and interaction, game research in HCI also has three main three topics: (i) player study (Klimmt et al., 2009), (ii) game design and development (Sykes and Federoff, 2006), and (iii) research on the player-game interaction (Caroux et al., 2015).

However, the design philosophy of video games seems to have stagnated, which is related to the lack of a unified meta-design vision. Specifically, entertainment games take sensory stimulation, narrative content, image style, gameplay, etc., as the main design objects and aim to pursue creating extreme experiences (Mekler et al., 2014). Serious games are scattered in different fields, such as medical care, marketing, health & fitness, and education. These games are designed for different specific design goals (Dörner et al., 2016) but lack important thinking about the mission of game design. What's worse, games cause the antibiosis issue between humans and computers, in which the "engagement" caused by the games does not benefit humans. People become worried about games' negative effects that lead to violence, addiction, obesity, etc. Therefore, applying HEC theory to guide the game design is urgently necessary. The introduction of the HEC perspective is promising in promoting theoretical, empirical, and practical research on games.

1.2 Video Games and Game Difficulty

Game difficulty (or game challenge) is one of the essential components in game design (Adams, 2014; Schell, 2019) and has been reported its impacts on entertainment (Bostan and Öğüt, 2009) and different serious goals (Dörner et al., 2016; Jacobs et al., 2020; Anguera et al., 2013). More importantly, it is one of the most related components that promotes human engagement in interactions (Chen, 2007). Therefore, the research scope of this dissertation is about the game difficulty in HCI and how HEC theory can guide the better design of game difficulty. To be clear, the difficulty in this dissertation basically refers to the difficulty in single and multiplayer video games for one player.

The research on game difficulty in HCI mainly focuses on the influence of game difficulty on players (Alexander et al., 2013; Burke et al., 2009) and how to design better game difficulty (Hunicke, 2005; Paraschos and Koulouriotis, 2023). However, despite the flourish of game difficulty research, there is still a lack of broad consensus on the concept of game difficulty (Dziedzic and Włodarczyk, 2018). The first issue is confusing game difficulty and game challenge concepts. "Challenge" is a more general concept (Lomas et al., 2017) and is considered a nontrivial task (Adams, 2014), a synonym for difficulty (Chen, 2007; Orvis et al., 2007), a kind of player experience (Denisova et al., 2020), a type of motivation related to pleasure (Ryan et al., 2006; Schell, 2019), etc. Even though we agree challenge is a valuable concept when discussing a design vision on creating an effort-necessary play experience (Juul, 2011), it lacks academic rigor and is more a general perception when we carefully check its connotation. In contrast, game difficulty can be more specific and refined for conceptualization.

Currently, there are still three different opinions about game difficulty concepts. Correspondingly to the three research topics in HCI, game difficulty is usually regarded as: (1) the subjective challenging experience of players (Frommel et al., 2018; Denisova et al., 2020), (2) the attributes of the game task (e.g., speed, time, etc) that causes difficulty in playing (Qin et al., 2010; Pedersen et al., 2010), and (3) the level of demand the game imposes on players' skills (Robinson, 2001; Guadagnoli and Lee, 2004; Orvis et al., 2008; Aponte et al., 2011b). The work of quantifying and measuring game difficulty also lacks accepted standards but corresponds to the mentioned three aspects: (1) players subjective feelings are measured by physiological tools (e.g., eye movement, EEG, etc.) or self-report (Chanel et al., 2011; Spiel et al., 2019; Peng et al., 2023; Ryan et al., 2006; Deci and Ryan, 1985; Vahlo and Karhulahti, 2020), (2) the difficulty-related game task attributes are quantified by the complex degree (Wheat et al., 2016; Qin et al., 2010), and (3) the demand of task on players is evaluated and predicted by the player performance, e.g., failure rate (Aponte et al., 2011b; Constant and Levieux, 2019).

Researchers have recently tried to clarify the concept of game difficulty. To clearly illustrate these concepts, Constant et al. (2017) suggested the use of subjective game difficulty (SGD) and objective game difficulty (OGD) to title the first type and the third type of difficulty. For the second type, referring to Liu and Li (2012), "task complexity" should be used to describe the aggregation of these task attributes rather than "task difficulty". Therefore, "game difficulty" is a concept misuse for this type, and "game complexity" is the more suitable concept. Referring to Dziedzic and Włodarczyk (2018), who adopted the interaction perspective and proposed that task difficulty in the game "involves the interaction between task, task performer, and context characteristics". Similarly, "game difficulty" (no matter OGD or SGD) can be considered to occur in players' interaction with games and changes over time (Caroux et al., 2015), but the "game complexity" seems produced by game design and can exist without players.

Although these efforts refine the concepts of game difficulty and promote a better understanding of these concepts, further questions emerge as this conceptual development.

1.3 Challenges in the Current Game Difficulty Research

The first issue is the lack of clear definitions of OGD and SGD. When we adopt the interaction perspective, the current definitions of the difficulty are static and hard to present the connotations. This also makes the relationship between SGD and OGD unclear. It is natural to assume that OGD and SGD match each other for one player (Constant et al., 2017) because they form the two parts of the original concept. This assumption means players would experience corresponding difficulty feelings based on how much their skills meet the game's demands. They are also mixed in use in current research (Alexander et al., 2013; Ang and Mitchell, 2017). However, research has shown a more complex result than this assumption. Hunicke (2005) found that player perception of game difficulty does not correlate with player performance. This result was also indicated by Aponte et al. (2011a), who further explained that players may have more complex evaluation patterns of SGD. Constant et al. (2017) proposed that players seem to easily underestimate OGD and become overconfident about future success, indicating that player self-efficacy (i.e., confidence on the specific task; Bandura and Wessels, 1994) is related to this mismatch. These findings challenge this assumption and require us to rethink these two kinds of game difficulty fundamentally.

Due to the limitation in definitions, current OGD and SGD measuring methods are not as effective as expected. For OGD, the function of failure probability proposed by Aponte et al. (2011b) seems to lack standard form across games. Another universal method, measuring the player's failure rate (Hocine et al., 2015; Anguera et al., 2013; Kitakoshi et al., 2020), is not effective and precise in measuring the overall OGD in one attempt (i.e., only get the result of success or failure) and the real-time OGD in the playing process. In addition, this method also cannot measure whether and how players progress in multiple failure attempts. These issues make it still difficult to effectively measure OGD for design and research but rely more on designers' practice experience or costly game testing. In comparison, SGD is directly measured by self-report and physiological measures. However, physiological measures still need to be based on the self-report data as the interpretive material, e.g., Chanel et al. (2008), while the current self-report SGD measures usually only have a few simple questions or are based on the vague challenge concept (Aponte et al., 2011a; Wheat et al., 2016; Blom et al., 2019; Vahlo and Karhulahti, 2020). These methods fail to reflect the comprehensive connotation of SGD and measure SGD ineffectively.

The research on game difficulty design is also affected by the variety of game difficulty concepts and measuring methods. Designers usually think that through designing OGD (more precisely, designers are to create specific game complexity to realize expected OGD), they can create the expected player difficulty experience (SGD) (Schell, 2019; Adams, 2014). However, if SGD does not match OGD, the impact of OGD design on SGD would be vague and incapable of evaluation. This is undoubtedly disastrous for game designers because "creating a great experience is the goal of game design" (Schell, 2019). Additionally, the assumption of natural matching between OGD and SGD has been used by default in many game difficulty studies. Most designers agree that a well-designed game with a balanced challenge is expected to encourage players to get into the Flow state (Fullerton, 2014; Schell, 2019), i.e., an ideal psychological state of being highly immersed and engaged proposed in Flow theory (Csikszentmihalyi and Csikzentmihaly, 1990).

Dynamic Difficulty Adjustment (DDA) is considered a promising game difficulty mechanism that aims to create skill-challenge balance following the guidance of Flow theory, usually by measuring the failure rate (i.e., one of the measuring methods of OGD; Zohaib, 2018). However, it should be noted that the challenge proposed by Flow theory is a "perceived challenge" (i.e., SGD) and is usually measured by self-report of the Flow state or players' competence (Csikszentmihalyi and Csikzentmihaly, 1990; Norsworthy et al., 2021; Ryan, 1982; Ryan et al., 2006). Therefore, most DDA mechanisms would lose the rationality to achieve an "SGD-related balance" by adjusting OGD without this matching relationship. According to the research, current DDA designs were not as effective as expected in creating a better player experience indeed (Alexander et al., 2013; Ang and Mitchell, 2017; Smeddinck et al., 2016; Salehzadeh Niksirat et al., 2017). Therefore, some researchers have attempted to improve it beyond the Flow (Masanobu et al., 2017). However, little attention has been paid to how to design DDA based on the understanding of the concept of game difficulty.

Furthermore, the previous findings on the impact of game difficulty on player-related design factors also become unclear. For example, it is hard to determine whether the excellent player experience depends more on a good OGD design or the specific SGD experience of the player (Juul, 2009; Smeddinck et al., 2016; Allart et al., 2017). Engagement, which refers to the player's willingness to participate in future play, is another critical issue. Player engagement was believed to be influenced by game difficulty (Lomas et al., 2017), but OGD was more considered for engagement design in present studies (e.g., Roohi et al., 2020). Self-efficacy is also highly concerned as another difficulty-related factor, especially in some serious games (Hung et al., 2014; Khalili-Mahani et al., 2020). Even though self-efficacy was reported as having a negative relationship with game difficulty (Power et al., 2020; Nuutila et al., 2021), the respective roles of OGD and SGD remain unclear (Constant et al., 2017; Constant and Levieux, 2019). These issues also cause confusion in game difficulty design. For example, the DDA designs that are based on different game difficulty understandings gain mixed results in enhancing player experience (e.g., Smeddinck et al., 2016; Wang et al., 2016; Ang and Mitchell, 2017; Xue et al., 2017; Akbar et al., 2019).

In summary, there are three main challenges in the current game difficulty research:

- Different definitions and understandings of game difficulty are mixed in use, and the conceptual development of game difficulty into OGD and SGD introduces more questions;
- No standard and effective methods of measuring OGD and SGD for the game difficulty measurement;
- Design research on game difficulty, especially the DDA design, is restricted by the lack of conceptual clarifying studies and the unclarity of the relationship and influences of OGD and SGD.

We found that player-game interaction is a promising perspective for providing solutions to these challenges. Therefore, this dissertation reframes game difficulty in player-game interaction in three aspects: concept, measurement, and design. The contributions of this dissertation are three-fold: (1) enhancing the theoretical fundamentals of game difficulty by clarifying the concepts' connotations and relationships; (2) clarifying the relationship between concept to measurement and proposing effective measuring methods for game difficulty; (3) exploring the game difficulty's impacts on players and rethinking the DDA mechanism to provide practical methodology and implications for game difficulty design and HEC game design.

The dissertation is organized as follows (see Fig. 1.2). Chapter 2 sorts out various concepts of game difficulty, its current measuring methods, its impacts on players, and design issues about the DDA mechanism. Chapter 3 builds a model to illustrate how SGD and OGD occur in the player-video game interaction and redefine them. Chapter 4 explores the relationship between SGD and OGD and reveals that they only partially match. This finding calls for an improvement in the measuring methods of both OGD and SGD. Therefore, Chapters 5-6 focus on the measurement of game difficulty and aim to provide solutions to the second challenge. Chapter 5 develops and validates a new OGD measuring method that quantifies OGD with formulas using the input time and incorrectness factors. Chapter 6 develops a new scale, SGDS, to measure the six dimensions of SGD through the standard three steps. Based on these works, Chapters 7-8 attempt to address the design issues of game difficulty, especially issues in the DDA mechanism design. Chapter 7 redefines DDA and proposes a new DDA design methodology, including a design framework and a 6-step design process. Chapter 8 validates our DDA design methodology through a case study, in which we design a cognitive training game with DDA for the elderly. Finally, Chapters 9-10 discuss the findings in the concept and measurement of game difficulty and implications for game difficulty design. Limitations, future work, and conclusion for this dissertation are also provided.

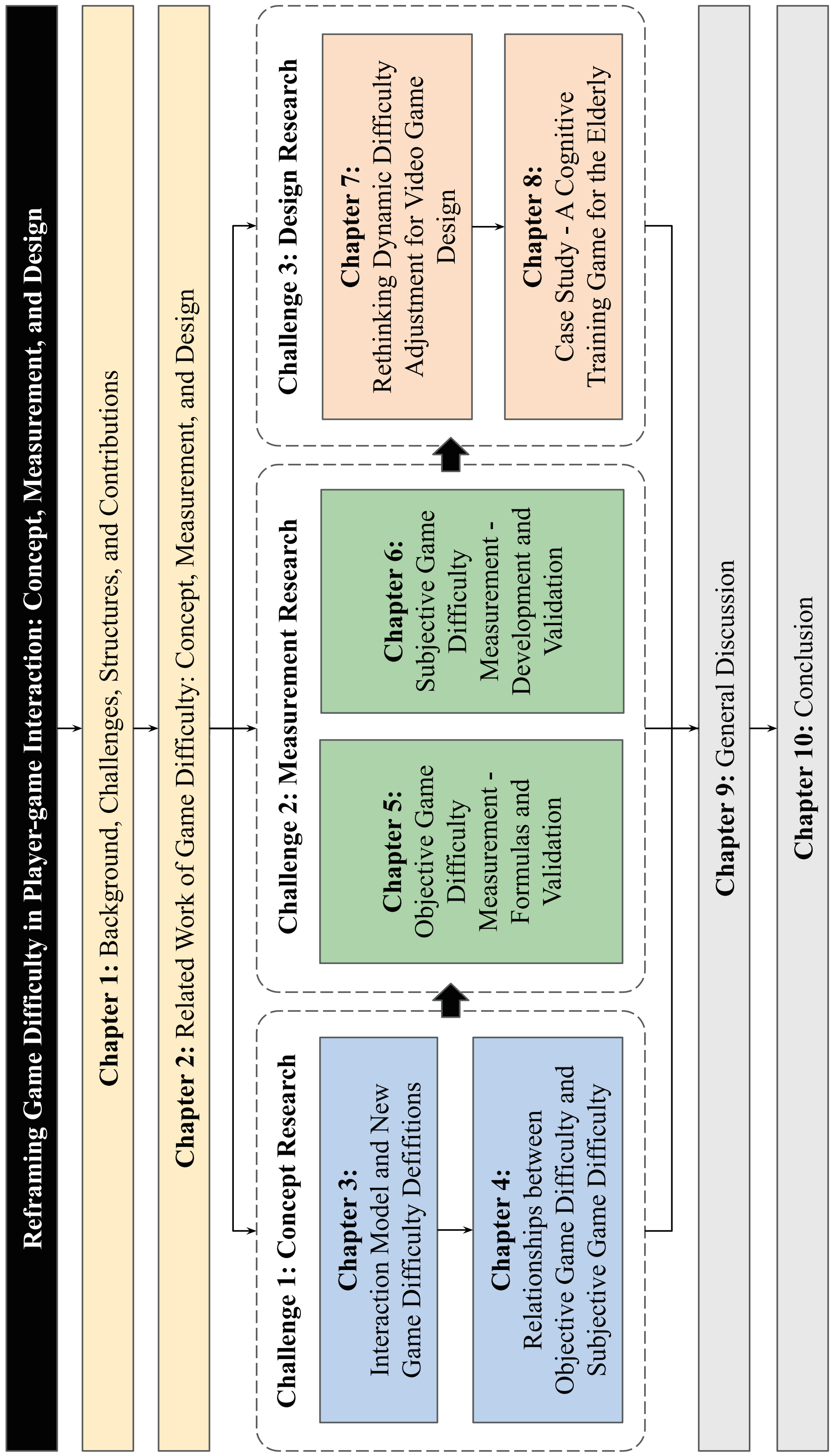

Fig. 1.2 Dissertation Outline

Chapter 2

Related Work

This chapter first reviews the development of the game difficulty concept and sorts out the measuring methods of the two difficulties: objective game difficulty (OGD) and subjective game difficulty (SGD). We further illustrate the multiple dimensions of SGD. Subsequently, this chapter introduces the research on game difficulty's impacts and the Dynamic Difficulty Adjustment (DDA) mechanism.

2.1 Game Difficulty: Original Concept and its Evolving

The concept of game difficulty originates from various fields and shows a trend toward subdivision. We elaborated on this trend and introduced the interaction perspective to understand this concept.

2.1.1 How Video Game Definition Relates to Task

Game difficulty naturally comes from the game, thus, understanding the game is helpful in clarifying the concept of game difficulty. Based on psychological and anthropological perspectives, play (or games) are a kind of human activity (Tudge and Winterhoff, 1993; Huizinga, 2014). This view is also supported by some game designers, for example, Adams (2014) proposed that "games are a type of play activity, conducted in the context of a pretended reality, in which the participant(s) try to achieve at least one arbitrary, nontrivial goal by acting in accordance with rules". Schell (2019) also believed that games are problem-solving activities.

However, with the development of video games, another perspective sees games as systems that trigger interactions between players and games. A well-known definition comes from Tekinbas and Zimmerman (2003), "A game is a system in which players engage in artificial conflict, defined by rules, leading to quantifiable outcomes." McGonigal (2011) proposed that games have four characteristics, i.e., goals, rules, feedback systems, and voluntary participation. Fullerton (2014) believes the game is a closed and formal system that allows players to engage. Based on these views, Sánchez et al. (2012) proposed that video games are considered interactive systems that use various digital devices as carriers and aim to entertain players. We agree that games are interactive systems that are designed to allow players to interact under specific game rules and goals.

The structure of goals and rules of games constitutes the task. Objective task complexity is usually used to describe the form of a task with attributes (Campbell, 1988). Liu and Li (2012) described task complexity as an objective aggregation of task characteristics of goal, input, process, time, and presentation. Therefore, the multiplicity of factors and elements of game tasks should be defined as “game complexity” rather than “game difficulty”.

2.1.2 Current Definitions of Game Difficulty

The definition of game difficulty still lacks a broad consensus in precise meaning, and it is usually mixed with the challenge concept in use. Pusey et al. (2021) suggested “difficulty” is more related to players’ failure in a task, while “challenge” is related to players’ effort toward “in-game” task success. Therefore, it is worth determining the meaning of “task difficulty” before proceeding. The term “task difficulty” is commonly used in the fields of psychology, HCI, and education. This concept refers to the level of demand the task imposes on performers and how well the performer meets the demand (Robinson, 2001; Guadagnoli and Lee, 2004; Orvis et al., 2008). Therefore, task difficulty is related to task complexity and performer ability. This view can be a valuable reference to game difficulty because it distinguishes difficulty from task characteristics and can be quantified by the performer’s performance.

Without a clear distinction, many game design studies still use attributes of game task (i.e., complexity) to quantify and design game difficulty (e.g., Wehbe et al., 2017; Hsu et al., 2007; Klimmt et al., 2009). To improve this situation, Aponte et al. (2011b) proposed that game difficulty can be represented as a probability function of player failure at a specific time. This quantifying method of difficulty relates more to the evaluation and prediction of player performance and is applied in recent studies (Sarkar and Cooper, 2019; Roohi et al., 2020). However, it should be noted that they do not provide a clear definition of game difficulty. Therefore, by referring to the definitions of task difficulty, we suggest that the difficulty in these studies refers to the objective game difficulty (OGD). We further summarize the current definition of OGD as **the level of demand the game imposes on player skills**. In this case, the degree to which players' skills meet the game's demands is rational to be quantified by player performance: a better performance (or a lower probability of failure) means a closer match between skills and demands and a lower OGD.

However, this way of defining and measuring game difficulty did not reflect the subjective difficulty feelings of the players (Aponte et al., 2011a), so more researchers have begun to deconstruct the game difficulty concept. Perceived difficulty has been used in early psychology research and can be referred to people's opinion about whether a behavior is difficult to perform or not (Trafimow et al., 2002). Li et al. (2014) proposed that game difficulty consisted of objective difficulty and player-perceived difficulty. They further argued that objective difficulty is determined by the demands of operation speed and task complexity, but perceive difficulty varies from person to person. Constant et al. (2017) proposed that SGD was a psychological construct of the player. Denisova et al. (2020) believed SGD should be described as the player's experience of challenge in different aspects. Therefore, we summarized the current definition of SGD as **the player's perception of game difficulty from the playing experience**.

2.1.3 Game Difficulty in the Interaction

Although game difficulty becomes clear conceptually, there is still a need to explain how game task demands relate to the task structure and the interaction process. If we adopt a static perspective, such demands are determined objectively by the elements of the game tasks and are, therefore, unchanging. As mentioned in Section 2.1.2, this is the classical way to quantify OGD and it has been applied in some game design research (Wehbe et al., 2017; Hsu et al., 2007; Klimmt et al., 2009). However, this view now seems to be relatively rigid since the introduction of the dynamic interaction perspective. Firstly, it needs to be clarified whether the difficulty of the game exists independently of the player, i.e., whether the game task has a level of difficulty regardless of any player interaction. As we mentioned before, the concept of game complexity is sufficient to express the game task's static and objective attributes. Game difficulty must be related to the game player (Dziedzic and Włodarczyk, 2018). Therefore, an actual interaction between game tasks and players is a necessary condition for identifying and measuring game difficulty.

More specifically, OGD and SGD are both related to the player and thus cannot be determined solely by the game task's attributes. For OGD, player skills are essential in determining the level of difficulty (Adams, 2014; Denisova et al., 2020). Game skills (or game expertise) are more specialized skills and can be acquired only by playing games (Cox et al., 2012; Lee and Heeter, 2017; Deterding, 2015; Linehan et al., 2014). However, player skills vary from player to player and will improve over the playing process through learning (Huniche and Chapman, 2005; Jennings-Teats et al., 2010; Martin, 2014; Linehan et al., 2014). OGD thus can be regarded as a changeable interaction result between player skills and game task demands (Aponte et al., 2011b; Pavlas, 2010). For SGD, players have their own complex SGD evaluation patterns, which may be based on their game experience, self-efficacy, and other cognition-related characteristics (Hunicke, 2005; Aponte et al., 2011a; Juul, 2009; Bandura and Wessels, 1994). In addition, different playing processes also affect the player's SGD evaluation (Denisova et al., 2020). Thus, SGD is also changeable among players and in different playing processes. The static perspective is weak in explaining these complex and changeable game difficulties. The dynamic perspective becomes necessary to understand OGD and SGD in each particular interaction process.

This interaction perspective from HCI provides the basic view that a specific interaction process is mainly about how human input and computer output happen and interact (Helander, 2014). Caroux et al. (2015) introduced how interaction happens between players and video games and discussed the input and output research of games. Specifically, the game-playing process can thus be regarded as the player-game interaction process in which players receive and understand game tasks through output devices and use specific input devices to complete game tasks. The interaction perspective focuses more on this changeable process of game-play. Game difficulty appears in an interaction process and thus it is dynamic. This notion of "dynamic process" contains at least three levels of connotation: 1) Difficulty is different for a single player at different moments during one interaction, 2) Difficulty is different for each of a single player's multiple plays, and 3) The difficulty of different players doing the same game task is also different. In Chapter 3, we further built a model to combine OGD and SGD in the interaction process and redefined these two difficulties in more detail.

However, questions emerged as the game difficulty was separated into subjective and objective parts and were regarded as dynamic. Based on the quantifying way proposed by Aponte et al. (2011b), Constant et al. (2017) found SGD might not match OGD because of individual bias of the player in evaluating SGD. Previous findings supported this view but researchers only roughly attributed it to various players' complex difficulty evaluations and explanation patterns (Aponte et al., 2011a; Hunicke, 2005). There remains a lack of research aimed at deeply clarifying the relationship between OGD and SGD and their impacts on players. Therefore, we explored this relationship by experiment in Chapter 4.

2.2 Multidimensional Structure of SGD

To form the perception of SGD, players may evaluate the game task, the interaction process, and their playing states. Therefore, we reviewed the literature on these three interaction elements and summarized the six dimensions of SGD perception.

Regarding the game task element, SGD is related to the complexity of the game and the possibility of game completion. Liu and Li (2012) defined subjective task complexity as the task performer's feeling about the complex degree of the task. Similarly, players may have a perception about how complex the game is, i.e., game complexity (Dziedzic and Włodarczyk, 2018). Constant et al. (2017) measured the SGD by the player's estimate of their likelihood of failure. We suggested that this perception can be explained as the player's expectation of whether they can complete the game task;this may be called game completion difficulty.

Regarding the interaction element, the player perceives the SGD by 1) experiencing the specific process of interacting with the game tasks and 2) evaluating their performance during game play. Researchers usually identify the difficulty of the process by classifying game challenge types. Vahlo and Karhulahti (2020) summarized game challenge inventory into twelve types and four core factors of challenges: physical, analytical, socioemotional, and insight. Similarly, Denisova et al. (2020) divided the challenges into four categories: emotional, performative, cognitive, and decision-making challenges. These types of challenges represent the different aspects of game task demands that can be imposed on players. The SGD for interaction is tightly related to player perceptions of these demands (Robinson, 2001; Kim, 2009; Byström and Järvelin, 1995). Moreover, players would evaluate their performance and produce a perception of their competence in performing game tasks (Ryan et al., 2006; Johnson et al., 2018). A player's sense of competence is about whether players believe they can play well in the game. This perception overlaps with the concept of self-efficacy, which refers to the task performers' opinions about their capabilities and performance on specific tasks (Bandura and Wessels, 1994; Liu and Li, 2012). In summary, player SGD for the interaction process has two aspects, namely, game-playing difficulty and player competence.

For the player element, SGD is related to their sense of effort and the negative emotions aroused by the difficulty of the playing process (Liu and Li, 2012; Juul, 2009). According to research in the fields of psychology and HCI, mental workload (i.e., the mental demand imposed on the performer by the task; Gopher and Donchin, 1986) was introduced in quantifying task difficulty (Ayaz et al., 2012; Martin, 2014; Wickens et al., 2015; Hsu et al., 2007). Perceived effort and perceived pressure are commonly used to measure mental workload through the self-report method (Hart and Staveland, 1988; Hart, 2006; Robinson, 2001; Steele, 2020). In game research, player effort and frustration were also mentioned when describing more direct feelings of difficulty (Pusey et al., 2021; Juul, 2009). It is natural to assume that the harder the game, the more effort is needed and the more negative feelings players experience (Lomas et al., 2017). Therefore, there are two SGD aspects to a players element, namely player effort and player pressure.

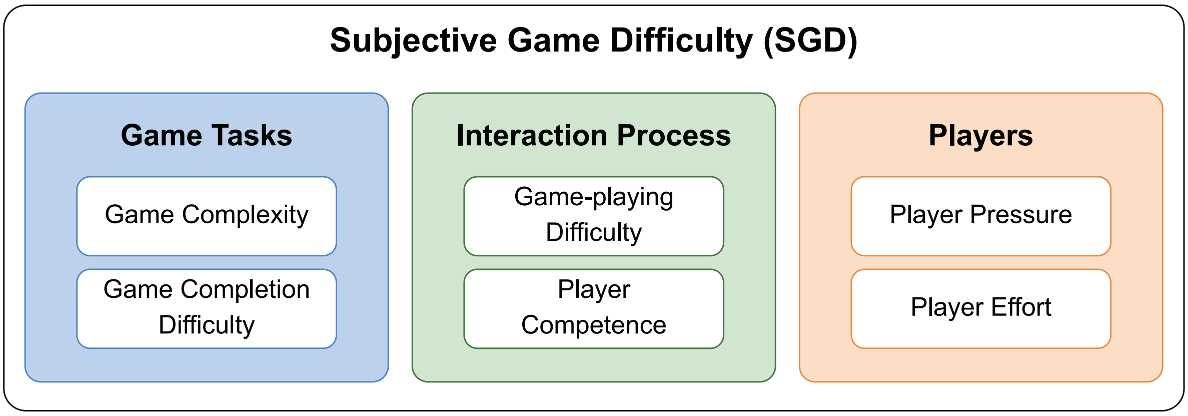

Fig. 2.1 The structure of subjective game difficulty (SGD). SGD consists of perceptions on three parts: game tasks, interaction process, and players, and six dimensions: game complexity, game completion difficulty, game-playing difficulty, player competence, player pressure, and player effort.

Combining these views, we structured SGD as three parts and six dimensions, see Fig. 2.1. The *game tasks* part has two dimensions, namely, game complexity and game completion difficulty. These two dimensions represent the player's perceptions of the game task complexity and the difficulty of completing this game, respectively. The *interaction process* part has two dimensions, namely, game-playing difficulty and player competence. They represent the player's two aspects, namely, player perceptions of the playing difficulty and their personal performance. The *players* part has two dimensions, namely, effort and pressure. The player effort dimension represents the player's perception of how much effort they invest in the game. Besides the feeling of stress, the pressure dimension further includes the player's perception of other negative feelings related to the pressure, such as nervousness and frustration. These dimensions were used in the SGD measuring in Chapter 4 and the development of the SGD scale in Chapter 6.

2.3 Measurement of OGD and SGD

2.3.1 Measuring Methods of OGD

Measuring dynamic game difficulties is challenging. Measuring OGD still lacks effective methods. In comparison, several SGD measuring methods and techniques have been developed to measure both real-time SGD and overall SGD (Cox et al., 2012; Denisova et al., 2020; Martin, 2014; Charles and Nixon, 2019). Due to the dynamic nature of game difficulty, measuring OGD by the degree of game complexity is not effective (Wehbe et al., 2017; Hsu et al., 2007; Klimmt et al., 2009). More precise tools are required to measure OGD.

Game researchers have tried to detail basic game challenges to make the game difficulty concept more operable. For example, Adams (2014) split game challenges into more than thirty types of atomic challenges, i.e., the basic types of game challenges. On this basis, Vahlo and Karhulahti (2020) summarized twelve types of challenge and four core challenging factors: physical, analytical, socioemotional, and insight. Denisova et al. (2020) divided challenges into four categories: emotional, performance, cognitive, and decision-making challenges. However, although these challenge lists indicated OGD sources, they helped little in measuring OGD.

As a reference, in HCI, task difficulty was objectively measured by user performance, e.g., performance time and error rate (Fitts, 1954; Zhai et al., 2004; Wobbrock et al., 2008). Therefore, some researchers tried to quantify OGD based on the failure to successfully complete the game task. Usually, OGD is quantified as the player's failure rate in performing game tasks directly (Hocine et al., 2015; Anguera et al., 2013; Kitakoshi et al., 2020). It is a commonly used method to measure the overall OGD of the specific game task, usually by calculating the average failure rate of one player's multiple plays or one play of multiple players. However, this method cannot reflect the changing or dynamic nature of OGD in an attempt (i.e., real-time OGD) and it is imprecise over a few attempts. In more detail, "Overall OGD" refers to the general OGD of the whole play of one player or a group of players. "Real-time OGD" is more about how difficulty changes during one player's play, i.e., it relates to dynamic changes in difficulty during an interaction. As the interaction perspective was adopted, measuring the real-time OGD has become necessary.

To improve OGD measurement, Aponte et al. (2011b) proposed that OGD can be represented as a probability of player failure function at a specific time. This method is a development using failure rate to measure the real-time OGD and it has been applied in recent studies (Constant et al., 2017; Sarkar and Cooper, 2019; Roohi et al., 2020). However, it still lacks adequate operability, not providing a clear OGD measuring method: it remains unclear how to determine the computational ways of player's failure probability in most types of games. Referring to this method proposed by Aponte et al. (2011b), some studies also utilized the incompletion rate to quantify real-time OGD (Khoshkangini et al., 2021; Kristensen and Burelli, 2022). In comparison, the incompletion rate can be used to measure the real-time OGD and overall OGD. However, the usability of this method depends on whether a game's task completion state can be defined. For example, it seems hard to determine the task completion state of a player in a GO game.

Other measuring methods have also been developed. Gallego-Durán et al. (2018) suggested defining OGD as the learning progress over time. Pusey et al. (2021) suggested that the "time taken to solve the puzzle" and the "number of incorrect/failed attempts" can be used in analyzing the OGD. However, these methods were still limited in clearly and precisely measuring the two kinds of OGD.

In summary, although OGD is identified in the player-game interaction, it remains unclear how to quantify real-time OGD and overall OGD for measurement. We believe that a deeper understanding of game tasks and game interactions will be helpful. Therefore, in Chapter 5, we propose a new measuring method of OGD and validate our method by the experiment.

2.3.2 Measuring Methods of SGD

There are currently three main methods for measuring SGD: simple self-report through questions, structured self-report through scales, and physiological measurement. In game research, a common method is to ask players to rate their perception of the difficulty of the game via a few questions (Wheat et al., 2016; Wehbe et al., 2017; Demediuk et al., 2019). However, this simple self-report method is ineffective in providing details but may help understand the player's overall perceptions of SGD. The other methods include self-report scales, e.g., asking players to report the challenges in the game (Denisova and Cairns, 2015), and measuring the players' physiological states through targeted indicators (Charles and Nixon, 2019). However, we have found that SGD has a multi-dimensional structure in Section 2.2, and different players may have different evaluation patterns (Li et al., 2014). Current methods only focused on some aspects of the six dimensions. Therefore, we reviewed the measuring approaches for each proposed dimension.

Regarding game complexity, we didn't find any instruments to measure game complexity directly. In other fields, task perception was used to describe the task performers' subjective interpretations of the task's attributes and demands (Luyten et al., 2001). Therefore, this subjective interpretation of the game task can be used to measure the game's complexity. For game completion difficulty, Constant et al. (2017) suggested it could be measured by the player's estimate of their chance of failure. However, how to estimate players' perception of the game completion difficulty in more detail remains unclear.

For game-playing difficulty, the simple self-report method is commonly used (Wheat et al., 2016; Wehbe et al., 2017; Demediuk et al., 2019). Recently, Denisova et al. (2020) developed and validated a game challenge scale, *Challenge Originating from Recent Gameplay Interaction Scale* (CORGIS), to measure the players' perceptions of the difficulty of the playing process. As mentioned in Section 2.3.1, they classified the challenges into four types, namely, emotion, performance, cognition, and decision-making. Therefore, measurement of game-playing difficulty may refer to the difficulty caused by these aspects. For player competence, it is necessary to first refer to the research of task performer competence. In the fields of psychology, design, and education, this perception can be measured by asking performers to evaluate their performance (Hart and Staveland, 1988; Hart, 2006; Young et al., 2008; Gray, 2014; van Dinther et al., 2014). In game research, player competence is usually measured by the competence subscales of the Player Experience of Need Satisfaction (PENS) scale, or the Intrinsic Motivation Inventory (IMI) scale (Rigby and Ryan, 2007; Ryan et al., 2006).

For player effort and player pressure, mature instruments in other fields can be referenced. In HCI and psychology, workload measurements include using subjective reports and physiological techniques to measure (Bevana et al., 1991; Chen et al., 2011; Veltman and Gaillard, 1998). The subjective report instruments (e.g., NASA-Task Load Index; Hart and Staveland, 1988), require task performers to report how much effort they invested and how frustrated they were during task performing. Regarding physiological techniques in effort and pressure measurement, indicators like heart rate, respiration, and blink rate are used to measure the objective state of investment/stress of performers (Charles and Nixon, 2019). In game research, the IMI scale provides two subscales, effort and pressure, for measuring these two dimensions. In addition, in IMI, the pressure from game difficulty is measured not only by the feeling of pressure but also by the related emotions of nervousness and anxiety. However, little study has applied this scale to measure the SGD except for the motivation while playing.

In summary, current measuring methods are not comprehensive and inclusive; rather, they focus on the different aspects of SGD. Based on these instruments, we developed a new scale that measures the six dimensions of SGD in Chapter 6.

2.4 The Impacts of Game Difficulty on Players

Game difficulty has been reported to impact player motivation, experience, engagement, and self-efficacy (Allart et al., 2017; Lomas et al., 2017; Power et al., 2020). However, as different concepts of game difficulty were used in research, the respective impacts of OGD and SGD on players were hard to distinguish.

Motivation is used to explain where players' enjoyment comes from. Researchers indicated that difficulty is an important source of game fun (Juul, 2009) by fulfilling players' challenge needs to trigger game-play behavior (Malone, 1981; Yee, 2006). Malone (1981) proposed a famous early view that challenge is one of three game motivations. Self-determination theory (Deci and Ryan, 1985, 2000) also discusses challenge needs and proposes that players want to feel competent and capable when playing games (Ryan et al., 2006). Challenge needs are then discussed and integrated into achievement needs by Yee (2006) and Bostan and Öğüt (2009); however, Bostan and Öğüt (2009) further suggested that sensual needs contain relaxation and the need to amuse oneself, which explains the enjoyment of casual games. In addition, different players have different needs will be motivated in different ways (Tondello and Nacke, 2019).

Player Experience is the player's individual and personal experience born from the whole game process (Wiemeyer et al., 2016). Player experience theories focus on 1) interpreting players' experience contents and how to measure it subjectively and 2) identifying good player experience and how to create it. For example, Flow theory and immersion experience describe similar good experiences in which players forget time and themselves, becoming immersed in games (Chen, 2007; Jennett et al., 2008). Related scales are also used in measuring such experiences (Fang et al., 2013; IJsselsteijn et al., 2013). Flow theory indicated a good design in the perceived challenge (i.e., SGD) is promising to create a fancy player experience. However, most researchers regarded the challenge in Flow theory as the OGD and believe there is an optimal OGD for the great player experience (Bostan and Öğüt, 2009; Chen, 2007). Other theories such as playability (González Sánchez et al., 2009; Paavilainen, 2017, 2020), Game Experience Model (Suovuo et al., 2020), etc., have also been proposed for describing the player experience and supporting better experience design. More studies focused on the effects of OGD on player experience but indicated that the effects seem to be more complex. Juul (2009) proposed that the possibility of failure is primary to player enjoyment of games and that it makes success meaningful. However, Klimmt et al. (2009) reported that the easier the game (OGD), the more enjoyment players reported. Lomas et al. (2017) also reported similar results that the easiest games (OGD) were the most motivating. Therefore, it still remains unclear the impacts of OGD and SGD on the player experiences.

Engagement theories focus on player engagement in games and there are two definitions. One is the short-term engagement. It somewhat overlaps with player experience theory but pays more attention to how to objectively describe the player's state (e.g., cognitive, behavioral, and emotional state) of partial immersion in the game playing (Jennett et al., 2008; O'Brien and Toms, 2008; Sharek and Wiebe, 2014; Przybylski et al., 2010). Another one is the long-term engagement, which refers to a player's willingness to participate in future play (Huang et al., 2024). Current research quantifying the long-term engagement through the rate of replay and churn (Burke et al., 2010; Khajah et al., 2016; Xue et al., 2017; Roohi et al., 2020; Alan et al., 2022). To distinguish player engagement from player experience, we adopted the long-term engagement in our following studies. According to research, players' preferences (Karpinskyj et al., 2014) and needs (Malone, 1981; Ryan et al., 2006; Bostan, 2009) regarding game difficulty affect their motivation to play and engagement. For example, Xue et al. (2017) designed a new DDA mechanism for maximized engagement; they used the replay and churn rate to represent engagement. We argue that a player's motivation to play again is more practical to predicting players' future game possibilities and this has been used in recent game research (Alan et al., 2022; Roohi et al., 2020). However, there is a lack of clear evidence on whether this willingness to future playing is produced based on OGD or SGD.

Self-efficacy is the player's confidence on the specific game task, which is also believed to interact with player OGD (Vancouver et al., 2002) and SGD (Constant and Levieux, 2019; Power et al., 2020; Nuutila et al., 2021). According to research, self-efficacy is positively related to performance (negatively related to OGD; Stajkovic and Luthans, 1998). Findings also suggest that self-efficacy is not the perceived difficulty (SGD), even though they are similar (Rodgers et al., 2008). Power et al. (2020) provided a clearer illustration that OGD is negatively associated with self-efficacy and mediated by mastery experience. On the other hand, Nuutila et al. (2021) explained that increased perceived difficulty may lower player self-efficacy, which implies the impact of SGD. In summary, OGD and SGD seem negatively related to player self-efficacy, but further study is still necessary to determine the effects of these two difficulties separately.

In short, due to the evolving game difficulty concept, it is necessary to study whether OGD and SGD impact player motivation, experience, engagement, and self-efficacy, respectively. In Chapter 4, we explored the impacts of OGD and SGD on these factors through an empirical experiment.

2.5 Dynamic Diffusely Adjustment Mechanism

According to Zohaib (2018), DDA is *"a method of automatically modifying a game's features, behaviors, and scenarios in real-time, depending on the player's skill, so that the player, when the game is very simple, does not feel bored or frustrated, when it is very difficult"*. This definition regards the challenge-skill balance as the primary goal of DDA, which is the consensus in current DDA research (Hunicke, 2005; He et al., 2010; Baldwin et al., 2013; Alexander et al., 2013; Karpinskyj et al., 2014; Denisova and Cairns, 2015; Silva et al., 2015; Lach, 2017; Pfau et al., 2020). Corresponding to the skill and challenge, DDA consists of two basic components: (1) a player evaluation mechanism for measuring player performance and (2) a difficulty adjustment mechanism to change the level of game difficulty (Adams, 2014; Yin et al., 2015; Demediuk et al., 2017).

2.5.1 Difficulty Adjustment Mechanism

Difficulty adjustment mechanism is about changing the real-time challenges dynamically. Due to the natural differences between different types of games, the difficulty adjustment mechanism is diverse in various games. There are three main kinds of adjustment techniques according to the adjustment methods: adaptive game AI technique, adaptive content generation technique, and adaptive content adjustment technique.

Adaptive game AI technique aims that controls the interaction and competition between Non-player Characters (NPCs) and players. This technique mainly changes the difficulty by adjusting the strength of game AI agents (He et al., 2010). Artificial intelligence algorithms such as Self-Organizing Systems (Ebrahimi and Akbarzadeh-T, 2014), Artificial Neural Networks algorithm (Yin et al., 2015), and Monte Carlo Tree Search (MTCS; Demediuk et al., 2017; Pratama and Krisnadhi, 2018; Moon et al., 2022) are used in improving this technique. Adaptive content generation generally refers to the automatic generation of different game levels (Bakkes et al., 2014), and this kind of technique is usually used in platform games (Jennings-Teats et al., 2010). Recently, machine learning algorithms such as Bayesian-based Intelligent Trial-and-Error Algorithm (IT&E) are also applied in this technique to generate game levels (Risi and Togelius, 2020; González-Duque et al., 2020). Adaptive content adjustment techniques usually adjust the game's contents more directly. For example, the Hamlet System designed by Hunicke (2005) tried to limit the player's health value to a specific range by adjusting the player's weapon damage and the donations of health packs. Some serious games also adopt this technique in DDA design, for instance, adjusting the response time (Anguera et al., 2013; Sampayo-Vargas et al., 2013) or the contents need to be remembered (Kitakoshi et al., 2020).

However, a good DDA design with precise difficulty adjustment should depend on effective player evaluation. Therefore, more research focuses on sdtudying the DDA's player evaluation mechanism.

2.5.2 Player Evaluation Mechanism

According to different evaluations of players, the player evaluation mechanism of DDA can be divided into OGD evaluation technique, SGD physiological evaluation technique (including physiological techniques and self-report method).

The OGD evaluation technique is widely used for its association with challenge-skill balance. This technique assesses the player's performance by analyzing players' in-game behavior or players' success rate on tasks (or win rate of competitions). For DDA with adaptive game AI, researchers are usually concerned with whether the designed AI agent can achieve a 50% win rate against a real player (He et al., 2010; Demediuk et al., 2017). While for DDA with adaptive content generation, evaluation is usually achieved through modeling the players' real-time behaviors in the game (Jennings-Teats et al., 2010). And for DDA with adaptive content adjustment techniques, an early study assessed players based on their behavior, such as data on the player's location, number of deaths, and number of system interventions (Hunicke, 2005). Other studies used task success rates, such as the success rates on signal response tasks (Anguera et al., 2013) and correct response rates (Sampayo-Vargas et al., 2013; Kitakoshi et al., 2020). However, the findings of DDA that employing this technique implied its contribution to good game experience seem weak (Hunicke, 2005; Orvis et al., 2008; Smeddinck et al., 2016; Salehzadeh Niksirat et al., 2017).

Recently, researchers have paid attention to evaluating the SGD of players. Flow theory suggests that players will have different feelings when difficulty is changed. Based on this point, some researchers have developed physiological techniques for assessing players' emotions, such as Galvanic Skin Response (GSR), blood pressure, and EEG (Chanel et al., 2008, 2011). Facial temperature recognition (Moniaga et al., 2018) and facial expression recognition (Akbar et al., 2019) also used for player emotion evaluation. The physiological evaluation technique is adaptive to all adjustment techniques for its more related to the players, but not the games. Studies applied these methods in DDA design and have shown good results in improving player experience (Liu et al., 2009; Stein et al., 2018). Only a small body of research uses the self-report method, typically involving players judging the game difficulty and reporting during play (Pedersen et al., 2010; Alexander et al., 2013; Frommel et al., 2018). However, studies have not found this method very effective in improving the player experience. Recently, more studies have begun combining more than two techniques to improve the SGD evaluation (Ang and Mitchell, 2019; Ozkul et al., 2019; Moon et al., 2022).

In summary, research indicates that DDA's effects on creating a great player experience cannot always be ensured. It is necessary to clarify how to design DDA when taking into account OGD and SGD. Therefore, we redefined DDA and proposed a DDA design methodology by rethinking its theoretical fundamentals in Chapter 7. We further validated our DDA design methodology by a case study in Chapter 8.

Chapter 3

Redefine OGD and SGD Based on an Interaction Model

As mentioned in Chapter 2, a game is a system that involves tasks for players to interact. Game difficulty occurs in the interaction between players and game tasks. Therefore, this chapter introduces the three components of the game task, players, and interaction to clarify how game difficulty occurs in specific processes and proposes an interaction model of game difficulty.

3.1 The Components of Player-game Interaction

3.1.1 Tasks, Game Tasks, and Task Complexity

Even though tasks are so commonly used in psychology and education areas, most researchers tend to use this concept by describing it without defining it (e.g., Simon task, multi-tasking, speaking task, etc.; De Houwer, 2003; Fulcher and Reiter, 2003; MacPherson, 2018). According to Winne (1985), tasks present a collection of initial conditions and set a goal. Problem is a task-related concept usually used in psychology and education. Similar with task-performing, problem-solving is considered to involve an interaction of a person's experience and the demands of the task (Martinez, 1998). Dunbar (2017) proposed that a problem includes the components of an initial state, the goal state, actions or operations, and task environment (or task rules). Dunbar (2017) also mentioned representation is one of the key elements for solving it. The reason is that solvers should construct their understanding of problem features based on the problem statement and features of the task environment. These descriptions are helpful in understanding tasks and task-performing.

As we mentioned in Chapter 2, HCI field uses the definition of tasks as "activities that people should conduct to move their work and life on" (Liu and Li, 2012; Li and Belkin, 2008). This definition provides a general description of tasks but lacks enough details about how to analyze and construct tasks. It should be noted that in HCI, understanding tasks is for designing and evaluating tasks in the human-computer interaction process. For example, usability is the ease of use and acceptability of a system or product by evaluating the task-performing process (Bevana et al., 1991). Therefore, we proposed the game tasks refer to **tasks that players must perform in order to complete, advance or reach a specific goal in the game**, and adopted the game complexity to describe task attributes in more detail.

Campbell (1988) believed objective task complexity was contributed by four objective task characteristics: paths, end-states, interdependence, and uncertainty. This view is more of an analysis perspective about why a task is complex. Liu and Li (2012) described task complexity as an objective aggregation of task characteristics of goal, input, process, time, and presentation. Their opinion is more helpful in constructing tasks and explainable list of characteristics proposed by Campbell (1988). It also partially overlaps with problem components (goal state: goal, actions: input, rules: process and time, and representation: presentation).

To summarize, game tasks can be defined as task that should be conducted in games and can be designed in specific forms with four basic elements: goals, rules, states, and presentation. Goals are the desired future state realized by performing tasks; rules specify how to accomplish tasks, task failure conditions and task performing environment; state includes the initial state, ongoing state, and end state of tasks (success or failure), and the ongoing state refers to the problem space that contains different task paths; presentation contains all information types, contents, and presentation rules of tasks and tasks performing, and therefore specifying task description (or problem statement), how information output during task performing and uncertainty of task information (Diaper and Stanton, 2003; Dunbar, 2017; Liu and Li, 2012).

3.1.2 Player Factors and Game Difficulty

As pointed out by researchers, game difficulty is also highly related to the player's characteristics (Dziedzic and Włodarczyk, 2018; Denisova et al., 2020; Cox et al., 2012). Game challenges are usually regarded as an important source of game fun (Juul, 2009) and thus become one of the player motivations to trigger game-play behavior (Malone, 1981; Yee, 2006). More specifically, player motivations are important for them to evaluate and overcome difficulties in the game. It has been shown that players' preferences

(Karpinskyj et al., 2014) and needs (Ryan et al., 2006; Bostan, 2009) on game difficulty affect their playing and future engagement. Therefore, some researchers attempted to keep players motivated by providing optimal difficulty dynamically by DDA mechanism (Hunicke, 2005; Zohaib, 2018). Studies also used player preferences modeling to provide personalized game difficulty (Yu and Trawick, 2011; Bakkes et al., 2014), but more DDA researchers focus on players skills (He et al., 2010; Yun et al., 2010; Jennings-Teats et al., 2010; Alexander et al., 2013).

Player skills are essential in understanding game difficulty (Adams, 2014; Denisova et al., 2020) and can be divided into cognitive skills (Bostan and Öğüt, 2009; Denisova et al., 2020) and game skills (or game expertise) (Cox et al., 2012; Lee and Heeter, 2017; Deterding, 2015). Cognitive skills are basic abilities of humans and are related to how we precept, memorize, understand, and process information during task-performing (Sternberg and Kaufman, 2011; Carroll, 1993). As mentioned in Chapter 2, Game skills are more specialized skills, and game difficulty is regarded as the interaction between player skills and game task demand (Adams, 2014). Player skills are commonly quantified by player performance in some game studies (Jennings-Teats et al., 2010; Zook and Riedl, 2012; Demediuk et al., 2017). According to Sweller (Sweller, 1994), learning is to store knowledge and construct automated patterns (they call them schemas) of knowledge in long-term memory. Therefore, acquired knowledge of games (i.e., game skills) can reduce the need for processing resources of game tasks and lower the game difficulty (Martin, 2014).

Game experience refers to players' previous experience of playing games and is also used as one of the indicators of game skills in some studies (e.g., Yun et al., 2010). In addition, researchers suggested game experience affected subjective game difficulty (Pato and Delgado-Mata, 2013; Cechanowicz et al., 2014), which can be explained by self-efficacy theory that different game experience probably provides different explainable background knowledge (Bandura et al., 1999). Self-efficacy of players was also believed to interact with player performance (Vancouver et al., 2002) and subjective game difficulty (Constant and Levieux, 2019; Power et al., 2020; Nuutila et al., 2021), and thus affected player experience (Lin et al., 2018). Even though many game researchers focus on player experience in research of game design and game difficulty evaluation (Jennett et al., 2008; Altimira et al., 2014; Khajah et al., 2016; Smeddinck et al., 2016), it should be noted that player experience comes from the game interaction process and is not the characteristics of the player itself.

To summarize, each player has different self-efficacy, motivation, skills, and game experience. These factors affect objective and subjective game difficulties by influencing the player-game interaction process.

3.1.3 Understanding Player-game Interaction

The interaction perspective can be adopted in video games. It is based on player-game interaction (Caroux et al., 2015), which belongs to a type of human-computer interaction. The specific interaction process in HCI mainly consists of how human input and computer output. For computer output, the mainstream output types are visual and auditory by digital devices to provide enough information on interaction tasks (Jacko, 2012). In addition, haptic feedback (e.g., vibration; Keates et al., 2000) is also used in some interactions to support output (MacLean, 2000). For human input, there are kinds of classical devices (e.g., mouse and keyboard; Jacko, 2012) and new techniques (e.g., brain control and gaze; Tan and Nijholt, 2010; Rozado et al., 2015)) to make input and usually can be divided into input types of pointing, stroke, gesture, motion and others roughly (Jacko, 2012; Schmidt et al., 2018). Caroux et al. (2015) introduced how interaction happens between players and video games in their review paper and discussed input and output research of game. As they mentioned, visual and auditory output were commonly used but game researchers focused more on new input techniques rather than traditional devices in player-game interaction. However, it should be noted that the mouse & keyboard, touch screen and controller are still more popular devices and are often compared in game research (Lee et al., 2015; Oshita and Ishikawa, 2012; Brown et al., 2015).

Game playing process thus can be regarded as the player-game interaction process, in which players receive and understand the game tasks through output devices and use specific input devices to complete tasks. The interaction perspective focuses more on this changeable process of players' game-playing. A well-known issue is that, ideally, player skills improve as the game progresses (Huniche and Chapman, 2005; Jennings-Teats et al., 2010), which constructs the basis of Flow-based DDA. More than skills, however, player motivation, experience (whether game experience or player experience), and many other factors also change as the game interaction progresses. Game difficulty appears in such a process and thus is dynamic but not just a result.

3.2 Difficulty Interaction Model

Based on our elaborations of these three components, we built a model to present how game difficulty occurs in the player-game interaction. In this interaction process, game tasks provide visual and auditory outputs to players, while players need input to meet the task demand. See Fig. 3.1.

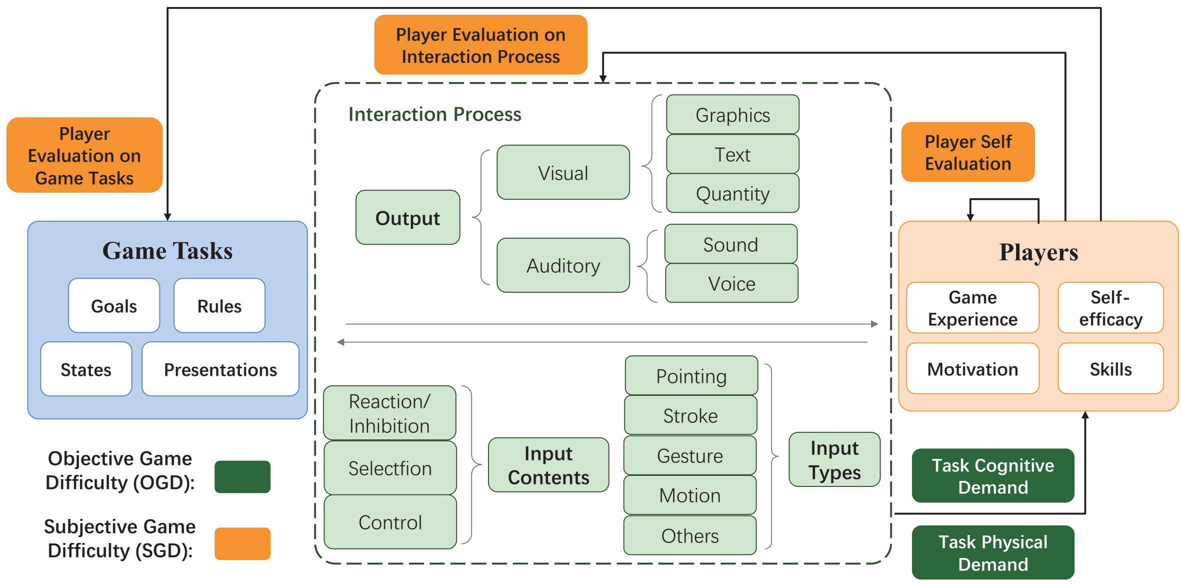

Fig. 3.1 The game difficulty model in the player-game tasks interaction. In this model, objective game difficulty is the cognitive and physical demands imposed on the player by game tasks; while SGD is the player's evaluation of game difficulty based on the structured perceptions of the game task, interaction process, and their game states.

More specifically, OGD and SGD occur in this dynamic process. Tasks, based on their structure and complexity, impose cognitive and physical demands on the players and provide dynamic feedback by checking the players' input. In addition, the structure of tasks is usually kept static unless specially designed for adapting. In contrast, players' characteristics are dynamically changing in this process. In more detail, players' skills and game experience will develop, while motivation and self-efficacy will correspondingly change due to needs satisfaction and mastery experience. Therefore, OGD refers to how well the player's skill meets the demands and is thus dynamic during the interaction. In comparison, SGD is evaluated by players based on three aspects: the game task, the interaction process, and the player's own. Therefore, SGD is also dynamic but more complex than OGD in structure. In addition, players can evaluate SGD at different stages: 1) before specific interaction but have constructed the task perception, 2) during the interaction process, and 3) after the interaction process. However, the OGD can only be assessed during and after the interaction process.

The tasks given to players by video games become the prerequisite for the game difficulty and game difficulty appears in the interaction process. Therefore, we further defined OGD as **"OGD is the dynamic meeting result of the player's skill to the game task demand during gameplay"**; while SGD is defined as **"the player's subjective evaluation of game difficulty based on their structured perceptions of the game task, game-playing process, and their game states."**

Chapter 4

Relationship between OGD and SGD

In this chapter, we explored (1) whether there is a mismatch between OGD and SGD and (2) how OGD and SGD affect player experience, engagement, and self-efficacy. We first built a research framework and proposed seven hypotheses and then we conducted an experiment to test these hypotheses.

4.1 Research Framework and Hypotheses

4.1.1 Research Framework and Measuring Methods

Even though previous studies implied a complicated relationship between OGD and SGD (Hunicke, 2005; Aponte et al., 2011a), it is still necessary, by more comprehensive verification, to determine whether there is a mismatched relationship. Therefore, we established a research framework to clarify our research process, see Fig. 4.1. When game difficulty is separated into OGD and SGD, four situations between OGD and SGD are expected to emerge: 1) low OGD and low SGD, 2) high OGD and high SGD, 3) high OGD but low SGD, and 4) low OGD but high SGD. If OGD matches SGD, there should be the first two situations but no third or fourth situations. Furthermore, regardless of whether OGD and SGD match each other, we wanted to study how they interact to impact player experience, engagement, and self-efficacy.

Before conducting the experiment, a critical issue is to determine the measuring methods of OGD and SGD. As mentioned in Chapter 2, game difficulty happens in the interaction process and keeps changing over time. Changes occur because of the development of players' skills (Johanson et al., 2019), the advancement through the game levels, and changes in player experience. However, it is very challenging to measure the SGD and OGD and assess their match relationship during the interaction process. A possible

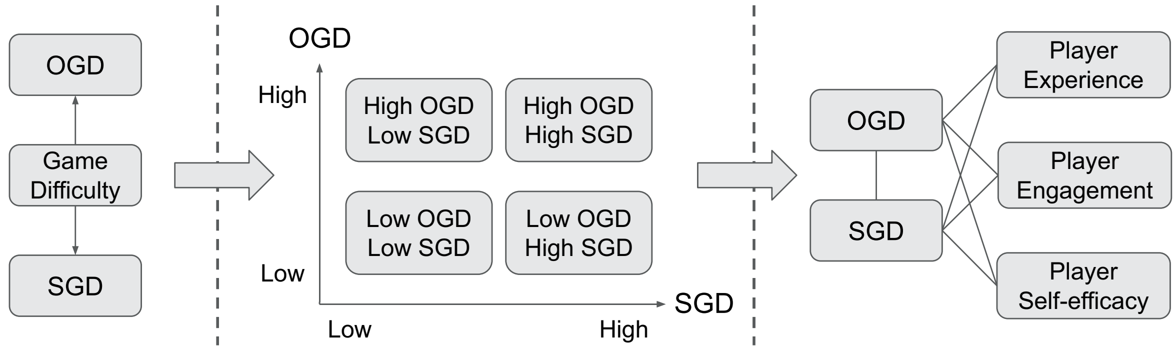

Fig. 4.1 Our research framework. We first separate game difficulty into objective game difficulty (OGD) and subjective game difficulty (SGD). Subsequently, we assume there are four situations combining OGD (Low or High) and SGD (Low or High) that should be validated. Finally, we investigate the respective impacts of OGD and SGD on player experience, engagement, and self-efficacy.

option is to measure OGD and SGD after the game play. Regarding SGD, Chapter 2 has reviewed the six-dimensional contents, and the self-report method for measurement is suitable. For OGD, based on the evaluation and prediction of player performance, there are two choices for measuring: one is the function of failure probability, and the other is the failure rate. However, we could not find a universal failure probability function for all games; the function seems to depend on the game design, various preset parameters, and player performance (Aponte et al., 2011b; Jennings-Teats et al., 2010; Constant and Levieux, 2019). Furthermore, this method is not suitable for measuring OGD after game play.

Another possible post-game form of this method is the degree of game completion, the higher degree of completion, the closer to success and the lower the OGD. However, compared to the degree of completion, the failure rate still fulfills the requirements of this study more satisfactorily. The reason is that it is difficult to control the degree of game completion for each player when conducting the experiment, and thus it is hard to create expected conditions for validation (e.g., a high or low OGD condition). In any case, the failure rate is still more convincing and universal when we need to select only one measuring method (Hocine et al., 2015; Anguera et al., 2013; Kitakoshi et al., 2020).

After determining the measuring methods for OGD and SGD, it is still necessary to illustrate how we can validate the existence of the four situations. We would lose scientific rigor if we let players play a game freely and search whether there are assumed situations existing by checking their OGDs and SGDs. Therefore, a more rigorous approach is to control one difficulty (OGD or SGD) and test whether the other difficulty is as expected. However, as the subjective feelings of each player (SGD) are individual and fluid, they cannot be purposefully designed. More importantly, if there is a mismatch between OGD and SGD, it becomes almost impossible to control the SGD precisely. By contrast, OGD is about player performance and is easier to control. Therefore, creating all possible OGD conditions and validating the corresponding SGD in each condition is more operable. It should be noted that they are only ideal situations by assumption because OGD and SGD are usually more complicated than quantified as the high or low degrees. However, these situations still well reflect typical possibilities in reality and are easy to conduct in the validating experiment.

To create expected OGD conditions, it is necessary to first define the scope of a low OGD and a high OGD. Because our OGD measuring method is defined by the failure rate, the scope should be defined by the failure rate value. However, there is no standard of a fixed value scope for low or high OGD settings, and relying on experience to set the value subjectively is not convincing. For example, if the high OGD scope is set to range from 75% to 100%, can we consider a player with a 75% failure rate (succeeding once in four attempts) and a player with a 100% failure rate (no success) to have experienced the same OGD condition? Skill development is another problem when we employ the failure rate. Multiple repeats will develop players' skills in the specific game. However, the failure rate can hardly represent the skill development process but the last result of the playing process. For example, if a player develops his/her skills and finally succeeds through multiple attempts, the failure rate will be high at that time (high OGD); however, he/she is already skilled and more likely to succeed in future attempts. In this case, the relationship between OGD and SGD would hardly be explained.

Therefore, although skill development is more in line with the realistic game process, limited by the OGD measuring method and experimental needs, we decided to control skill development in this study, i.e., we minimized the number of player attempts in the game. Considering our required experimental conditions, we decided to employ the failure rate as a dichotomous variable in this research by letting players try only once. In this way, we created high OGD (player fails, failure rate = 1) and low OGD (player succeeds, failure rate = 0) conditions accordingly. Although a dichotomous failure rate may not be enough to represent all OGD situations, it fulfills our requirement of objectively scoping the OGD into low and high and is adequate for the OGD and SGD's matching relationship validation.

In summary, based on our measuring methods, only if the following three situations are true, we can prove that OGD matches SGD:

(1) OGD (with low or high failure rate values) corresponds to SGD (with low or high player ratings);

(2) SGDs (player ratings) have no differences between the same OGDs (with low or high failure rate values);

(3) SGDs (player ratings) have differences between low OGD (with low failure rate value) and high OGD (with high failure rate value).

4.1.2 Research Hypotheses

To be comprehensive, we would study this relationship through two experimental settings: OGD and SGD in one trial (the data of playing one game level, see Section 4.2) for testing the first situation and OGDs and SGDs in two trials (the data of playing two game levels, see Section 4.2) for testing the second and third situations. Based on these situations, we established our hypotheses 1-2, which assume a matching relationship between OGD and SGD, see Fig. 4.2.

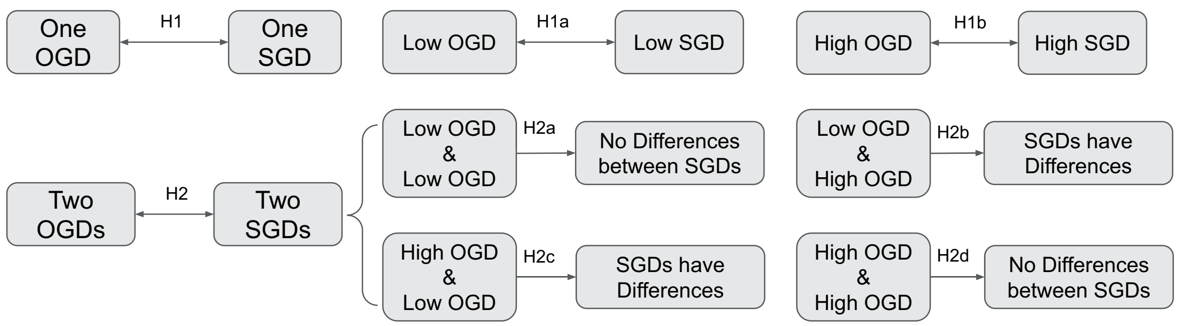

Fig. 4.2 Hypotheses 1-2 (H1 and H2) and their sub-hypotheses (H1a, H1b, H2a, H2b, H2c, and H2d).

H1. OGD and SGD match each other in one trial.

H1a. Low OGD matches to the low SGD in one trial.

H1b. High OGD matches to the high SGD in one trial.

Hypothesis 1 corresponds to situation 1 and is divided into two sub-hypotheses, 1a and 1b. We can only accept hypothesis 1 if the two sub-hypotheses are accepted.

H2. OGDs and SGDs match each other in two trials.

H2a. There are no differences between SGDs in the two low OGD conditions.

H2b. The SGDs in the two OGD conditions, low and then high, have differences.

H2c. The SGDs in the two OGD conditions, high and then low, have differences.

H2d. There are no differences between SGDs in the two high OGD conditions.

H2 corresponds to situations 2 and 3 and is divided into four sub-hypotheses, H2a to H2d. We can only accept H2 if the four sub-hypotheses are accepted. It should be clarified that we established H2b and H2c based on situation 3 for considering the order effect (Success first and then failure, or reverse). Considering the complexity of the game-playing process and individual differences among players, we believe that H1 is likely to be true, but H2 is highly likely to be rejected.

Even though there may not be a matching relationship between OGD and SGD, a correlation between them is still expected. “Correlation” and “match” here are not completely the same in connotation. Specifically, the matching relationship between OGD and SGD means that if the player has a high OGD (high failure rate), they will feel the high SGD (regard the game as difficult); if the player has a low OGD (low failure rate), they will feel the low SGD (regard the game as easy). However, the correlation between OGD and SGD describes a trend, and the positive correlation means that as the player’s failure rate (OGD) increases (or decreases), the player feels the game becomes harder (or easier). Based on this expectation, we established H3 and believed it is likely to be true.

H3. OGD and SGD are positively correlated.

To study the relationships between OGD and SGD and their impacts on players, we argue that SGD, player experience, player engagement, and player self-efficacy are all individual factors that can be affected by OGD. OGD has been reported to affect engagement positively but it negatively affects experience and self-efficacy (Bostan, 2009; Klimmt et al., 2009; Stajkovic and Luthans, 1998). Hence, we hypothesize the following.

H4. OGD and player experience are negatively correlated.

H5. OGD and player engagement are positively correlated.

H6. OGD and player self-efficacy are negatively correlated.

In addition, according to the finding that SGD affects other factors (Chen, 2007; Lomas et al., 2017; Power et al., 2020), SGD seems to be a mediated factor. Therefore, we hypothesize that SGD mediates OGD effects on these three factors as follows, see Fig. 4.3.

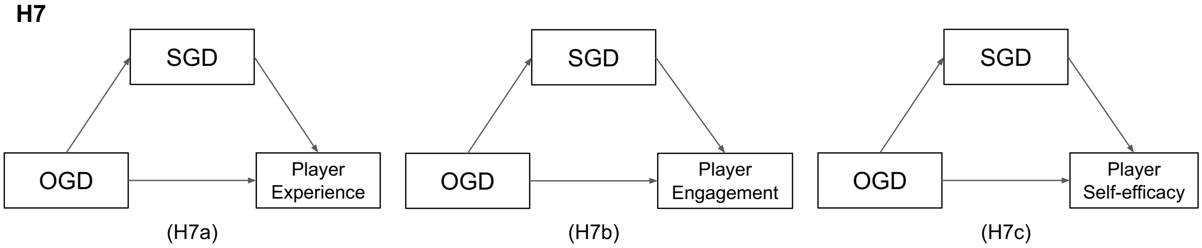

Fig. 4.3 Hypotheses 7 (H7) and its three sub-hypotheses (H7a, H7b, H7c).

H7. SGD mediates the effect of OGD on other players' individual factors.

H7a. SGD mediates the effect of OGD on player experience.

H7b. SGD mediates the effect of OGD on player engagement.

H7c. SGD mediates the effect of OGD on player self-efficacy.

H7 is divided into three sub-hypotheses: H7a concerns player experience, H7b concerns player engagement, and H7c concerns player self-efficacy. We can only accept H7 if the three sub-hypotheses are accepted, and we believe that H7 is likely to be accepted.

In summary, we have established seven hypotheses and their sub-hypotheses. Among them, H1 and H2 assume a matching relationship between OGD and SGD. H3 assumes a positive correlation between OGD and SGD. H4 to H7 assume correlations between individual player factors (including experience, engagement, and self-efficacy) and OGD, with SGD mediating this correlation.

4.2 Method

This section includes game design and experimental design parts. Our experiment aims to test our hypotheses to explore the relationship between OGD and SGD based on the research framework. An experimental game was accordingly designed; participants' failure rate in the game was partially manipulated, and their six-dimension SGD, player experience, engagement, and self-efficacy were measured.

4.2.1 Game Design

To test our hypotheses, we designed and developed a "Match 3" game, *Eat Them All*, by Unity3D Engine 2021.3. Match 3 games are casual puzzle games in which the player needs to manipulate tiles to make them disappear according to matching rules, also known as matching tile games (Juul, 2007). We selected our experimental game based on the following three considerations. (1) The experimental game should easily control the failure rate; therefore, we excluded the games that have no clear success or failure (e.g., simulation games of construction) or the games that need a lot of replays. (2) Our control of the failure rate should be reasonable and inconspicuous because any unnatural or obvious control of OGD would interfere with the players' real SGD evaluation processes. Therefore, we did not adopt games with performative challenges (requiring players to act or react quickly and correctly; e.g., action games and shooting games; Denisova et al., 2020). This is because it is hard to create unavoidable failure without it being noticed by players in these games, and it may make findings difficult to interpret if we adopt any "cheating" designs in the games (e.g., fake bullets with no damage; Zhang, 2021). (3) To control the skill development process, the game should be easy to understand and play without a long learning process to improve skills. This is because we need players with different experiences and skills to show similar game performance as expected, and quick learning with few replays is necessary. This means the game should have simple rules and gameplay, so we thus excluded complex strategy games and chess games. In addition, games designed based on standard psychological tasks or paradigms or Stroop paradigm (e.g., simple reaction time task; Hultsch et al., 2002; Logan et al., 1984) are another alternative. This type of game is easy to understand and play and suitable for performance measuring. However, they do not satisfy the first and second points because they usually contain performative challenges (e.g., reacting to the signal) that can hardly control the OGD inconspicuously, and they usually need a lot of replays to measure the player's performance. Finally, we settled on casual games.

We adopted the Match 3 games in this study after investigating the mainstream casual games. This type of game has clear and simple rules, i.e., matching tiles, and the gameplay only needs the player to choose the tiles. More importantly, the complexity of this game mainly depends on tile positions, numbers, and types, which are easy to design and control. Therefore, by modifying these complexity elements, expected OGD conditions can be created inconspicuously. In detail, the matching rule in our game is that players need to find three of the same food tiles (the matching tiles). When the game begins, many food tiles will be placed by layers on the screen. The food tiles on the top layer are bright and interactive, while those covered by other tiles are gray and not interactive. Players need to use the mouse to click on the bright food tiles, which will be moved into the vertical column on the right automatically. There are seven areas for storing food tiles in the vertical column, and any three identical food tiles in the column will be eliminated, see Fig. 4.4. The game fails when the vertical column stores seven food tiles but cannot eliminate them. The game is successful when all the food tiles are eliminated. All the art assets used in this game were modified from paid online resources*[1] to fit the requirements of game design.

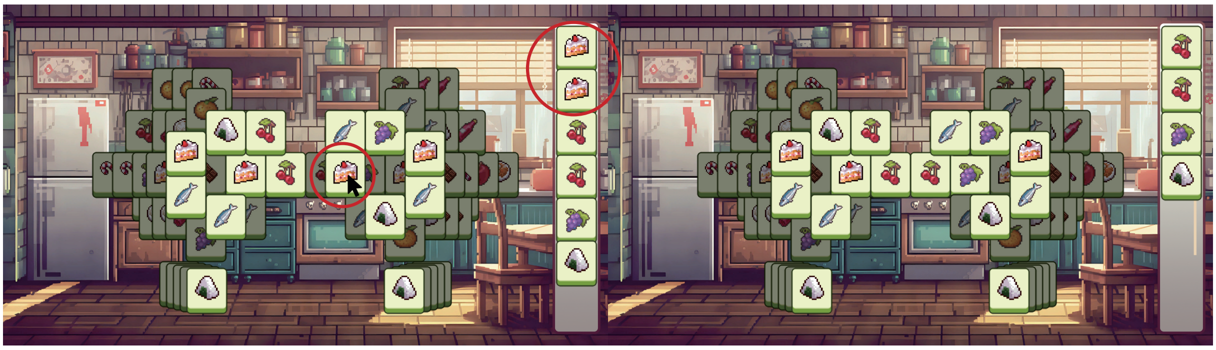

Fig. 4.4 Screenshots of the *Eat Them All* game. Players need to find three of the same food tiles and click to move them into the vertical column on the right to eliminate them.

According to our research framework and hypotheses, it is necessary to design (1) two game levels with low OGD conditions and high OGD conditions and (2) different game levels with four conditions that combine low and high OGDs. The challenge of this game design is that these OGD conditions should be created intentionally but cannot be noticed by players. In more detail, gameplay is inevitably designed to produce failure (or success) at specific game levels, but players should not suspect the game result is intentionally designed. Due to the differences in player skills, we designed a player division mechanism in our game to adapt game difficulty to their respective game skills. Players will be automatically assigned to normal mode or hard mode based on their performance while playing. Players will only play this game once and will not be informed of this normal/hard division. Players need to play four game levels in the normal mode and five in the hard mode. These game levels are designed to create the required experimental conditions, see Fig. 4.5.

We argue that if our hypotheses are valid, SGD cannot be affected by factors other than matches with the OGD. Therefore, players are designed to play different game levels but experience the same success and failure process. Specifically, all players need to play this game starting from level 0. This level is a guide level that teaches players

*[1] https://www.aigei.com/

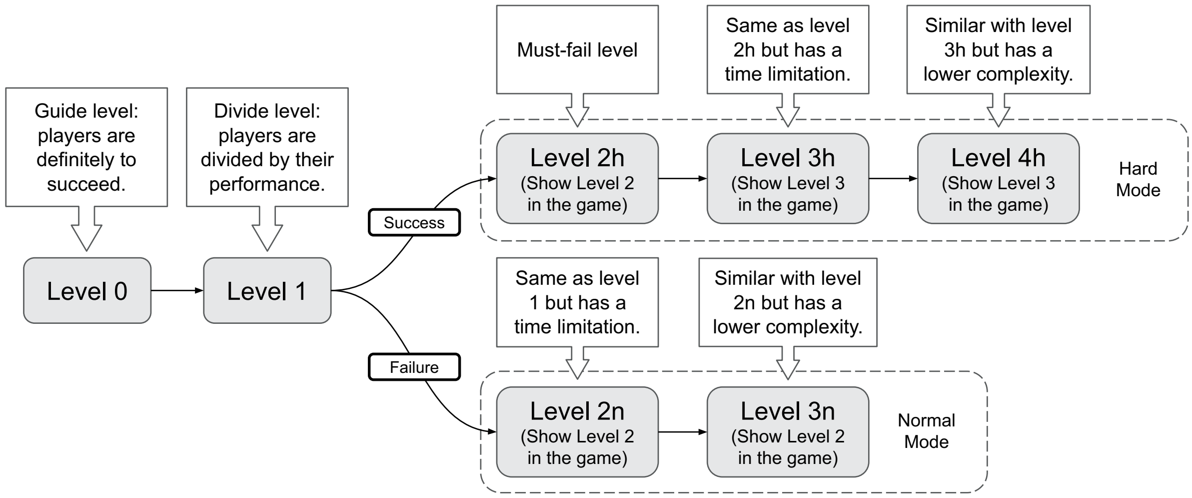

Fig. 4.5 Game level design. All players need to play this game starting from level 0 and will be divided into normal or hard mode based on their success/failure results in level 1. Players in normal mode will play level 0 (guide level), level 1 (divide level), level 2n (add a time limitation based on level 1), and level 3n (lower the complexity based on level 2n). Players in hard mode will play level 0 (guide level), level 1 (divide level), level 2h (must-fail level), level 3h (add a time limitation based on level 2h), and level 4h (lower the complexity based on level 3h).

about gameplay; players will definitely win in this level. Players then play level 1 and are divided into either normal or hard mode based on their success/failure results. Players in hard mode should play level 2h (i.e., representing level 2 in hard mode) in the next, which is a must-fail level (with very high complexity where it is almost impossible to win). The content settings of level 3h are the same as level 2h (though food tile types are different), but level 3h includes a time limitation. Level 4h looks almost the same as level 3h but it has fewer tile types to lower the complexity of the content of level 4h and make it winnable. Creating must-fail levels is unnecessary for players in normal mode because they fail in level 1. Therefore, the subsequent level is level 2n (i.e., representing level 2 in normal mode), which has the same content settings as level 1 (though food tile types are different) 2n includes a time limitation. Finally, level 3n is also modified in complexity to be more winnable. In summary, our design is such that all players, in normal or hard mode, should experience a game process of success - failure - failure - success. This process covers all the conditions we expect to create in this experiment.

We created the OGD conditions mainly by changing the layers, types, and numbers of food tiles, see Table 4.1. In addition, time limitations were added in the later levels.

Time limitation is regarded as a perceived difficulty element, but it is not related to specific gameplay (Qin et al., 2010). The time allowance setting was adequate to win the level. It is an intentional design factor that aims to help players distinguish between level 3h and level 2h (or level 3n and level 2n) because these two levels have a similar design for the tiles. We adopted a fixed-order design rather than a random-order design in the game levels. There were two reasons. On the one hand, the game was designed to acquire real and comprehensive opinions of players on SGD. Applying a random-order design is inconsistent with reality and seems strange (imagining the guiding level is arranged to play last). Therefore, the complexity of the game levels (increasing in general) and the orders of success and failure (success-failure-success) were designed to simulate a real game process. On the other hand, our design included two-trial conditions that cover all the order possibilities to weaken the order effect; e.g., players in hard mode were expected to experience: (1) two successes (level 0 and level 1), (2) succeed and then fails (level 1 and level 2h), (3) two failures (level 2h and level 3h), and (4) fail and then succeed (level 3h and level 4h). Therefore, we believe this process was well designed to control all the factors we were concerned about and promised to create our expected experiment conditions.

Table 4.1 Content of each game level. Each level has different layers, types, and numbers of food tiles to create different complexity to achieve the required OGD conditions.

Levels	Tile layers	Food tile types	Food tile numbers	Description
Level 0	2	3	18	Guide level
Level 1	6	11	99	Divide level
Level 2n	6	11	99	Time-limited level based on level 1
Level 3n	6	9	108	Lower-complex level based on level 2n
Level 2h	8	15	135	Must-fail level
Level 3h	8	15	135	Time-limited level based on level 2h
Level 4h	8	12	144	Lower-complex level based on level 3h

To test our design, we conducted a pretest on the developed game. The goals of the pretest were to test whether: (1) the game was well-developed for playing and the data could be auto-collected correctly, (2) all the expected OGD conditions could be created in both modes, and (3) participants would not notice the intentional control in the game design and would not feel any other doubts about the game. Three researchers aged 27-30 (M = 28.33, SD = 1.57) with different game skills in our team participated in this pretest. The game was proven to run well, and the results showed that one of them played the normal mode, while two played in the hard mode. All participants experienced the OGD conditions as expected and did not notice the control features, nor did they have any questions about the game, which meant our game passed the pretest and was ready for the formal experiment.

4.2.2 Experiment Design

We designed an experiment to investigate our research framework and hypotheses. This experiment has a within-subject design; the independent variable is OGD, and the dependent variables are SGD, player experience, player engagement, and player self-efficacy. To control our independent variable, as the research framework suggested, we created six conditions by varying the complexity of our game. In short, we compared the SGDs and other player factors between different game levels, and these game levels were designed to create the six OGD conditions accordingly.

To clearly illustrate how we created the six experimental conditions in this game, we list them in Table 4.2. More specifically, one level corresponds to one trial, and the combination of two levels corresponds to two trials. For the one-trial experimental situation, the low OGD condition corresponds to level 0, and high OGD corresponds to failed levels, i.e., level 1 in normal mode and level 2h in hard mode. For the two-trial experimental situation and the low & low OGD condition, the normal mode has no condition-fit levels, but in the hard mode, this condition is realized by a combination of level 0 and level 1, in which players win twice. Other conditions are similarly created according to a combination of players' success and failure situations which we shall not describe in detail further.

Participants We adopted a within-subject experimental design, therefore, an equal division of participants into the modes of normal and hard was not necessary. We regard the OGD conditions realized in the two modes as the same. However, to ensure enough data for each OGD condition, we planned to recruit forty participants. Finally, thirty-six participants (25 males and 11 females) were recruited from the university. Our participants aged from 20 to 57 (M = 26.86, SD = 6.53), and their game experience ranged from 0 to 20 years (M = 11.78, SD = 6.53). For the playing frequency and game skills, our participants played games ranging from 0 to 35 hours (M = 8.21, SD = 9.13) per week, most of our participants played games within a few weeks (27 of 36) and rated their game skills as ordinary (25 of 36). Their favorite game genres were action games

Table 4.2 Six experimental conditions. In the one-trial experimental situations, there are two OGD conditions, namely, low and high. In the two-trial experimental situations, there are four OGD conditions, namely, low & low, low & high, high & low, and high & high. The game levels of normal and hard modes are designed to correspond to these six conditions.

Trial situations	OGD conditions	Normal mode	Hard mode
One trial	Low	Level 0	Level 0
	High	Level 1	Level 2h
Two trials	Low & Low	Null	Level 0 and Level 1
	Low & High	Level 0 and Level 1	Level 1 and Level 2
	High & Low	Level 2n and Level 3n	Level 3h and Level 4h
	High & High	Level 1 and Level 2n	Level 2h and Level 3h

(27 of 36) and shooting games (25 of 36). Half of them (18 of 36) had played casual games before. The nationalities of our participants included Chinese (18 of 36), Japanese (9), Thai (7), Czech (1), and Bangladeshi (1); all used English, Chinese, or Japanese as their first or second language. Therefore, gameplay introduction was presented in these three languages, as were questionnaires, and interviews; professional workers handled the translation between the various languages. Demographics of participants in different groups or conditions used in this study are listed in Table 4.3.

Table 4.3 Participant demographics in different groups or conditions used in this study.

Groups or conditions	Numbers	Ages	Years of game	Experienced in casual games
All participants	36	20-57 (M = 26.86, SD = 6.53)	0-20 (M = 11.78, SD = 6.53)	50% (18 of 36)
Hard mode	26	20-57 (M = 26.42, SD = 7.29)	0-20 (M = 12.73, SD = 6.02)	50% (13 of 26)
Normal mode	10	21-33 (M = 28.00, SD = 4.03)	0-20 (M = 9.30, SD = 7.47)	50% (5 of 10)
Low	36	Same as All participants group		
High	35	20-57 (M = 26.91, SD = 6.62)	0-20 (M = 11.54, SD = 6.47)	48.57% (17 of 35)
Low & Low	26	Same as Hard mode group		
Low & High	35	Same as High condition		
High & Low	20	20-35 (M = 25.55, SD = 3.98)	2-20 (M = 13.75, SD = 5.61)	45% (9 of 20)
High & High	25	20-57 (M = 26.48, SD = 7.43)	0-20 (M = 12.44, SD = 5.96)	48% (12 of 25)

Materials and Apparatus After each level, the game results (success or failure) were provided as feedback, and participants were asked to complete a 10-item Likert questionnaire; see Fig. 4.6. The 10-item questionnaire after each level was developed based on our literature review and existing scales, which we list in Table 4.4. The question responses in this questionnaire ranged from 1 (strongly disagree) to 7 (strongly agree). Only after participants answered all the questions could they continue the game to the next level.

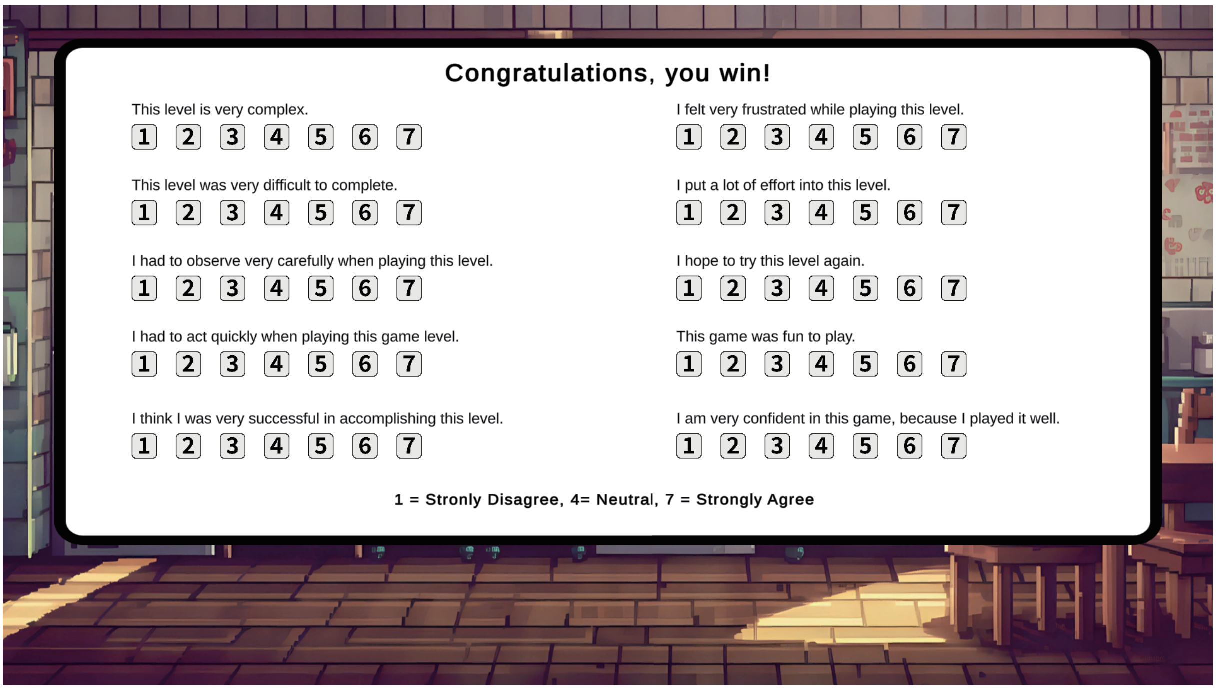

Fig. 4.6 Feedback and questionnaire after each game level.

After finishing all game levels, participants attended semi-structured interviews to provide their opinions. We list seven questions used in the interview in Table 4.5. Questions 1-3 focus on OGD and SGD, while Questions 4-7 ask for opinions about player experience, engagement, and self-efficacy. Question 3 was intentionally designed. This question specifically asks for the participant's opinion of the relationship between OGD and SGD. Questions 1 and 2 first ask participants to recall and evaluate the difficulty of each level. After answering these two questions, the participants can compare question 3's opinion with their views generated from the game experience, which we think can promote more keen insights into this relationship.

The experiment was conducted through a computer in a single room at the university. The game was built using the Unity3D engine and was run on the PC for this experiment. Specifically, the game ran on a 3.4 GHz Inter Core I7 CPU PC with Windows 10 and was played using a Logitech G203 mouse with 800 DPI. A 23" LG LCD

Table 4.4 The 10-item Likert questionnaire used in the experiment. This questionnaire is set in the game, and participants are required to answer it after playing each game level.

Dimensions	Question Contents
Game Complexity (SGD)	This level is very complex.
Game Completion Difficulty (SGD)	This level was very difficult to complete.
Game-playing Difficulty (SGD)	I had to observe very carefully when playing this level.
Game-playing Difficulty (SGD)	I had to act quickly when playing this game level.
Player Competence (SGD)	I think I was very successful in accomplishing this game Level.
Player Pressure (SGD)	I felt very frustrated while playing this level.
Player Effort (SGD)	I put a lot of effort into this level.
Player Engagement	I hope to try this level again.
Player Experience	This game was fun to play.
Player Self-efficacy	I am very confident because I played this game well.

Table 4.5 Semi-structured interview outlines.

Question No.	Question Contents
1.	You have just finished playing. Can you describe the content of these levels?
2.	How would you rate the difficulty of each level? Why?
3.	Do you think your success means this game level is easy and your failure means this game level is difficult? How does your performance (success or failure) in a particular level affect your perception of the game's challenge and difficulty? Why?
4.	Do you prefer the difficulty level to be easy, medium, or hard? Why?
5.	How does your perception of the difficulty level affect your enjoyment of the game? Why?
6.	In what situation would you want to try this game level again? Why?
7.	Do you feel confident in this game after playing? Why?

screen with a resolution of 1920 by 1080 was used. The data on player performance was collected in-game automatically, and the data analysis utilized IBM SPSS 26. The interview data were collected by audio recordings, transcribed, and manually coded following the thematic analysis protocol to identify themes (Braun and Clarke, 2006; Blandford, 2013). Two independent raters randomly coded all the open-ended answers and identified the themes based on each other's coding.

Procedure All participants were introduced to the content and procedure of this experiment after which they all signed the informed consent form. However, the OGD conditions designed in this experiment were not revealed to prevent impacting their

SGD. Subsequently, a form was required to be filled out to collect participants' demographic and game experience information, and the formal experiment followed. There is no practice before the formal experiment to avoid the influence of practice effect on OGD and SGD. Instead, the experimenter explained the game rules by showing how to play level 0.

Participants read level rules before each level and played from level 0 (i.e., the guide level) to the final level. After finishing level 0, the experimenter asked each participant to confirm whether they had understood how to play. Only after confirming their understanding could participants continue to level 1 and all levels could only be played once. There is no random order level-playing setting to control experimental conditions but there was a preset playing procedure (see Fig. 4.5).

All 36 participants completed the experiment; 26 were in hard mode and played five levels, and 10 were in normal mode and played four levels. All participants answered the questionnaire after each level and attended the interview after finishing the game. The entire experiment process for each lasted approximately 30 minutes.

4.3 Results

This section presents the following results: descriptive statistics of all OGD conditions and Wilcoxon matched-pairs signed rank test (Rosner et al., 2006) of the two-trial OGD conditions, the correlation analysis between OGD and other factors, mediation effect analysis of SGD on OGD, and interviews.

4.3.1 OGD Conditions in the One-trial Situation

The results of this subsection were used to test the hypotheses of H1, H1a, and H1b. In more detail, H1 assumes OGD and SGD match each other in one trial, H1a assumes low OGD matches to the low SGD in one trial, and H1b assumes high OGD matches to the high SGD in one trial. Because our SGD questionnaire's score range is 1-7, we regard the SGD score of 1-3 as subjectively easy and 5-7 as subjectively difficult. Therefore, to accept the H1a and H1b, the average SGD score by participants should be below 3 in the low OGD condition, while above 5 in the high OGD condition.

All thirty-six participants won level 0 and achieved the Low OGD condition in the one-trial situation, the mean scores of each dimension of SGD were: Game Complexity (M = 1.19, SD = 0.40), Game Completion Difficulty (M = 1.14, SD = 0.35), Game-playing Difficulty (M = 1.88, SD = 1.48), Player Competence (M = 6.31, SD = 1.41), Player Pressure (M = 1.36, SD = 0.96), Player Effort (M = 1.31, SD = 0.75); the mean score of all dimensions of SGD (scores of Player Competence are the reverse, the same in the following) was M = 1.49, SD = 1.11. Therefore, participants provided a low SGD score in the low OGD condition, which means H1a is supported.

Thirty-five participants lost level 1 (10 of 35) or level 2h (25 of 35) and achieved the High OGD condition in the one-trial situation. The mean scores of each dimension of SGD were: Game complexity (M = 4.57, SD = 1.80), Game Completion Difficulty (M = 4.71, SD = 1.45), Game-playing Difficulty (M = 4.24, SD = 1.97), Player Competence (M = 2.63, SD = 1.44), Player Pressure (M = 2.46, SD = 1.38), Player Effort (M = 4.17, SD = 1.65); the mean score of all dimensions of SGD was M = 4.25, SD = 1.86. Therefore, participants provided a middle SGD score in the high OGD condition, which means H1b is not supported.

In summary, H1 is only partially supported, which means OGD and SGD do not match each other exactly in the one-trial situation, see Fig. 4.7. Specifically, in the one-trial situation, low OGD matches low SGD, while a high OGD may not cause a high SGD.

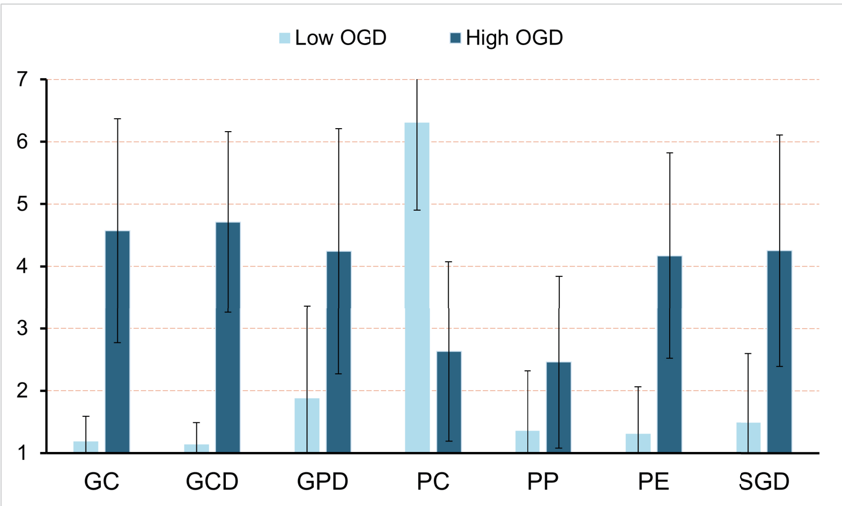

Fig. 4.7 The mean SGD scores and its six dimensions in the two conditions (low and high OGD) of the one-trial situation. GC: Game complexity, GCD: Game Completion Difficulty, GPD: Game-playing Difficulty, PC: Player Competence, PP: Player Pressure, PE: Player Effort.

4.3.2 OGD Conditions in the Two-trial Situation

The results of this subsection were used to test the hypotheses of H2, H2a, H2b, H2c, and H2d. Correspondingly to H2a-H2d, there are four OGD conditions in the two-trial situation: Low & Low, Low & High, High & Low, and High & High. Before the data analysis, we first conducted the Kolmogorov-Smirnov test (K-S test; Berger and Zhou, 2014) to determine whether our results obeyed the normal distribution and fit for the paired samples t-test. However, the results did not meet the requirements for using the t-test, so we decided to employ the Wilcoxon matched-pairs signed rank test (Woolson, 2007; Rosner et al., 2006) as an alternative method. Additionally, the Bonferroni correction (Armstrong, 2014) was also adopted to correct for multiple comparisons. Therefore, for the 0.05 level (2-tailed) and seven assumptions for each condition, there is significance only if p for each assumption in the 0.007 level (2-tailed). The results of testing are shown in Table 4.6.

Twenty-six participants of the hard mode won at level 0 and level 1 and achieved the Low & Low OGD condition. Results showed that there were significant differences in Game Complexity (z = 3.949, p < 0.001), Game Completion Difficulty (z = 3.886, p < 0.001), Game-playing Difficulty (z = 5.128, p < 0.001), and Player Effort (z = 3.894, p < 0.001). However, there was no significant difference in Player Competence (z = 1.952, p = 0.051) and Player Pressure (z = 1.387, p = .165). For SGD (all dimension scores were used and the Player Competence scores are reversed, the same in the following), there were significant differences between these two levels (z = 8.691, p < 0.001), even though players won at both levels. Therefore, H2a (there is no difference between SGDs in the two low OGD conditions) is not supported for SGD scores show differences between these two levels. Only the two dimensions of Player Competence and Player Pressure show no differences between these two levels and support this hypothesis.

Thirty-five participants won at level 0 and then lost at level 1 (10 of 35, all in the normal mode) or won at level 1 and then lost at level 2h (25 of 35, all in the hard mode), and they achieved the Low & High OGD condition. Results showed that there were significant differences in all six dimensions: Game Complexity (z = 4.798, p < 0.001), Game Completion Difficulty (z = 5.042, p < 0.001), Game-playing Difficulty (z = 4.375, p < 0.001), Player Competence (z = 4.839, p < 0.001), Player Pressure (z = 3.805, p < 0.001), and Player Effort (z = 4.299, p < 0.001). For SGD, there were also significant differences between these two levels (z = 10.929, p < 0.001). Therefore, H2b (SGDs in the two OGD conditions, low and then high, have differences) is supported in all dimensions and SGD.

Table 4.6 The results of Wilcoxon matched pairs signed rank test of four OGD conditions in the two-trial situation. The Bonferroni correction was used for multiple comparisons. Notes: *Denotes significant at the 0.007 level (2-tailed). GC: Game complexity, GCD: Game Completion Difficulty, GPD: Game-playing Difficulty, PC: Player Competence, PP: Player Pressure, PE: Player Effort. Subsequent tables use the same abbreviations.

OGD conditions	N	Variables	First level M(P_{25}, P_{75})	Second level M(P_{25}, P_{75})	Median M difference (First level - Second level)	z	p
Low & Low	26	GC	1.0 (1.0, 1.0)	3.0 (2.0, 4.0)	-2.0	3.949*	<0.001
		GCD	1.0 (1.0, 1.0)	2.0 (1.3, 3.0)	-1.0	3.886*	<0.001
		GPD	1.0 (1.0, 2.0)	3.0 (1.0, 5.0)	-2.0	5.128*	<0.001
		PC	7.0 (6.3, 7.0)	6.0 (5.0, 7.0)	1.0	1.952	0.051
		PP	1.0 (1.0, 1.0)	1.0 (1.0, 2.0)	0.0	1.387	0.165
		PE	1.0 (1.0, 1.0)	2.0 (2.0, 3.8)	-1.0	3.894*	<0.001
		SGD	1.0 (1.0, 1.0)	2.0 (1.0, 4.0)	-1.0	8.691*	<0.001
Low & High	35	GC	2.0 (1.0,4.0)	5.0 (3.5,6.0)	-3.0	4.798*	<0.001
		GCD	2.0 (1.0,3.0)	5.0 (4.0,5.0)	-3.0	5.042*	<0.001
		GPD	3.0 (1.0,4.0)	5.0 (3.0,5.8)	-2.0	4.375*	<0.001
		PC	6.0 (5.0,7.0)	3.0 (1.0,3.0)	3.0	4.839*	<0.001
		PP	1.0 (1.0,2.0)	2.0 (1.0,4.0)	-1.0	3.805*	<0.001
		PE	2.0 (1.0,3.0)	4.0 (3.0,5.0)	-2.0	4.299*	<0.001
		SGD	2.0 (1.0,3.0)	5.0 (3.0,6.0)	-3.0	10.929*	<0.001
High & Low	20	GC	6.0 (4.8,7.0)	6.0 (3.8,6.3)	0.0	2.121	0.034
		GCD	6.0 (4.8,7.0)	5.0 (3.8,6.0)	1.0	3.094*	0.002
		GPD	6.0 (5.0,7.0)	6.0 (5.0,7.0)	0.0	1.311	0.190
		PC	3.0 (1.0,3.0)	6.0 (5.8,7.0)	-3.0	3.839*	<0.001
		PP	2.0 (1.0,4.0)	1.5 (1.0,3.0)	0.5	1.466	0.143
		PE	5.0 (4.0,6.0)	6.0 (5.0,7.0)	-1.0	1.523	0.128
		SGD	5.5 (4.0,7.0)	5.0 (2.0,6.0)	0.5	3.507*	<0.001
High & High	25	GC	5.0 (4.0,6.0)	6.0 (4.0,7.0)	-1.0	2.153	0.031
		GCD	5.0 (4.0,6.0)	6.0 (5.0,7.0)	-1.0	2.306	0.021
		GPD	5.0 (3.0,6.0)	6.0 (5.0,7.0)	-1.0	4.594*	<0.001
		PC	2.0 (1.0,3.0)	3.0 (1.0,3.0)	-1.0	0.161	0.872
		PP	2.0 (1.0,4.0)	2.0 (1.0,4.0)	0.0	0.459	0.647
		PE	4.0 (4.0,5.0)	5.0 (4.0,6.0)	-1.0	2.058	0.040
		SGD	5.0 (3.0,6.0)	5.0 (4.0,7.0)	0.0	5.335*	<0.001

Only twenty participants of the hard mode achieved the High & Low OGD condition by losing level 3h but winning level 4h. Results showed that there were significant differences in Game Completion Difficulty (z = 3.094, p = 0.002), Player Competence (z = 3.839, p < 0.001). However, there were no significant differences in Game Complexity (z = 2.121, p = 0.034), Game-playing Difficulty (z = 1.311, p = .190), Player Pressure (z = 1.466, p = .143), and Player Effort (z = 1.523, p = .128). For SGD, there were also significant differences between these two levels (z = 3.507, p < 0.001). Therefore, H2c (SGDs in the two OGD conditions, high and then low, have differences) is supported for SGD scores show differences between these two levels. The two dimensions of Game Completion Difficulty and Player Competence also support this hypothesis.

Twenty-five participants of the hard mode lost Level 2h and Level 3h and achieved the High & High OGD condition. Results showed that there were significant differences in Game-playing Difficulty (z = 4.594, p < 0.001). However, there were no significant differences in Game Complexity (z = 2.153, p = 0.031), Game Completion Difficulty (z = 2.306, p = 0.021), Player Competence (z = 0.161, p = .872) and Player Pressure (z = 0.459, p = .647), Player Effort (z = 2.058, p = 0.040). For SGD, there were significant differences between these two levels (z = -5.77, p < 0.001). Therefore, H2d (there are no differences between SGDs in the two high OGD conditions) is not supported for SGD scores show differences between these two levels. However, besides the Game-playing Difficulty dimension, the other five SGD dimensions still support this hypothesis.

In summary, H2 (OGDs and SGDs match each other in two trials) is only half supported, which means OGDs and SDGs do not completely match each other in the two-trial situation. Furthermore, results show that SGDs differ between the two levels in all conditions, which implies players may not assess subjective difficulty merely based on their failure rate.

4.3.3 The Correlation Analysis Between OGD and Other Factors

All the data of all levels from participants in the experiment (N=170) were used in the correlation analysis. The OGD data were based on player performance, and their failure rate on each level was used for correlation analysis. The correlation analysis was conducted to test the hypotheses of H3, H4, H5, and H6. In more detail, H3 assumes OGD and SGD are positively correlated, H4 assumes OGD and player experience are negatively correlated, H5 assumes OGD and player engagement are positively correlated, and H6 assumes OGD and player self-efficacy are negatively correlated.

We conducted Pearson correlation to analyze the relationships between pairs of

OGD, the six dimensions of SGD, and SGD, see Table 4.7. The results of the correlation analysis showed that OGD was positively correlated to the five SGD dimensions of Game Complexity, Game Completion Difficulty, Game-playing Difficulty, Player Pressure, and Player Effort. By contrast, OGD was negatively correlated to the Player Competence dimension. OGD is also positively correlated to SGD. In addition, the six dimensions were also correlated, and each pair was positively correlated except for the pairs containing the Player Competence dimension. To summarize, H3 is supported and OGD is positively correlated to SGD.

Table 4.7 The results of Pearson correlation analysis (N=170) on the correlated relationships between pairs of OGD, the six dimensions of SGD, and SGD. Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

	OGD	GC	GCD	GPD	PC	PP	PE	SGD
OGD	1							
GC	0.452***	1						
GCD	0.578***	0.931***	1					
GPD	0.352***	0.731***	0.761***	1				
PC	-0.717***	-0.489***	-0.571***	-0.347***	1			
PP	0.219***	0.354***	0.364***	0.394***	-0.249**	1		
PE	0.356***	0.818***	0.796***	0.719***	-0.401***	0.451***	1	
SGD	0.569***	0.919***	0.939***	0.827***	-0.651***	0.548***	0.880***	1

Another Pearson correlation analysis was conducted to analyze the relationships between pairs of OGD, SGD, player experience, player engagement, and player self-efficacy, see Table 4.8. The results showed that OGD was positively correlated to engagement but negatively correlated to self-efficacy. However, no significant correlation (r = 0.072) existed between OGD and player experience. SGD was positively correlated to both player engagement and experience but negatively correlated to player self-efficacy. Furthermore, player engagement was positively correlated to player experience but negatively correlated to player self-efficacy, while there was no significant correlation (r = 0.126) between player self-efficacy and experience. In conclusion, OGD is positively correlated to SGD and player engagement and negatively related to player self-efficacy. However, OGD has no correlation relationship with player experience. Therefore, H5 and H6 are accepted, but H4 is rejected.

Table 4.8 The results of Pearson correlation analysis (N=170) on the correlated relationships between pairs of OGD, player experience (PX), player engagement (EN), player self-efficacy (PS), SGD. Subsequent tables use the same abbreviations. Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

	OGD	SGD	PX	EN	PS
OGD	1				
SGD	0.569***	1			
PX	0.072	0.352***	1		
EN	0.440***	0.597***	0.537***	1	
PS	-0.474***	-0.448***	0.126	-0.155*	1

In summary, H3, H5, and H6 are supported but H4 is not supported. Specifically, OGD is positively correlated to SGD and player engagement and negatively correlated to player self-efficacy. However, there is no correlation between OGD and player experience; instead, SGD and player experience are positively correlated.

4.3.4 Mediation Effect Analysis

All the data of all levels from participants in the experiment (N=170) were used in the mediation effect analysis. A bootstrap (5000) resample procedure (Preacher and Hayes, 2004; Wen and Ye, 2014) calculated the direct and indirect effects of OGD or SGD on player experience, engagement, and self-efficacy, see Table 4.9 and Fig. 4.8.

The result of the mediation analysis on OGD, SGD, and player experience showed that OGD significantly affected SGD (Path a; B = 1.870, p < 0.001), and SGD significantly affected experience (Path b; B = 0.480, p < 0.001). The indirect effect was significant (Path a*b; B = 0.898, p < 0.001), and the direct effect was also significant (Path c'; B = -0.650, p = 0.031). However, the total effect was not significant (Path c; B = 0.248, p = 0.349). Because the total effect was not significant but the indirect effect and direct effect were significant and had opposite signs, we suggested there was a suppressing effect (ratio of effects: 132.19%; Wen and Ye, 2014). The suppressing effect means that the negative effect of OGD on player experience is suppressed by SGD. Because the suppressing effect still can be regarded as a kind of mediation effect

Table 4.9 Mediation effect analysis (N=170). Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

Variables	Pathways	B	SE	95%CI	Description	Ratio of Effects
PX	OGD → SGD → PX (a×b)	0.898***	0.057	[0.155, 0.375]	Suppressing	138.19%
	OGD → SGD (a)	1.870***	0.209	[1.451, 2.279]	Effects	$\lvert a \times b/c' \rvert$
	SGD → PX (b)	0.480***	0.091	[0.241, 0.658]		
	OGD → PX (c')	-0.650*	0.298	[-1.235, -0.065]		
	OGD → PX (c)	0.248	0.264	[-0.270, 0.766]		
EN	OGD → SGD → EN (a×b)	1.358***	0.052	[0.194, 0.399]	Partial	66.33%
	OGD → SGD (a)	1.870***	0.209	[1.461, 2.279]	Mediation	$a \times b/c$
	SGD → EN (b)	0.726***	0.106	[0.519, 0.933]		
	OGD → EN (c')	0.690*	0.347	[0.009, 1.371]		
	OGD → EN (c)	2.048***	0.323	[1.416, 2.680]		
PS	OGD → SGD → PS (a×b)	-0.571***	0.051	[-0.249, -0.046]	Partial	31.59%
	OGD → SGD (a)	1.870***	0.209	[1.461, 2.279]	Mediation	$a \times b/c$
	SGD → PS (b)	-0.305**	0.093	[-0.488, -0.123]		
	OGD → PS (c')	-1.236***	0.306	[-1.836, -0.637]		
	OGD → PS (c)	-1.807***	0.259	[-2.314, -1.300]		

(MacKinnon et al., 2000), H7a (SGD mediates the effect of OGD on player experience) is supported (Fig. 4.8a).

The result of the mediation analysis on OGD, SGD, and player engagement showed that OGD significantly affected SGD (Path a; B = 1.870, p < 0.001), and SGD significantly affected motivation (Path b; B = 0.726, p < 0.001). The indirect effect (Path a*b; B = 1.358, p < 0.001), the direct effect (Path c'; B = 0.690, p = 0.049), and the total effect (Path c; B = 2.048, p < 0.001) were all significant, see Fig. 4.8b. Therefore, there is a partial mediation effect and the ratio of effects is 66.33%; H7b (SGD mediates the effect of OGD on player engagement) is supported (Fig. 4.8b).

The result of the mediation analysis on OGD, SGD, and player self-efficacy showed that OGD significantly affected SGD (Path a; B= 1.870, p < 0.001), and SGD significantly affected self-efficacy (Path b; B= -0.305, p = 0.001). The indirect effect (Path a*b; B= -0.571, p < 0.001), the direct effect (Path c'; B= -1.236, p < 0.001), and the total effect (Path c; B= -1.807, p < 0.001) were all significant, see Fig. 4.8c. Therefore, there is a partial mediation effect and the ratio of effects was 31.59%; H7c (SGD mediates the effect of OGD on player self-efficacy) is supported (Fig. 4.8c).

In summary, SGD suppresses OGD's negative effect on player experience and par-

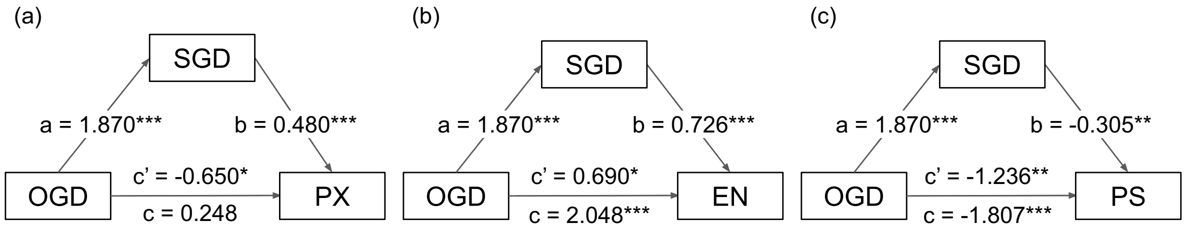

Fig. 4.8 Mediation effects of SGD to OGD on player experience (PX), player engagement (EN), and player self-efficacy (PS) (N=170). SGD suppresses the negative effect of OGD on PX and partially mediates the positive effect of OGD on EN and the negative effect of OGD on PS. Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

tially mediates OGD's positive effect on engagement and its negative effect on self-efficacy. Therefore, H7 (SGD mediates the effect of OGD on other players' individual factors) is accepted.

4.3.5 Interview

All 36 participants (hard mode 26 and normal mode 10) participated in our interview. The coding of open-ended answers identified the themes of: (1) OGD and SGD, (2) OGD, SGD and player experience, (3) OGD, SGD and player engagement, and (4) OGD, SGD and player self-efficacy. The identified themes are described in the following subsections.

OGD and SGD **The players' subjective assessments of each level's difficulty are not completely related to their success or failure at this level.** All participants regarded level 0 as easy; 88.4% of participants in hard mode thought level 1 in which they succeeded was easy, while participants in normal mode, who lost in this game level, differed in their opinions (3 easy, 4 medium, and 3 hard). 25 of 26 hard-mode participants lost in level 2n, but only 53.8% (14 of 26) thought the level was hard, 34.6% (9 of 26) thought it was medium, and 11.54% (3 of 26) considered it to be easy. Even though all hard-mode participants failed in level 3n and most of them (20 of 26) succeeded in level 4n, the majority of hard-mode participants (20 of 26) still rate the two levels as similarly hard. 8 of 10 normal-mode participants won level 2h and level 3h, but 4 participants thought these levels were easy, while 4 thought they were medium.

the players' SGD evaluations were based on various factors rather than merely OGD. Only 2 participants agreed with the statement in Question 3, i.e., that success means this game level is easy, and failure means it is difficult; 58.3% (21 of 36) participants partially agreed, and 36.1% (13 of 36) disagreed. Those participants who partially agreed provided their opinions from two aspects: success does not always mean easy, and failure does not always mean difficult. Some participants attributed their success to luck, effort, or skill enhancement, not low difficulty levels. *"The final level is still hard, but I am much better after practice (in the previous level), and I learned skills from my failures"* (P16). Some participants denied the level was difficult even though they lost; they attributed the failure to their own carelessness or lack of experience: *"I failed because I was careless just now, this level is not difficult"* (P29); *"This level is not difficult, I just don't play games like this very often"* (P7). Another participant said, *"Although I lost, I almost won; therefore, this level is not hard"* (p22). Most participants (9 of 13) who disagreed with this statement believed game difficulty is based on the content (e.g., game design, level of complexity, etc.) but not their performance. *"The last level is inherently difficult, whether I win or lose"* (P11); *"The more tiles there are, the harder it is, making it more possible to lose, but my performance does not represent the difficulty"* (P33). Other opinions include, *"Whether I win or not, it is easy because I am good at this game"* (P20), and *"There are many factors that affect whether I win or not, such as luck, difficulty cannot be represented by my win or lose"* (P33).

To summarize, the opinions of most participants did not support the existence of a matching relationship between OGD and SGD. On the contrary, participants introduced other factors, like game complexity, carelessness, effort, etc., to evaluate SGD in addition to their performance. These factors well matched the six dimensions we proposed.

OGD, SGD, and player experience **Players prefer games with challenges, however, the challenge is not directly determined by OGD but by the players' SGD interpretation.** Interestingly, nobody liked purely easy games, and most participants (33 of 36) preferred medium to hard game difficulty levels; 3 participants thought that increasing difficulty from easy to hard during play was better. When they were asked why they love challenging games. The most common reasons given were that they *"like challenges"* or that *"high difficulty makes the game fun"*. The reasons for liking medium levels of difficulty differ but can be summarized thus: they didn't want too many failures, but easy games are boring. One participant mentioned, *"I don't like to* *win too quickly, but I also hate losing all the time"* (P26). Other participants provided more detailed perspectives: *"Nearly winning a game (but failing in the end) excites me to keep trying"* (P4, P31), *"Medium-hard is not enough unless this game is very interesting"* (P35). Therefore, whether they win the game or not, the player experience varies according to how participants interpret their game results.

OGD, SGD & player engagement **Failing but believing in the possibility of success in the next attempt is an important reason for retrying.** Participants showed high consistency when answering this question: 77.8% (28 of 36) said that failure would make them retry. In more detail, they indicated that failure was hard to accept and they believed they would win on the next attempt, especially if they thought this game was easy. *"I don't want to admit this failure, I think I could win and want to try again"* (P33). *"I want to prove that I can win"* (P11). *"I will give up the next try if I think success is impossible"* (P4). Some also believed there were no unwinnable games, so one more try was worth it. *"Because I don't think there's any game that's impossible to win"* (P24). Other retry reasons included making progress, doing better than others, and sustained interest. Therefore, a high OGD (failure in the game) can be a good reason for players to sustain their engagement in the game, under the condition of not extremely high SGD (where the game would be considered unwinnable).

OGD, SGD, and player self-efficacy **OGD and SGD both affect player self-efficacy.** 69.4% (25 of 36) of participants agreed that the results whether success or not (OGD) had influenced their confidence, while others didn't think so. Of participants who were influenced, those who won the final level (20 of 25) felt confident, while the other five had no confidence because of failure. Some participants who denied the impact of the results pointed out that they had no confidence even after winning the easy level (SGD). They believed everyone would win at this easy level, so winning could not enhance their confidence. *"Maybe many people can do it, I'm not a special case. So, I am not so confident"* (P21). P28 said, *"I am not so confident (even though I finally won) because I made a lot of effort"*, which is related to SGD but not OGD. Three participants (P13, P17, P27) also said they had no confidence even after winning the final level. However, seven participants (2 losers and 5 winners of the final level) said they always had confidence, whether winning or losing. In short, to most participants, OGD and SGD have impacts on player self-efficacy.

Success after failures may strengthen confidence. Over half (12 of 20) of the winners at the final level in hard mode mentioned that success after failures strengthened their confidence, e.g., *"Winning the game after losing gave me a lot of confidence"*(P9). The previous failures seemed to strengthen the self-attribution of the following success. P33 said, *" I believe the success (after failure) is because I played very carefully this time."* P34 explained, *"The success after failure makes me confident because I proved I can do it if I try hard."* P21 also expressed the same view. In summary, failures contribute to the following success, and success after failure is helpful for players to strengthen their self-efficacy.

4.4 Discussion

This section discusses how our findings reveal the relationship between OGD and SGD and how these two game difficulty concepts affect players regarding experience, engagement, and self-efficacy.

4.4.1 The Relationship Between OGD and SGD

Our experiment results indicate that 1) SGD positively correlates to OGD, and 2) SGD partially matches OGD. In more detail, the data analysis shows SGD and its six dimensions are correlated to OGD, and the interview results also support this point: all these dimensions were mentioned as related factors when discussing the game difficulty. This finding is consistent with previous research (Adams, 2014; Fulmer and Tulis, 2013; Denisova et al., 2020; Ryan et al., 2006; Juul, 2004). However, both quantitative and qualitative results imply that SGD does not exactly match OGD, but there is a more complex relationship. For example, the interview results showed that participants proposed luck, effort, level of complexity, etc., as probably affecting their difficulty evaluations. Therefore, it can be considered that players evaluate difficulty not only according to their success or failure but more factors are considered.

We here discuss the experiment's results further and use the following abbreviations for the six dimensions of SGD: Game Complexity (GC), Game Completion Difficulty (GCD), Game-playing Difficulty (GPD), Player Competence (PC), Player Pressure (PP), and Player Effort (PE). For the high OGD condition of one trial, the average SGD score was around the middle, even though they failed. From the scores of different dimensions, players noticed their failure and provided low PC scores, but they still thought the game was not very complex or hard to complete (middle GC and middle GCD scores); they thought the playing difficulty was medium (middle GPD scores) and made appropriate efforts (middle PE scores). In addition, they found the pressure while playing the level was low (low PP scores). For the low & low OGD condition of the two trials, the results showed that all dimensions were regarded as different besides PC and PP, which means they noticed their success in these two levels and thought the pressures were similarly low. The results for low & high OGD conditions are as expected, and there is nothing to discuss. However, for the high & low OGD condition, even though SGDs were different, only the scores in dimensions of GCD and PC supported this difference. This result is still rational because the two levels had similar complexity and showed that participants experienced similar tough play processes. However, participants mentioned similar SGD for the two levels in the interview. This contradiction implies that, compared to assessing SGD by general feelings, the six-dimensional SGD evaluation method can provide more details. Finally, for the high & high OGD condition, SGDs were different, but only the significant difference in the GPD dimension supports it. This result indicates that even though the failure was the same, participants found one of them seemed to be more difficult in the process.

Interestingly, SGDs are significantly different from each other in every condition of the two-trial situation. This result is rational if we adopt a perspective other than OGD because the SGDs we compared are about different game levels in each condition. We also noticed that the Player Competence dimension highly matched OGD in all conditions. This dimension is about how players evaluate their performance. Therefore, due to adopting performance as OGD in this study, this result shows that players can assess their performance well based on the game results. This finding challenges the opinion of Constant et al. (2017), i.e., players are overconfident about game difficulty. We believe Constant and colleagues probably confused SGD with confidence about the future. As Huang et al. (2024) reported, player perceptions of game success in the present and the distant future differ. We argue that players attempt to stay positive about the game and future play (Klimmt et al., 2009), but they can still understand their failure and assess the game's difficulty rationally.

Based on these findings, we suggest that the main reason for the partially matching relationship between OGD and SGD is that OGD and SGD share different structures: SGD is multidimensional, but OGD is only quantified as player performance. This study made the first effort to adopt six dimensions to evaluate SGD. The results proved that these dimensions can better clarify players' views on various aspects of game difficulty, and the composed SGD by these dimensions also well represents the participants' general difficulty views. However, OGD is only about evaluating and predicting player performance with a simpler structure. Aponte et al. (2011a) proposed similar views when explaining this mismatch. More specifically, if the perception of Player Competence is adopted as the SGD, the two kinds of difficulty can match each other well. However, using player performance as the OGD is somewhat oversimplified when a more complex SGD composition is considered. Therefore, we recommend redefining OGD to contain more dimensions so that we properly represent its (natural) correlation with SGD. Based on our findings, it can be more comprehensive to include game task complexity and game completion progress into OGD.

In conclusion, there is a partial match between OGD and SGD due to the different structures of the two concepts. We propose to address this problem by redefining and expanding the concept of OGD.

4.4.2 How Game Difficulty Affects Players

Data analysis results show that 1) SGD is correlated to and influences player experience, engagement, and self-efficacy, and 2) OGD is only correlated with engagement and self-efficacy, but its influences, including on player experience, can be mediated by SGD. The results of the interview provide more details. Players enjoy medium or harder difficulty, but this difficulty is personally evaluated, and thus SGD; pure game results of success or failure do not affect their enjoyment directly. However, the majority of participants think the game results directly affect their retry motivation and confidence.

Regarding player experience, OGD seems to have no influence on it in general but SGD can mediate this influence. In more detail, OGD has a little negative direct impact on player experience, but the positive impact of SGD suppresses the effect of OGD. Our findings are consistent with some previous studies (Klimmt et al., 2009; Juul, 2009; Petralito et al., 2017). Klimmt et al. (2009) found players can keep a positive experience despite low performance. Juul (2009) proposed that "game enjoyment derives from player failure theory" but not from failure itself. It has been shown that negative events have the potential to form positive and meaningful experiences for players (Petralito et al., 2017). Considering SGD is also part of player experience (Ryan et al., 2006), this finding is rational. The game's result causes players to assess and explain SGD, and such subjective difficulty opinion directly affects their experience. Failure itself may have some negative effects on experience, but understanding and evaluating this failure towards future success promotes a positive player experience.

Regarding player engagement, OGD influences it positively, which means the harder, the more motivated players to keep playing. However, Lomas et al. (2017) found that increasing OGD decreases motivation, which shows a negative relationship between engagement and OGD. According to the interview results, participants mentioned that retry promoted by failure is based on their greater likelihood of success on the next try, which we believe is the pre-condition of this positively correlated relationship. It also indicates how SGD mediates this process from OGD to engagement: if players evaluate the game as unwinnable, they may lose motivation to challenge it. This finding supports the view of Juul (2009) that meaningful failure is crucial for players and partially explains the popularity of hardcore games like *Dark Souls Seris* (FromSoftware, 2012, 2014, 2016). Additionally, engagement is positively correlated to player experience, which implies that an increase in motivation can also enhance the player experience.

Regarding player self-efficacy, this is negatively affected by OGD and SGD, and SGD partially mediates the influence of OGD. This result shows that the more difficult the game is, whether game results or player perceptions, the less confident the player will be. This finding is consistent with recent research (Power et al., 2020; Nuutila et al., 2021) and with the mastery experience factor of self-efficacy theory (Bandura and Wessels, 1994; Stajkovic and Luthans, 1998). The results of the interviews provided additional information that although confidence is related to mastery experience, the two are not simply related. Failure before success will significantly enhance and consolidate the player's confidence caused by this success. This may be due to the mediation effect of SGD: failure forces players to evaluate the SGD seriously, while success positively reinforces this evaluation. We tend to liken it to a "spring effect," where applying excessive pressure to a spring can lead to its breaking (losing tenacity); however, the releasing spring will exhibit a remarkable rebound (substantial enhancement in confidence).

In summary, SGD, as a mediator, positively suppresses OGD's negative influence on player experience and mediates OGD's positive impact on engagement and negative impact on self-efficacy.

4.5 Conclusion

Our work explored the relationship between subjective game difficulty (SGD) and objective game difficulty (OGD) and explored their impacts on players through an experimental study. We found that OGD and SGD only partially match each other, which may due to their structure differences. Our findings support that SGD mediates the OGD's effect on player experience, engagement, and self-efficacy and indicate that SGD has an indispensable role in influencing players.

Chapter 5

OGD Measurement - Formulas and Validation

To measure OGD, we need to build a conceptual definition, identify its quantifying factors, and establish its operational definition and computational forms in steps. Having built its conceptual definition in Chapter 3, this chapter identifies its quantifying factors and proposes the operational definition and computational formulas of OGD as its measuring method. We test this new OGD measuring method with the experiment. The results showed that our method is effective in measuring OGD and has better validity than the other two methods.

5.1 Investigation on Game Tasks in Commercial Video Games

Since OGD is related to player interaction with game tasks, it is necessary to identify the quantifying factors from this interaction process. However, game tasks are often complex and involve a complex combination of game content and gameplay information. Therefore, this section describes how we conducted an investigation to summarize the basic game tasks of typical commercial games and then we identified two OGD quantifying factors based on players' interaction forms with these basic game tasks.

5.1.1 Investigation Procedure

It is necessary to first determine the scope of our investigation before it is conducted. Considering the goal of this investigation, we decided to investigate mainstream genres of commercial games but exclude serious games. The reason is that serious games are usually designed for specific serious goals and may confuse the game task classification.

Referring to the research that introduced typical game genres (Sellers, 2017; Heintz and Law, 2015; Teoh et al., 2020), we first confirmed 16 main game genres. Based on these genres, we determined the corresponding games on the Steam store (Steam, 2023). Steam is one of the most popular game distribution platforms on personal computers (PC) (Lin et al., 2019). For each game genre, we investigated a popular commercial game that can be purchased and downloaded from Steam. There were four rules for game selection.

1. Each game genre selects one game from the top 10 games that appeared in the "Top sellers" or "Top rated" lists.
2. The selected games should represent typical game genres' contents.
3. The selected game of each genre should be different.
4. The selected games should avoid duplication with each other of game content, gameplay, etc.

These rules were set since some genres overlap each other, and a game may belong to different genres. For example, shooter games can also be considered as action games. All the pages of this online store were accessed in April 2023. Finally, we settled on 16 popular commercial games corresponding to confirmed genres, see Table 5.1.

The specific game investigation process included more than 3 hours of playing time for each game (excluding the tutorial part) and more than 1 hour of video investigation. The video investigation was done by searching the name of the game and the keywords "Game", "Live", "Tutorial", and "Introduction" on YouTube (YouTube, 2023). Two researchers in our team conducted the game investigation. They both had more than 15 years experience in the game and had played all the listed genres of games before.

All the games and videos were played or watched on a PC with a 3.4 GHz Inter Core I7 CPU and a Windows 10 system. The input devices were a Logitech G203 mouse with 800 DPI, a Logitech K200 keyboard, and an Xbox One wired controller. Five of the 16 games were played with a controller (marked in Table 5.1), while the rest used a keyboard and mouse. The output devices were a 23-inch LG LCD screen with a resolution of 1920 by 1080 and a Logicool z313 Speaker.

Our goal was to answer four pre-set questions (see Table 5.2). The questions were set to confirm the following contents of each game: 1) core game tasks and required player skills to complete the tasks; 2) how each core game task is represented by output; 3) how players meet the task demands by input. It should be noted that we only fo-

Table 5.1 Investigation list of typical game genres and games. Notes: *Denotes this game was played with a controller.

Game Genres	Typical Games	Brief Descriptions
Shooter Games	PUBG: BATTLEGROUNDS	Games of shooting by weapons, usually firearms.
Action Games	Street Fighter V*	Games that emphasize fast actions and reactions.
Role-playing Games (RPG)	DARK SOULS™ III*	Games that need players to take on an in-game role.
Music Games	A Dance of Fire and Ice	Games that need player to act with music.
Sports Games	EA SPORTS™ FIFA 23*	Games that simulate one or more real-life sports.
Multiplayer Online Battle Arena Games (MOBA)	Dota 2	Games that need a team of several players to cooperate and compete with other teams.
Racing Games	Forza Horizon 5*	Games of driving vehicles for competition.
Platformer Games	Hollow Knight*	Games that need players to jump between platforms.
Strategy Games	Sid Meier's Civilization VI	Games that need players to make decisions strategically.
Real-time Strategy Games (RTS)	Age of Empires II	Games that need players to make decisions strategically in real-time.
Simulation Games	Cities: Skylines	Games of simulating real-world activities.
Board and Card Games	Slay the Spire	Games that use cards and pieces to play.
Gambling Games	HD Poker: Texas Hold'em	Games that need players to wager.
Puzzle Games	The Room	Games that need players to solving puzzles.
Fiction Games	Doki Doki Literature Club!	Games that need players to progress story.
Casual Games	Plants vs. Zombies GOTY Edition	Games with simple rules, shorter sessions, and require less learned skill.

cused on the core game tasks (i.e., closely related to the game genre and game progress) without including all game tasks in the games. For example, a sports game may also include tasks of in-game role development, but this would not be included. The investigation conducted by the two researchers was independent to avoid any biases. When the researchers believed the pre-set questions had been answered, they would terminate the selected game's investigation. Their investigation results were combined to produce an overall result for discussion.

Table 5.2 Pre-set four questions for game investigation.

Question No.	Question Contents
1.	What core game tasks does this game contain?
2.	What are the required skills in performing these tasks?
3.	What is the output of these game tasks to the player?
4.	What do these tasks require players to input?

5.1.2 Investigation Results and Discussion

According to the investigation of these 16 commercial games, we found that all game tasks only contain two output forms and five output contents, namely graphics, texts, and quantities in visual form, and sounds and voices in auditory form. Moreover, the input forms of game tasks can also be divided into two types, discrete and continuous. Discrete input content can be divided into reaction (including inhibition) and selection, while continuous input content is all about control. Specifically, reaction input requires that players react or inhibit their reaction to specific signals. Tasks with reaction input require players to react correctly within a specified time. Selection means players should select from different options to complete these tasks, and whether the tasks' goals are achieved is related to the correctness of the selected option. Tasks with control input need players to control game objects' location, direction, and motion (e.g., driving a vehicle or using a gun to target). The correctness of control (i.e., control precision in time and space) affects whether goals are achieved.

By referring to relevant research on cognitive psychology (Neisser, 2014) and human intelligence (Sternberg and Kaufman, 2011), we have isolated seven core abilities required to perform game tasks: reaction and control, which are related to physical game skills; perception, comprehension, memory, calculation, and reasoning, which are related to cognitive game skills. These key abilities constitute game skills and play essential roles in the players' game process.

Based on these findings and referring to the atomic challenges listed by Adams (2014) and the video game challenge inventory provided by Vahlo and Karhulahti (2020), we deconstructed all game tasks into 15 atomic game tasks in our first attempt, see Table 5.3. Atomic game tasks represent the elemental tasks that need to be completed in the game. For example, in a shooter game, the game task that requires the player to defeat the enemy may include three atomic game tasks in order: 1) *visual search task*: players need to find the enemy, 2) *object control task*: target the enemy by controlling the crosshair, and 3) *visual reaction task*: quickly input the gun shooting instructions as long as targeting the enemy. Since composite game tasks can consist of atomic game tasks, we believe that the determined difficulty of atomic game tasks can serve as a basis to represent the OGD in all game tasks.

After investigating these atomic game tasks, we found game difficulty appears mainly in the players' cognition and the input process of the player-game interaction. Similar views could also be found in other research which indicate that game difficulty is

Table 5.3 List of atomic game tasks. This list is built based on the investigation of typical games and core game tasks.

Atomic Game Task Types	Core Skills Required	Brief Descriptions
Visual Reaction Task	Reaction	Reacting to the visual signals properly.
Visual Search Task	Perception	Searching for required objects (or features) by visual scan.
Auditory Reaction Task	Reaction	Reaction to the auditory signals properly.
Auditory Reasoning Task	Reasoning	Reasoning unknown information based on sound or voice.
Knowledge Recall Task	Memory	Recalling Knowledge acquired from the real-world or games.
Graphic Memory Task	Memory	Memorizing specific graphics.
Graphic Comprehension Task	Comprehension	Understanding information in graphics.
Graphic Reasoning Task	Reasoning	Reasoning unknown information based on graphics.
Object Control Task	Control	Controlling game objects' location, direction, and motion.
Text Comprehension Task	Comprehension	Understanding information in the text.
Text Memory Task	Memory	Memorizing text information.
Text Reasoning Task	Reasoning	Reasoning unknown information based on the text.
Quantity Memory Task	Memory	Memorizing quantity information.
Quantity Calculation Task	Calculation	Calculating to solve mathematical questions.
Quantity Reasoning Task	Reasoning	Reasoning unknown information based on quantity.

highly related to the game tasks' physical and cognitive demands (Adams, 2014; Vahlo and Karhulahti, 2020; Denisova et al., 2020). However, we found that the proposed atomic game tasks were still conceptual. This list is valuable for game task analysis, which can help designers check which atomic game tasks are included in the designed game tasks. However, it helps little to quantify and measure the OGD of game tasks. It is still necessary to further detail the demand from a task that affects the OGD.

By referring to the research of Aponte et al. (2011b) and Pusey et al. (2021), we found that time and input correctness could be the core indicators for measuring OGD. In any case, whether the game task is successfully completed always depends on input from the player. Therefore, based on the investigation results and related research from psychology and HCI (MacPherson, 2018; Schmidt et al., 2018; Diaper and Stanton, 2003), we made a second attempt to summarize the abstract forms of all game tasks. We first divided the tasks into simple tasks and multitasking and then classified them into seven different task types based on the types and forms of input, see Table 5.4.

Specifically, the basic types of game tasks include single tasks and multitasking. Single tasks have only one goal, and multitasking contains several single tasks and an overall goal. Single tasks include simple input tasks, serial input tasks, and mixed input tasks. Simple input tasks require players to input only once during the interaction. Se-

Table 5.4 Basic game tasks are classified by different input forms. We also provide examples of using a keyboard and mouse to input.

Basic Game Task Types	Input Forms	Examples
Simple discrete input task	Discrete	Pressing the space key on the keyboard once.
Simple continuous input task	Continuous	Moving the mouse to the target area.
Serial discrete input task	Discrete	Pressing the space key twice on the keyboard.
Serial continuous input Task	Continuous	Moving the mouse to two target areas orderly.
Mixed input task	Discrete and continuous	Moving the mouse to the target area and clicking.
Serial multitasking	Discrete and continuous	Moving the mouse to the target area and then pressing the space key.
Concurrent multitasking	Discrete and continuous	Moving the mouse to the target area and pressing the space key at the same time.

rial input tasks contain only one specific type of input (e.g., using a mouse to click), but they require players to input more than once. Mixed input tasks contain multiple types of input (e.g., using a mouse to move and click) and require players to complete these input actions. Multitasking includes serial multitasking and concurrent multitasking, which are both constructed by more than one single task. Serial multitasking allows players to complete single tasks one by one, while concurrent multitasking requires players to perform different single tasks simultaneously.

Therefore, if all the game tasks can be regarded under these seven types of tasks or composites of these seven types, it becomes clear how to quantify OGD by the input factor: the game task's demand on player skills (i.e., OGD) could be detailed as a demand on the correctness of the player's input. Furthermore, to quantify real-time OGD, it is also necessary to include the time factor. Combining the two factors, OGD can be regarded as a relationship between the player's input incorrectness with the game task's correctness demand in a time period. In short, we found input time and correctness incorrectness could be the quantifying factors for OGD.

Additionally, because these seven game types are universal, they are ideal for displaying OGD measuring examples and can be used for general OGD measurement validation. Therefore, we applied them in Section 5.2's OGD computational examples and Section 5.3's experiment.

5.2 A New OGD Measuring Method

This section first proposes the computational formulas and operational definition of OGD using the two identified quantifying factors. As a new OGD measuring method, we provide seven examples to present how to use it to measure OGD. As an extra finding, the computational formula of the player's learning based on our method is also presented.

5.2.1 Computational Formulas and Operational Definition of OGD

Currently, OGD is usually quantified by performance factors such as failure and time. It is commonly assumed that the more failures and the more time consumed in the play, the harder the game is (Aponte et al., 2011b; Pusey et al., 2021). According to our investigation results, we found that input incorrectness is a promising replacement for the failure factor in OGD measurement. The reasons are: 1) input incorrectness is one of the direct causes of game task failure, 2) input incorrectness is applicable to all basic game tasks, and 3) input incorrectness can be measured in the interaction process for real-time OGD measurement. Therefore, we provide the following OGD computational formulas based on the factors of input incorrectness and input time.

We first define the correctness and incorrectness of input as:

- Correctness $c(t)$ refers to the correctness of the player's input actions at interaction time t,
$$c(t) \in [0, 1]$$
- Incorrectness $i(t)$ refers to the incorrectness of the player's input actions at interaction time t,
$$i(t) = 1 - c(t), i(t) \in [0, 1]$$

Considering the definition of OGD as "the dynamic meeting of the player's skill to the game task demand", we then define the demand of input correctness, the tolerance of incorrectness, and the amended incorrectness as:

- Correctness Demand $d(t)$ refers to the task demanded input correctness at interaction time t,
$$d(t) \in [0, 1]$$

- Error Tolerance $e(t)$ refers to the tolerance degree of the task to the input incorrectness at interaction time t;

$$e(t) = 1 - d(t), e(t) \in [0, 1]$$

- Amended Incorrectness $a(t)$ refers to the incorrectness of the player's input actions that is beyond the task tolerance at interaction time t,

$$a(t) = i(t) - e(t), e(t) \in [0, 1]$$

Therefore, we utilize the input correctness and the task's demand of the correctness to define the player's completion and the game's completion demand during game play:

- Completion $C(t)$ refers to the completion of the task by players until interaction time t,

$$C(t) = \int_0^t c(t)\, dt$$

- Completion Demand $D(t)$ refers to the whole demand for the completion of the task until interaction time t,

$$D(t) = \int_0^t d(t)\, dt$$

We believe that the overall OGD is determined by the overall degree of incorrectness during the game until completion. The real-time OGD is also not merely about input incorrectness at any moment. Instead, it is an accumulation of the player's input that fails to meet the demand before the calculated moment.

Therefore, to represent these two kinds of OGD in one formula, we define OGD as:

$$O(t) = \int_0^t \frac{a(t)}{d(t)}\, dt = \int_0^t \frac{i(t) - e(t)}{d(t)}\, dt = \int_0^t \frac{(1 - c(t)) - (1 - d(t))}{d(t)}\, dt$$

$$= \int_0^t \frac{d(t) - c(t)}{d(t)}\, dt = 1 - \frac{C(t)}{D(t)}$$

In addition, if there is an interval (t_a, t_b) of t such that $c(t) \geq d(t)$, then let $c(t) = d(t)$ in this interval. This provision ensures that OGD ≥ 0. More specifically, if there is $d(t) = 1$ of all t, then we have OGD:

$$O(t) = 1 - \frac{C(t)}{t}$$

Based on the OGD formulas, we propose the following operational definition of OGD: **an integral ratio of the amended incorrectness of the player's input**

to the game task's required input correctness within a given time frame. Defining $d(t)$ and $c(t)$ in specific game tasks is primary for the application of these functions to quantify OGD. In addition, the value range of OGD is $(0, 1)$, and the value "0.5" indicates a medium difficulty for the player to complete the task. A larger OGD value means a greater proportion of incorrect player input and more difficulty in completing the game task. Whether the player 1) has not completed the task, 2) has completed the task but does not meet the ideal demand, or 3) has completed the task successfully, OGD values can be calculated based on the player's input correctness and playing time. Therefore, compared to other OGD measuring methods, this method has wider applicability in measuring real-time OGD and overall OGD.

5.2.2 Seven Examples for OGD Measurement

The seven basic game tasks we summarized in the investigation are ideal for presenting how to calculate OGD by our formulas. Therefore, we provide eight examples for these seven tasks from Fig. 5.1 to Fig. 5.3.

For simple discrete input tasks, we assume a single-click task that requires players to click the mouse once to react to the visual signal within time t. We then assume this single-click task has $d(t) = 1$ of all t, a player acts at time t_0 and completes this task successfully. For simple continuous input tasks, we assume a moving task that requires players to move the mouse cursor into the destined area within time t. We then assume there are two tasks with different $d(t)$. In the first task that has $d(t) = 1$ of all t, a player acts at time t_0 and completes this task successfully. In the second task that has $d(t) = 0.9$, a player acts at time t_0 and completes this task successfully at time t_1, after which we have $c(t) = d(t) = 0.9$. These three tasks are shown in Fig. 5.1.

For serial discrete input tasks, we assume a double-click that requires players to do the double-click action by mouse to react to the visual signal within time t. We then assume this double-click task has $d(t) = 1$ of all t, and both click input share this demand with $d(t) = 0.5$ for each. Therefore, if the player completes the first click, the $c(t)$ of the player would be $c(t) = 0.5$. In this example, a player does the first click at time t_0 and the second click at time t_1 and completes this task successfully. See Fig. 5.2.

For serial continuous input tasks, we assume a double-moving task that requires the player to move the mouse cursor into two different destined areas by order within time t. We then assume this double-moving task has $d(t) = 1$ of all t, and the two input

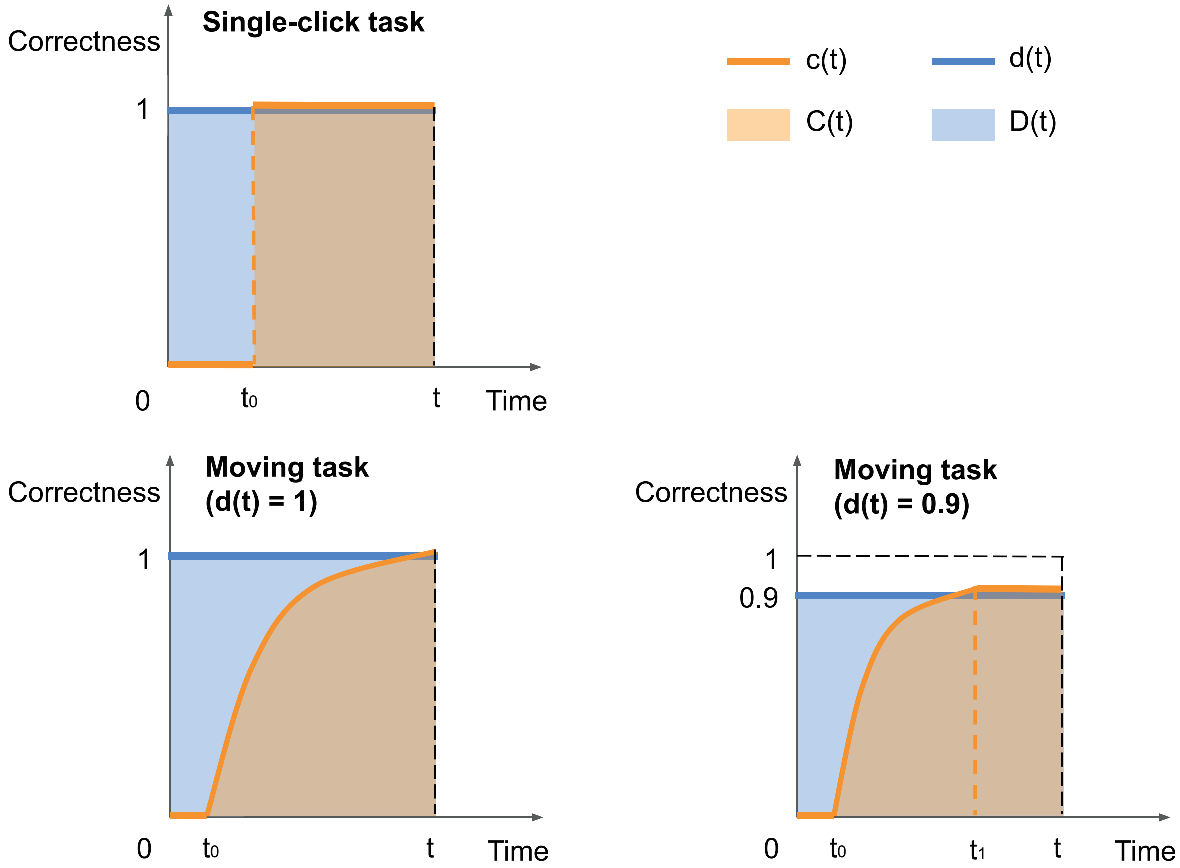

Fig. 5.1 There are three examples of single tasks: a single-click task (simple discrete input task), two moving tasks (simple continuous input task) with $d(t) = 1$ and $d(t) = 0.9$. In these three examples, a player acts at time t_0, and the tasks finish at time t. The orange color represents the player's input correctness, and the blue color represents the task's demand.

share this demand with $d(t) = 0.5$ for each, similar to the double-click task. In this example, a player begins to move at time t_0 and completes the first move at time t_1; the player then begins the second move, completes this move at time t_2, and successfully completes this task. See Fig. 5.2.

For mixed input tasks, we assume a moving & click task that requires players to input twice within time t: the first input is to move the mouse cursor into the destined area, and the second is to click the mouse once to react to the visual signal. Similarly, We assume this task has $d(t) = 1$ of all t, and the two input share this demand with $d(t) = 0.5$ for each. In this example, a player begins to move at time t_0 and completes the first move at time t_1; the player then takes the click action at time t_2 and successfully completes this task. See Fig. 5.2.

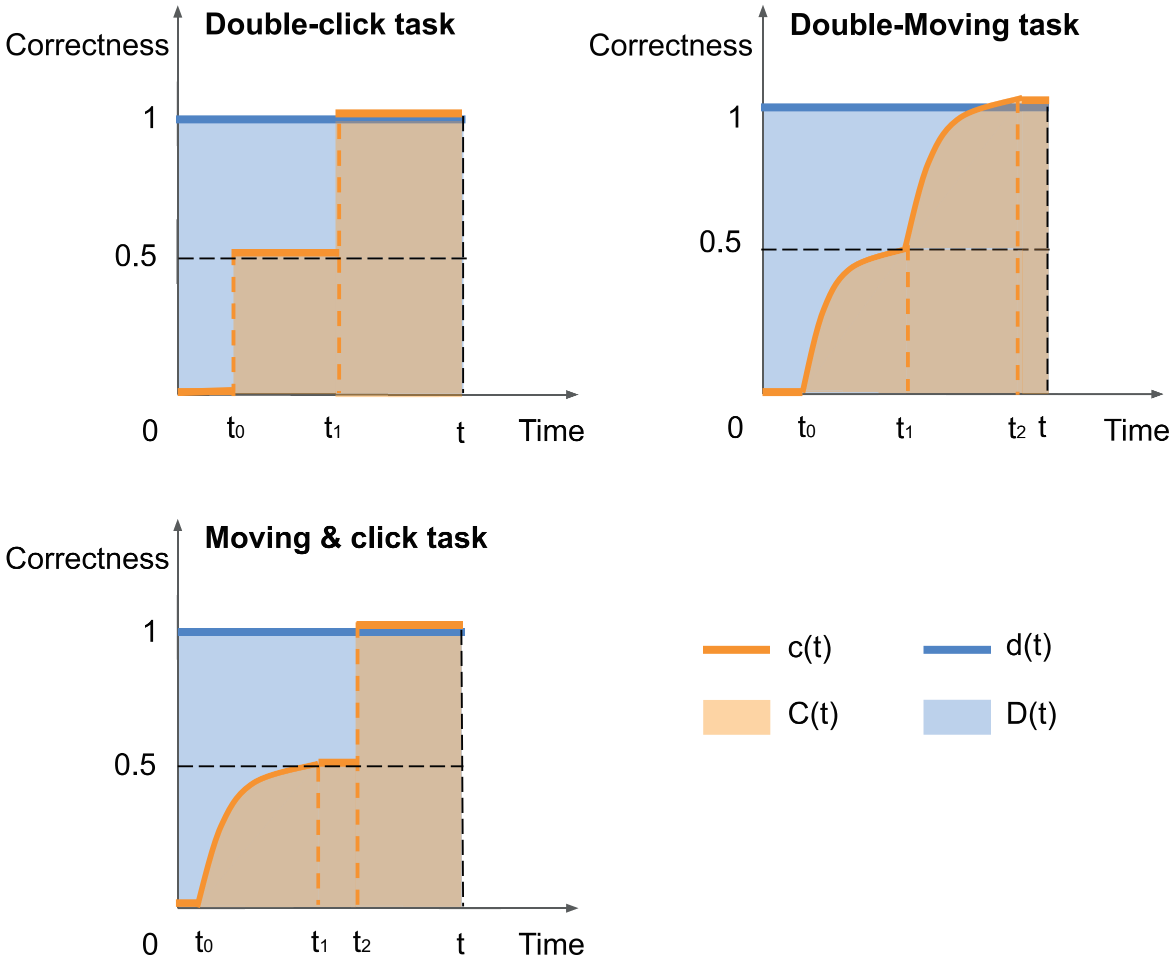

Fig. 5.2 There are two serial input task examples: a double-click task (serial discrete input task) and a double-moving task (serial continuous input task), and a mixed input task: a moving & click task. In these three examples, a player takes the first action at time t_0, then takes the second action at t_1, and the tasks finish at time t. The orange color represents the player's input correctness, and the blue color represents the task's demand.

For serial multitasking, we assume serial multitasking orderly combines a moving & click task and a single-click task. We then assumed that in this serial multitasking, the first task requires players to move the mouse cursor into the destined area and click the mouse once; the second task requires players to press the space key on the keyboard once to react to the visual signal within time t. This serial multitasking has $d(t) = 1$ of all t, and there would be three input that share this demand. The first two of them share this demand with $d(t) = 0.25$ for each, while the third input has the demand with $d(t) = 0.5$. In this example, a player begins to move at time t_0 and clicks at the first move at time t_2; this player makes the second click input at time t_3 and successfully completes the serial multitasking. See Fig. 5.3.

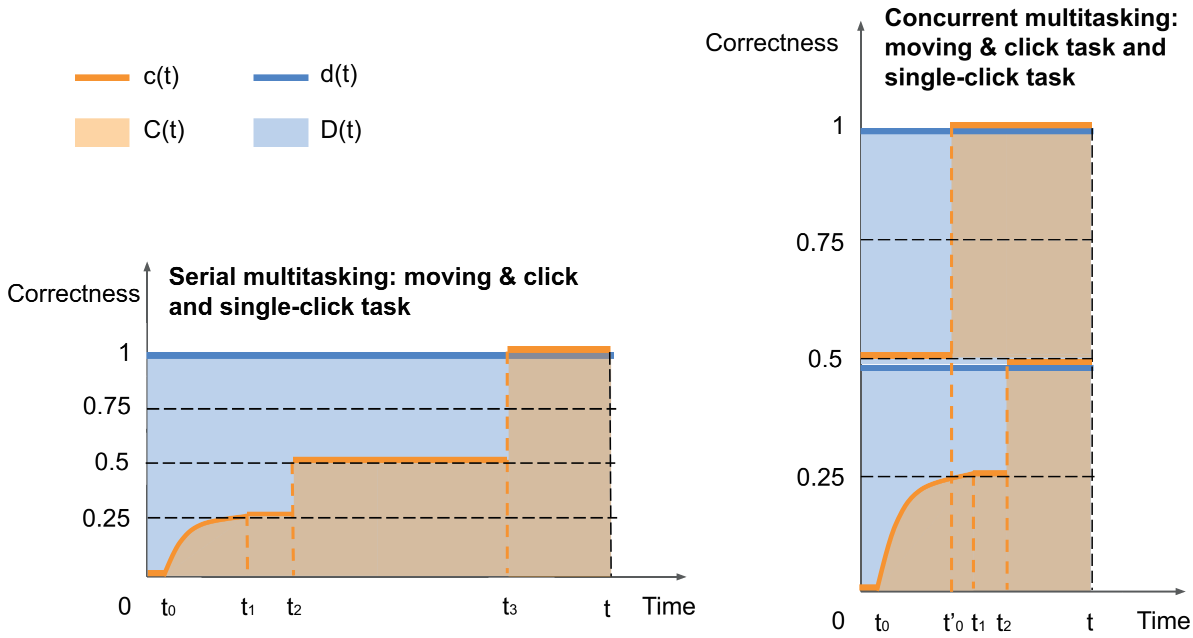

Fig. 5.3 Two examples of multitasking: serial multitasking and concurrent multitasking that both combine moving & click and single-click task. In these two examples, a player takes the first action at time t_0 and the tasks finish at time t. The orange color represents the player's input correctness, and the blue color represents the task's demand.

For concurrent multitasking, we assume concurrent multitasking also combines a moving & click task and a single-click task. We then assumed that within time t, this multitasking requires the player to move the mouse cursor into the destined area and click the mouse once, at the same time, the player is also required to press the space key on the keyboard once to react to a visual signal. This serial multitasking has $d(t) = 1$ of all t, and there would be three input that share this demand. The input in the moving & click task share this demand with $d(t) = 0.25$ for each, while input in the single-click task has the demand with $d(t) = 0.5$. In this example, for the moving & click task, a player begins to move at time t_0 and completes the first move at time t_1; the player then makes the one-click input at time t_2. For the single-click task, the player takes action at time t'_0 to press the key. Therefore, this player successfully completes the concurrent multitasking. See Fig. 5.3.

Based on these examples, it can also be found that graphically, OGD can be expressed as the ratio of the rest blue area that removes the orange area (i.e., $D(t)-C(t)$) to the entire blue area (i.e., $D(t)$). This is in line with our formulas. In short, our OGD measuring method is applicable to measuring OGD in these examples and has the potential to be applied to more complex games.

5.2.3 Computing Learning Based on Our Method

According to the literature, learning is to acquire specific cognitive constructs (schemas) for better automatic information processing (Sweller, 1994). For game players, learning can be simplified as "game skill enhancement". Based on our method, the learning process can be quantified by the Completion $C(t)$ changes measured over multiple game attempts. Therefore, we define learning as the change rate in player completion of the task and we assume the computational formula for learning $L(i)$ as the difference of completion between any two attempts in $i \geq 1$:

$$L(i) = C_i(t) - C_{i-1}(t), i \in N$$

In addition, if there is $i = 0$, then $L(i) = 0$, which means there is no learning before the first attempt.

Fig. 5.4 shows an example of a player's learning process in triple tries of the moving task. The player begins to move at time t_0 and t_1 in the first and second tries and fails to meet the d(t) at time t. Finally, the player successfully completes the task on the third try. From this example, we can find that learning is about how a player's completion portion of the task changes in different tries. When the completion portion of the task remains unchanged, no learning or progress is noted thus function $L(t) = 0$.

To sum up, we proposed an OGD measuring method. We quantified OGD by input incorrectness and time and provided quantifying formulas and an operational definition. In the next section, we report on our experiment to validate the effectiveness of our method by comparing our OGD measuring method with two other methods.

5.3 Experiment

To validate the effectiveness of our proposed method, we developed a game incorporating seven basic task types as summarized in Section 5.1 for experimental verification. In the experiment, we compared our method with the other two methods (failure rate and incompletion rate) in measuring the overall OGD and real-time OGD; self-report SGD results were also compared as an indirect standard. This section first presents

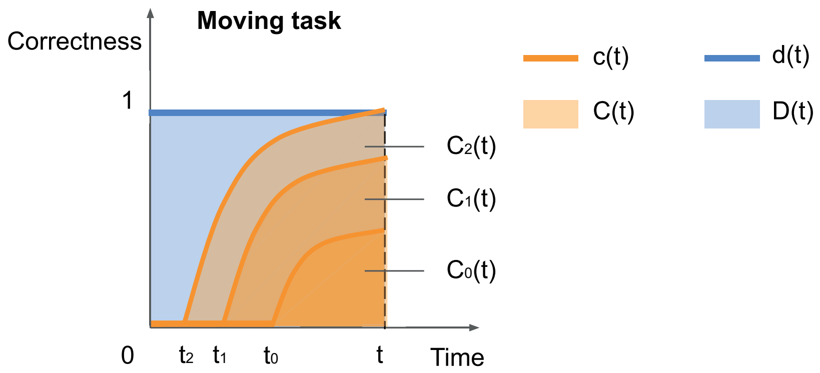

Fig. 5.4 Our method presents how a player learns through trying three times of the moving task. In the first try, a player acts at time t_0 and fails to meet the task demand at time t, and has a completion of $C_0(t)$. The player then tries two times and successfully completes the task on the third try ($C_2(t)$). The orange color represents the player's input correctness, and the blue color represents the task's demand.

our game design and experiment design, then we provide the experiment results and discussion.

5.3.1 Game Design

We designed a game, *Apple Farm*, using the Unity3D engine to validate our OGD measuring method. To ensure comprehensive validation, we adopted all the summarized seven basic tasks to design game levels based on our investigation results. For the same reason, we also designed three input modes by using the mouse & keyboard, controller, and touchscreen input to cover the mainstream input devices. Our game consists of a total of 7 levels, corresponding to the seven basic tasks. To include different levels of task complexity, the first 5 levels of single tasks all have three sublevels; while the final 2 levels, being multitasking levels, only have one sublevel. Players need to play this game in a linear order to complete a total of 17 sublevels, see Table 5.5.

Introductions to the gameplay were presented before each level. Countdowns were shown when players played each sublevel, and game results of either failure or success were presented after players finished each sublevel. SGD assessment was built into the game and displayed after each sublevel. Players needed to rate the difficulty of the sublevel from 1 to 7. See Fig. 5.5. In addition, all the performance and rating data of players during the game was collected automatically for OGD measuring.

Table 5.5 Basic game tasks with combination of different input forms.

Levels	Task Types	Sublevels	Level Contents
evel 1	Simple discrete	Level 1-1	React to a falling apple within 1.5s.
	Input Task	Level 1-2	React to a falling apple within 1s.
		Level 1-3	React to a falling apple within 0.5s.
Level 2	Serial discrete	Level 2-1	React to 4 falling apples within 1.5s.
	Input Task	Level 2-2	React to 5 falling apples within 1.5s.
		Level 2-3	React to 6 falling apples within 1.5s.
Level 3	Simple continuous	Level 3-1	Contact a big apple within 2s.
	Input Task	Level 3-2	Contact a medium-size apple within 2s.
		Level 3-3	Contact a small apple within 2s.
Level 4	Simple continuous	Level 4-1	Contact 4 apples in numerical order within 5s.
	Input Task	Level 4-2	Contact 5 apples in numerical order within 5s.
		Level 4-3	Contact 6 apples in numerical order within 5s.
Level 5	Mixed input	Level 5-1	Contact and hit 5 apples in numerical order within 10s.
	Task	Level 5-2	Contact and hit 10 apples in numerical order within 15s.
		Level 5-3	Contact and hit 15 apples in numerical order within 20s.
Level 6	Serial multitasking	Level 6	Finish a jigsaw and then contact and hit 15 apples in numerical order, complete these two tasks within 120s.
Level 7	Concurrent multitasking	Level 7	Contact 15 apples and catch them by a movable basket within 40s.

In more detail, all levels of *Apple Farm* were designed as follows:

Level 1: This level is about the simple discrete input task. The level sets an apple on a tree. After the game starts, the apple falls from the tree. As soon as the apple falls, players are required to react as fast as they can. This level has three sublevels: players must react to the falling apple within 1.5 seconds in level 1-1, 1 second in level 1-2, and 0.5 seconds in level 1-3.

Level 2: This level is about the serial discrete input task. The level sets some apples on the tree. After the game starts, all the apples fall from the tree at the same time. As soon as these apples fall, players are required to react to them as fast as they can. This level has three sublevels: there are 4 apples in level 2-1, 5 apples in level 2-2, and 6 apples in level 2-3, and players must react to these apples within 1.5 seconds in each of these sublevels.

Level 3: This level is about the simple continuous input task. The level sets an apple on the tree. As soon as the game starts, players are required to contact the apple

Fig. 5.5 The process of completing a game level of *Apple Farm*. Players were required to read the introductions to the gameplay before each level. The count-down was present in each sublevel, and game results were also presented after each sublevel. Players were then required to rate the SGD of each sublevel after confirming the game result.

as quickly and accurately as possible. This level has three sublevels: there will be one big apple in level 3-1, one medium-size apple in Level 3-2, and one small apple in level 3-3 and players must make contact within 2 seconds in each of these sublevels.

Level 4: This level is about the simple continuous input task. The level sets some apples on the tree. As soon as the game starts, players are required to contact these apples as quickly and accurately as possible. This level has three sublevels: there are 4, 5, and 6 apples in level 4-1, level 4-2, and level 4-3 respectively. Players must make contact within 5 seconds in each of these sublevels. Players must also contact these apples in numerical order from the smallest number to the highest number.

Level 5: This level is about the mixed input task. The level sets some apples on the tree. As soon as the game starts, players are required to contact these apples and hit them, as quickly and accurately as possible. This level has three sublevels: players will have 10, 15, and 20 seconds in level 5-1, level 5-2, and level 5-3, and 5, 10, and 15 apples in each level respectively. Players also must contact and hit these apples in numerical order from the smallest number to the highest number.

Level 6: This level is about serial multitasking. The level sets two tasks that need to be completed in order. The first task is to complete the jigsaw as quickly as possible. The second task is to contact some numbered apples and hit them, as quickly and accurately as possible. Players will be given 5 seconds to observe the jigsaw in the first task and then, after observing the jigsaw, 120 seconds are provided to complete the two tasks. There are 10 apples in the second task. Players also must contact and hit these apples in numerical order from the smallest number to the highest number.

Level 7: This level is about concurrent multitasking. The level sets two tasks that need to be completed simultaneously. One of the tasks is to contact the numbered apples on the tree. The other task is to catch falling apples by controlling a moveable basket on the ground. As soon as the game starts, players are required to contact these apples as quickly and accurately as possible; each apple will fall after being contacted. Players are required to control the basket and catch the falling apples as accurately as possible. The basket will auto-move randomly if there is no control from players; this ensures the concurrency of tasks. 40 seconds and 15 apples will be given at this level. Players must contact and catch these apples in numerical order from the smallest number to the highest number.

We designed the input forms for the three input devices to be as alike as possible. See Table 5.6. Specifically, except for level 7, discrete input through the keyboard & mouse is realized through the keyboard keys while continuous input is realized through the mouse movement; discrete input of the controller is realized through the buttons while continuous input is realized through the stick; discrete input of the touchscreen is realized through touch pointing while continuous input is realized through swiping on the screen. For level 7, to complete the multitasking in this level, two types of continuous input are required. Considered separately, the two types of input are the length of the key pressing time and the mouse movement in the keyboard & mouse mode, the length of the button pressing time and the stick control in the controller mode, and the length of the screen touching time and swiping in the touch screen mode.

To test our design, we conducted a pretest on the developed game. The goals of the pretest were to test: (1) whether the game was well-developed for playing and the data from input and SGD could be auto-collected accurately, (2) whether the complexity of each game level and sublevel is well designed with reasonable range to produce corresponding OGD. Three researchers aged 27-30 (M = 28.33, SD = 1.57) with different game skills in our team participated in this pretest. They played the game using all

Table 5.6 The game tasks' input design of three input modes of the seven game levels. "M&K" represents the input method of Mouse and Keyboard.

Levels	M&K Input Mode	Controller Input Mode	Touchscreen Input Mode
Level 1	Press the key once.	Press the button once.	Touch the screen once.
Level 2	Press the key multiple times.	Press the button multiple times.	Touch the screen multiple times.
Level 3	Move the mouse once.	push the stick once.	Swipe on the screen once.
Level 4	Move the mouse multiple times.	Push the stick multiple times.	Swipe on the screen multiple times.
Level 5	Move the mouse and press the key multiple times.	Push the stick and press the button multiple times.	Swipe on the screen and touch the screen multiple times.
Level 6	Click the mouse multiple times & Move the mouse and press the key multiple times.	Press the button multiple times & Push the stick and press the button multiple times.	Touch the screen multiple times & Swipe on the screen and touch the screen multiple times.
Level 7	Move the mouse multiple times & hold the key multiple times.	Push the stick multiple times & hold the button multiple times.	Swipe on the screen multiple times & hold the touch on the screen multiple times.

three input modes with different devices. The pretest was conducted through a PC and a Microsoft Surface Pro 7. The input devices for the PC were a Logitech G203 mouse with 800 DPI, a Logitech K200 keyboard, and an Xbox controller, while the input device of the Surface Pro is its touchscreen. The results showed that the game ran well on these three devices. In addition, the designs of game levels coped well with complexity and SGD levels. The three subjects succeeded in level 1 but failed levels 6 and 7, regardless of input mode. Therefore, they rated level 1 as very easy, level 6 and level 7 as very hard, and other levels as moderately hard. These results meant our game passed the pretest and was ready for the formal experiment.

5.3.2 Experiment Design

To validate the effectiveness of the proposed OGD measuring method, we conducted an experiment using the newly designed game. According to Kimberlin and Winterstein (2008), an effective measuring method should have good reliability and validity. Since the proposed method was based on theoretical deduction and different from other instruments like questionnaires, we did not adopt reliability tests (e.g., test-retest reliability or internal consistency) in our experiment. Instead, we attempted to test the validity of the proposed method.

For overall OGD measurement, the overall OGD means the measured OGD for any sublevel of the game for a certain participant. We tested the content validity, criterion validity, and discriminant validity of three OGD measuring methods, i.e., our method, failure rate method, and incompletion rate method. Content validation refers to the extent the method can measure each particular factor. Therefore, we compared the measuring range and measuring outliers of the three methods to test the content validation. Criterion validity is about the extent of correlation between the measurement results with the criteria. Due to the lack of direct criteria for OGD measurement, SGD (measured by self-report) is used as the indirect criterion for the correlation and regression analysis. Regarding the discriminant validity, we tested whether the OGD measurement results between sublevels with different complexity can be distinct. For real-time OGD measurement, the real-time OGD means the OGD per second in one sublevel of a certain participant. However, the failure rate method of measuring OGD and the self-report method of measuring SGD can hardly be used in real-time. Therefore, we test the content validity of the proposed method and the incompletion rate method by investigating their measuring range and measuring outliers.

This experiment is a within-subjects design; the independent variables are game complexity, and the dependent variables include three OGD results measured by 1) our method, 2) game incompletion rate, and 3) failure rate. SGD, as another dependent variable, is measured by self-report. The primary goal of the experiment is to validate (1) the proposed OGD measuring method as valid for measuring the overall OGD and real-time OGD, (2) the proposed method has better validity than the other two methods (i.e., failure rate method and incompletion rate method). We also used the experiment results to test the learning formulas proposed in Section 5.2.3.

5.3.3 Participants

Sixty participants (41 males and 19 females) were recruited and paid (1000 JPY per hour) from the university. Our participants were aged 20 to 57 ($M = 26.27$, $SD = 6.17$), and their game experience ranged from 0 to 22 years ($M = 12.32$, $SD = 6.01$). For the playing frequency and game skills, our participants played games ranging from 0 to 35 hours ($M = 7.63$, $SD = 8.87$) per week; most of our participants played games within a few weeks (48 of 60) and they rated their own game skills as ordinary (34 of 60). Their favorite game genres were action games (42 of 60) and shooter games (42 of 60). The nationalities of our participants varied, from China (33 of 60), Japan (18), Thailand (7), Czech (1), and Bangladesh (1); they use English, Chinese, or Japanese as their first or second language. Therefore, our game was presented in these three languages for the gameplay introduction and the SGD rating question; professional workers handled translations between the various languages.

5.3.4 Materials and Apparatus

The participants' SGDs for each sublevel were collected by The participants' SGDs for each sublevel were collected by asking participants the question, "You would rate the difficulty of the level you just played as...". Participants were required to rate their perceptions of SGD from 1 (very easy) to 7 (very hard). This rating was set in the game after each sublevel. Participants could not continue to the next level until they had finished the SGD rating.

The experiment was conducted through a PC and a Microsoft Surface Pro 7 in a single room at the university. The computer had a 3.4 GHz Inter Core I7 CPU with Windows 10 and a 23-inch LG LCD screen with a resolution of 1920 by 1080. The input devices were a Logitech G203 mouse with 800 DPI, a Logitech K200 keyboard, and an Xbox controller. The Microsoft Surface had a 2.4Ghz Inter Core I5 CPU and a 12.3-inch touchscreen with Windows 10. To create the same experimental condition, a resolution of 1920 by 1080 was also used in the Surface. The input device of the Surface was its touchscreen.

All the data on input, time, game results, and the SGD of players was collected automatically during game play. The OGD of each sublevel was calculated through three methods: (1) our proposed OGD, (2) failure rates, and (3) incompletion rate. More specifically, besides the OGD measured by our method, the failure rate is calculated by the goal achievement result, and the incompletion rate is calculated by the ratio of goal incompletion. For example, level 2-1 requires participants to catch 4 apples, if the participant catches 3 of them, the failure rate will be 100% and the incompletion rate will be 25%. For overall OGD, all the data of participants in all sublevels were used in computation for the average OGD value. While for the real-time OGD, we selected three participants' game data in playing Level 7 for real-time OGD analysis; one participant for each input device. We used Level 7 because it is the most comprehensive designed game level and contains all the task forms of lower game levels. In addition, the real-time OGD analysis focuses more on presenting the process of a certain participant. Therefore, it is not necessary to use all subjects' data. For the data analysis of the experiment results we utilized IBM SPSS 26.

5.3.5 Procedure

All participants were introduced to the content and procedure of this experiment, after which they all signed the informed consent form. Subsequently, a form was required to be filled out to collect participants' demographic and game experience information, after which the formal experiment was held. There was no practice before the formal experiment to avoid the learning effect or any other influence on their OGD and SGD. Instead, the experimenter explained the game rules by showing how to play each game sublevel.

Participants chose their preferred input devices to play the game. If they were familiar with more than one input device of these three, we would assign the input device to ensure an equal division between devices, see Fig. 5.6. It should be clarified that the input modes are not a variable or applied for grouping but for two purposes. One is to ensure this validation has a broader scope by including the mainstream input devices, while the other one is to minimize the impact on OGD and SGD of different players' familiarity with the devices.

Fig. 5.6 Participants played the game using their familiar input devices, which included a keyboard, a mouse, a controller, and the touchscreen of a Surface from left to right.

Participants read the introduction to the rules before each level. This introduction includes the rules for the whole level and for each sublevel in this level (where applicable). They were required to play from level 1 to level 7 and, except for level 7, each level could only be played once. Level 7 could be retried no more than 2 times (i.e., no more than three attempts in total) to test our formula in quantifying the learning effect in multiple attempts. This retry in level 7 is not forced but voluntary, which is to avoid measuring the OGD and SGD when the participant's motivation changes. We did not use the random-order playing design for two reasons: (1) each level (except for level 7) could be played only once, it is reasonable to expect increasing complexity of the game levels during play; (2) the natural improvement of players' game skills is measured in level 7 only. After finishing each sublevel, participants were required to rate the SGD of the sublevel.

All 60 participants completed the experiment. Each input device was assigned 20 participants. Among them, 26 tried level 7 three times. Participants rated the SGD of all sublevels after they finished or retried each sublevel. The entire experimental process for each participant lasted approximately 30 minutes.

5.4 Results

5.4.1 Overall OGD

For all participants for all sublevels, we first provided the values of overall OGDs measured by failure rate, incompletion rate, our proposed OGD, and the SGD results in Table 5.7. For level 7, only the data from participants' first attempts were used. Additionally, we standardized the SGD's data for better comparison with OGDs. SGD was measured by self-reporting in the score range from 1 to 7. Therefore, we used the function $f(x) = (x-1)/6$ to standardize the SGD's average values of each sublevel from the interval [1, 7] to [0, 1] to match the range of three OGDs.

For the content validity, the proposed method measured overall OGD values ranging from 0.289 to 0.879 and showed a more average value distribution from 0.4 to 0.6 (11 of 17 sublevels), with no outlier in the measurement results. In comparison, The failure rate ranged from 0 to 0.983, and most sublevels had failure rate values under 0.4 (13 of 17). The incompletion rate ranged from 0 to 0.633, and most sublevels had incompletion rate values under 0.2 (15 of 17). All participants succeeded in levels 1-1, 1-2,1-3, and 2-3. Therefore, the measuring results of failure rate and incompletion rate in these sublevels were the outlier of "zero". We also drew a figure to show the result trends of the three OGDs and the standard SGD for all game levels, see Fig. 5.7

For the criterion validity, we first performed a Pearson correlation analysis between the overall OGDs of the three methods and SGD based on the data of all sublevels, see Table 5.8. Sublevels 1-1, 1-2, 1-3, and 2-3 have zero values for failure rate and incompletion, so the analysis between them and SGD cannot be carried out. The results showed that for the data of all sublevels, the three OGDs were positively correlated with SGD. However, our method presented more positive correlations with SGD in separate sublevels (12 of 17) than the failure rate method (10 of 17) and the incompletion rate

Table 5.7 The calculated failure rate, incompletion rate (incompletion for short in this and the following tables), and proposed OGD, and reported SGD of all participants (N = 60) for all sublevels. The standard SGD is also calculated for comparison.

Sublevels	Failure Rate		Incompletion		Proposed OGD		SGD		Standard
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	SGD
Level 1-1	0	0	0	0	0.289	0.068	1.50	0.893	0.083
Level 1-2	0	0	0	0	0.385	0.074	1.62	1.121	0.103
Level 1-3	0	0	0	0	0.714	0.117	2.28	1.530	0.213
Level 2-1	0.017	0.129	0.004	0.032	0.513	0.090	2.33	1.174	0.222
Level 2-2	0.033	0.181	0.007	0.036	0.536	0.067	2.75	1.457	0.292
Level 2-3	0	0	0	0	0.563	0.057	3.25	1.590	0.375
Level 3-1	0.133	0.343	0.079	0.255	0.421	0.265	2.23	1.358	0.205
Level 3-2	0.117	0.324	0.092	0.283	0.470	0.224	2.45	2.556	0.242
Level 3-3	0.250	0.437	0.054	0.185	0.437	0.231	2.92	1.650	0.320
Level 4-1	0.133	0.343	0.042	0.114	0.374	0.141	2.62	1.391	0.270
Level 4-2	0.350	0.481	0.167	0.032	0.544	0.158	3.47	1.535	0.412
Level 4-3	0.367	0.486	0.169	0.272	0.527	0.186	3.80	1.695	0.467
Level 5-1	0.183	0.390	0.057	0.138	0.400	0.157	3.33	1.704	0.388
Level 5-2	0.450	0.502	0.097	0.134	0.467	0.134	4.32	1.578	0.553
Level 5-3	0.533	0.503	0.157	0.184	0.533	0.142	4.87	1.855	0.645
Level 6	0.767	0.427	0.633	0.382	0.789	0.167	5.97	1.104	0.828
Level 7	0.983	0.129	0.451	0.234	0.631	0.120	5.95	1.110	0.825

method (10 of 17).

We further conducted a multiple linear regression between the three overall OGDs and SGD to test the criterion validity. The analysis was to investigate the effects of OGDs in predicting SGD. Let variable failure rate be X_1, incompletion rate be X_2, proposed OGD be X_3, and the standard SGD be Y, the fitted regression model was: $Y = 0.728X_1 - 0.194X_2 + 0.453X_3 - 0.012$. The overall regression was statistically significant ($R^2 = 0.920$, adjusted $R^2 = 0.901$, F (3, 13) = 49.637, $p < 0.001$). It was found that X_1 (B = 0.728, $p < 0.001$) and X_3 (B = 0.453, p = 0.030) significantly predicted Y, but the predicted effect of X_2 (B = -0.194, p = 0.464) on Y was not significant. See Table 5.9. This result indicated that SGD could be predicted well by

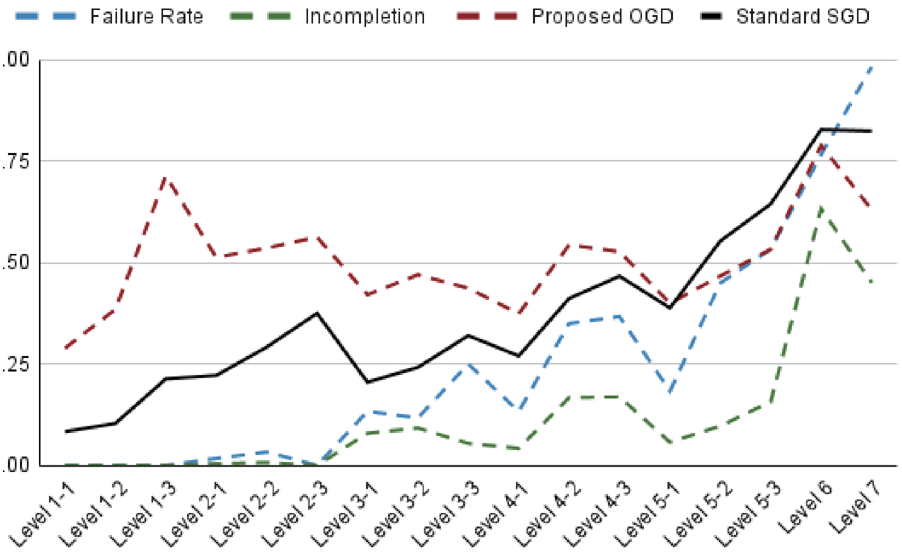

Fig. 5.7 The value trend of failure rate, incompletion rate, proposed OGD, and standard SGD of all participants ($N = 60$) for all sublevels.

combining the failure rate and our OGD measuring method, but the incompletion rate failed to take effect on the SGD prediction.

For the discriminant validity, Repeated Measures ANOVA and LSD post hoc tests were conducted to compare all the 60 participants' OGD results between sublevels within one game level. We analyzed Levels 1 to 5 because they all had three increasingly complex sublevels. See Table 5.10. The results showed that our method distinguished the sublevels' OGD in Levels 1, 2, 4, and 5, which was better than the other two methods; they were merely effective in distinguishing Levels 4 and 5.

5.4.2 Real-time OGD

We selected the participants based on their incompletion rate in Level 7 for better comparison. Three participants using different input devices with a 40% incompletion rate were selected. They rated the SGD of Level 7 as 6 (Mouse & Keyboard), 4 (Controller), and 6 (Touchscreen). Our results showed that the three participants' real-time OGDs by the two methods all decreased over the 40-second game time. We drew a figure to show the OGDs per second measured by our method and the incompletion rate method, see Fig. 5.8.

For the content validity, the proposed method measured real-time OGD values ranging from 0.689 to 1 (Mouse & Keyboard), 0.739 to 1 (Controller), and 0.685 to 1

Table 5.8 The Pearson correlation analysis results of failure rate, incompletion, proposed OGD (pOGD in the table for short), and SGD of all participants for all sublevels ($N = 60$). Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

Levels	Failure Rate × SGD		Incompletion × SGD		pOGD × SGD	
	r	p	r	p	r	p
Level 1-1	-	-	-	-	0.153	0.243
Level 1-2	-	-	-	-	0.122	0.352
Level 1-3	-	-	-	-	0.104	0.430
Level 2-1	0.298*	0.021	0.298*	0.021	0.328*	0.011
Level 2-2	0.161	0.220	0.161	0.220	0.257*	0.048
Level 2-3	-	-	-	-	0.112	0.396
Level 3-1	0.588***	<0.001	0.489***	<0.001	0.490***	<0.001
Level 3-2	0.264	0.410	0.195	0.135	0.251	0.053
Level 3-3	0.500***	<0.001	0.227	0.082	0.311*	0.016
Level 4-1	0.464***	<0.001	0.475***	<0.001	0.507***	<0.001
Level 4-2	0.602***	<0.001	0.565***	<0.001	0.515***	<0.001
Level 4-3	0.687***	<0.001	0.596***	<0.001	0.508***	<0.001
Level 5-1	0.493***	<0.001	0.494***	<0.001	0.487***	<0.001
Level 5-2	0.481***	<0.001	0.390**	0.002	0.349**	0.006
Level 5-3	0.550***	<0.001	0.451***	<0.001	0.450***	<0.001
Level 6	0.451***	<0.001	0.399**	0.002	0.381**	0.003
Level 7	0.112	0.393	0.418**	0.001	0.319*	0.013
All sublevels	0.649***	<0.001	0.565***	<0.001	0.465***	<0.001

(Touchscreen); while the real-time OGDs measured by the incompletion rate of the three participants all ranged from 0.4 to 1. In comparison to the incompletion rate method, our method provided more stable and fluent OGD values that change over time. In addition, the real-time OGD of the controller-use participant that was measured by the incompletion rate method showed a “zero” value between 0 to 6 seconds. These results indicate that our method had a better validity than the incompletion rate method in measuring real-time OGD.

Table 5.9 Summry of multiple linear regression analysis for variables of the three OGDs and standard SGD ($N = 17$). Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

	B	SE	β	t	p	95% CI for B
Constant	-0.12	0.089		-0.131	0.898	[-0.204, 0.181]
Failure rate	0.728	0.136	0.946	5.333***	$<$ 0.001	[0.433, 1.023]
Incompletion	-0.194	0.258	-0.150	-0.754	0.464	[-0.751, 0.362]
Proposed OGD	0.453	0.186	0.254	2.434*	0.030	[0.051, 0.855]

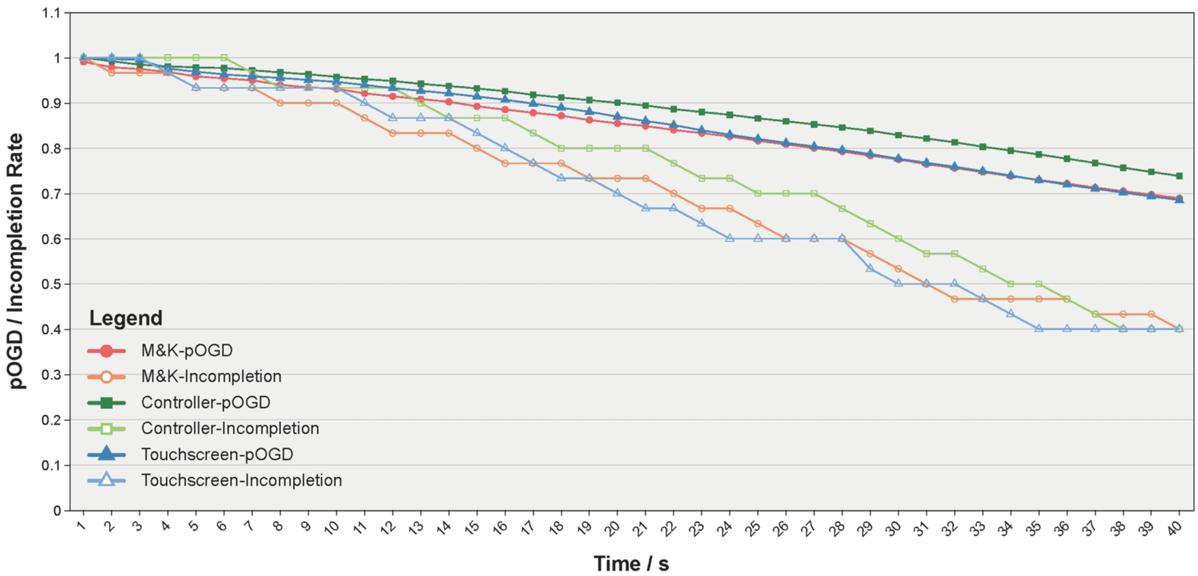

Fig. 5.8 The OGDs per second that were measured by the proposed method and the incompletion rate method of three participants ($N = 3$) using different input devices. "M&K" is the mouse and keyboard for short.

5.4.3 Learning

We calculated the mean learning value using our proposed formula $L(i)$ of the 26 participants in the three attempts in level 7. The results showed that the average function values of participants' learning on the three attempts changed from high to low: the first attempt ($M = 13.94$, $SD = 4.492$) was the highest, the second ($M = 2.99$, $SD = 3.871$) and the third ($M = 0.95$, $SD = 4.180$) attempts followed. This result indicated a lower speed in their learning with more attempts.

Regarding the learning that occurred while playing level 7, we also used the data

Table 5.10 The ANOVA results of sublevels' OGDs measured by the proposed OGD, failure rate, and incompletion methods ($N = 60$). For the values of M and SD of each sublevel please refer to Table 5.7. Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

Levels	Proposed OGD			Failure Rate			Incompletion Rate		
	F	*p*	*LSD*	*F*	*p*	*LSD*	*F*	*p*	*LSD*
Level 1	377.550***	<0.001	1<2, p<0.001	-	-	-	-	-	-
			1<3, p<0.001			-			-
			2<3, p<0.001			-			-
Level 2	7.014**	0.001	1<2, p=0.082	1.011	0.366	-	0.868	0.422	-
			1<3, p<0.001			-			-
			2<3, p<0.047			-			-
Level 3	0.646	0.525	-	2.3	0.103	-	0.377	0.687	-
			-			-			-
			-			-			-
Level 4	19.855***	<0.001	1<2, p<0.001	5.213**	0.006	1<2, p=0.008	6.028**	0.003	1<2, p=0.003
			1<3, p<0.001			1<3, p=0.004			1<3, p=0.003
			2<3, p=0.566			2<3, p=0.836			2<3, p=0.947
Level 5	12.555***	<0.001	1<2, p=0.012	9.157***	<0.001	1<2, p=0.002	6.442**	0.002	1<2, p=0.156
			1<3, p<0.001			1<3, p<0.001			1<3, p<0.001
			2<3, p=0.014			2<3, p=0.331			2<3, p<0.034

from the 26 participants to conduct a Repeated Measures ANOVA, see Table 5.11. We first compared the participants' three attempts in the proposed OGD and used LSD post hoc tests. There were significant differences in the three attempts for overall OGD measured by the proposed method (F (2, 75) = 6.725, p = 0.002). Additionally, the OGD in the first attempt is significantly greater than the second (p = 0.009) and the third (p = 0.001), but there is no significant difference between the second and third. This result was in line with the calculated results by our learning formula. However, the ANOVA results of SGDs in the three attempts showed no significant differences.

5.5 Discussion

This section first discusses our findings in the experiment. Subsequently, we discuss how to apply the proposed OGD measuring method in measurement, for designing difficulty, and for application in research.

Table 5.11 The ANOVA results of three attempts of level 7 about the incompletion, failure rate, proposed OGD, and SGD (N = 26). Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

Level 7	Attempts	*M*	*SD*	*F*	*p*	*LCD*
Incompletion	First	0.501	0.207	7.856**	0.001	1>2, p=0.012**
	Second	0.357	0.195			1>3, p<0.001***
	Third	0.283	0.204			2>3, p=0.196
Failure Rate	First	1	0	2.083	0.132	1>2, p=1.000
	Second	1	0			1>3, p=0.081
	Third	0.974	0.091			2>3, p=0.081
Proposed OGD	First	0.651	0.112	6.756**	0.002	1>2, p=0.009**
	Second	0.577	0.095			1>3, p=0.001**
	Third	0.553	0.108			2>3, p=0.398
SGD	First	6.23	0.951	0.627	0.537	1>2, p=1.000
	Second	6.23	0.652			1>3, p=0.335
	Third	6.00	0.938			2>3, p=0.335

5.5.1 The Validity of Our Method in OGD Measurement

In the measurement of overall OGD (Section 5.4.1), we tested the content validity, criterion validity, and discriminant validity of our OGD measuring method. For content validity, our method provides a relatively average distribution of OGD values and no outlier in measurement. For the criterion validity, we have investigated the correlation and prediction relationships between the measuring results of OGD and SGD. The results showed the measuring of OGD by our method was positively correlated with SGD in general, but not in all sublevels. According to Hunicke (2005) and Aponte et al. (2011a), players have more complex evaluation patterns of SGD. Our regression analysis results support this view and indicate that SGD can be well predicted by combining the failure rate and our measured OGD. For the discriminant validity, our method distinguishes most tested game levels (4 of 5) and shows good discriminant validity. However, our method failed to distinguish the sublevels in Level 3, which requires players to contact different sizes of apples. This may indicate that the factor of the apple sizes is not well controlled in the game complexity design. In short, the experiment shows that our method has good validity in measuring the overall OGD.

In the measurement of real-time OGD (Section 5.4.2), we tested the content validity of our OGD measuring method. The OGD real-time values measured by our method decrease over time with no outlier value and a fluent value change. This is because we adopt the integration method in OGD's computation; the OGD value calculated is about the entire playing process until any selected moment. From the conceptual definition of OGD, real-time OGD is about the dynamic meeting result of the player's skill to the task demand. Each input of the player changes this result during the game play. Therefore, as long as the player's input is not completely incorrect, the value of OGD will decrease as the player's play progresses over time.

In summary, our method is valid for measuring overall OGD and real-time OGD based on the experiment results.

5.5.2 Comparing Our Method to Other OGD Measuring Mehtods

We compared our method with the other two OGD measuring methods (i.e., the failure rate and incompletion rate methods) through the experiment. The failure rate method is more applicable for overall OGD measurement and is commonly used in two scenarios: when a single player attempts to complete a game level multiple times and when multiple players attempt to complete the same game level. However, this method is ineffective in measuring the overall OGD of one player in a single attempt. The incompletion rate method is better in this context. However, both measures have a value of zero when the player successfully completes the game level. In addition, these two methods are not very effective in distinguishing game levels with different complexity.

We believe the reason is that these two methods are only sensitive to the game results (i.e., results of the game task's goal achievement) in measuring the overall OGD. More specifically, the failure rate and incompletion rate methods do not contain process factors but depend on the game results in the overall OGD measurement. This may cause a process information loss while playing and thereby affect the OGD measuring results. Consider a scenario where two players successfully complete a game task in a single attempt, but with different completion times. In such a case, the OGD measurement results obtained using failure rate and incompletion rate methods would both be zero. However, if one of the players succeeds in the beginning while the other one completes the game in the final second, their OGD on the game should be different. In comparison, our method is established on the two interaction-related factors of input and time. In the above example, the overall OGD measured using our method would be significantly lower for players with less playing time. Therefore, our method is more progressively related to the process than the game results; this is better in discriminant validity and can avoid extreme values whether the participant succeeds or fails in a single attempt.

As another method to measure overall OGD, Pusey et al. (2021) suggested that overall OGD in puzzle games can be measured by the "time taken to solve the puzzle" or the "number of incorrect/failed attempts". Their method is similar to ours in that it includes both incorrectness and time factors to measure OGD. However, their method is not a universal approach but a specific analysis tool for puzzle games. Therefore, their method cannot measure the overall OGD precisely through computation. Furthermore, the demands of tasks are also not considered or clearly presented in their method, which fails to represent the concept of OGD.

For the measurement of real-time OGD, the method by Aponte et al. (2011b) quantifies OGD as a probability of player failure function at a specific time. However, they still adopted the failure rate method to compute the failure probability, which is unsuitable for real-time OGD measurement (Aponte et al., 2011a). The incompletion rate method, as an approach to reflecting the player's probability of failure, can measure OGD in real-time. More specifically, the incompletion rate method quantifies "how much the player does not complete at the moment" to measure real-time OGD. However, it merely measures how much of the task the player does not complete at any particular moment. By comparison, our method further quantifies the incompletion of the process more precisely by including the task demand and time. It can be found that the two methods provide widely different values in the experiment, for example, 0.739 (our method) and 0.4 (incompletion rate method) for the controller input in the final second. The incompletion result of value "0.4" denotes the player completing 60% of the task when finishing the game. Considering the connotation of OGD, the player is not skillful enough to successfully complete the task and OGD should be harder than a medium level. Therefore, the OGD value "0.739" seems more correct than "0.4" in the difficulty representation. Their SGD ($M = 5.33$, $SD = 1.15$) rating also indicates they agree that the difficulty of Level 7 is not low.

In summary, compared to other currently available methods, the proposed OGD measuring method has better validity and applicability in both overall and real-time OGD measurements.

5.5.3 Other Implications from the Experiment

What SGD Means to OGD Measurement Measuring OGD is usually used to predict the SGD without measuring it. Although it is natural to assume that OGD and SGD match each other (Constant et al., 2017), research has shown a more complex result than this assumption. Constant et al. (2017) found that players seem to easily underestimate OGD and become overconfident about future success. Furthermore, SGD may be affected by many factors (Ryan et al., 2006; Denisova et al., 2020), e.g., characteristics of players (Tondello and Nacke, 2019) or perceived effort (Ryan et al., 2006), and OGD is only one of the factors. Considering that SGD is the players' perceptions of game difficulty from the playing experience, it is reasonable that players may include various differing factors in their SGD evaluation.

Therefore, it is necessary to consider these two difficulties respectively. On the one hand, SGD can be a reference for OGD measurement but not a reliable criterion. Measuring OGD should reflect more on whether the results represent the connotation of OGD but not simply pursue matching the SGD. On the other hand, researchers and designers should be aware that relying solely on OGD measurement results may not accurately predict SGD. According to our experiment results, combining more measuring methods and results (e.g., our method and failure rate method) may improve the prediction effect on SGD.

OGD and Learning The results from the analysis of the three attempts at level 7 reflect the players' learning process during the game (Section 5.4.3). Although no learning effect is shown in the SGD, there are significant differences in our OGD results, especially between the first and second attempts. In addition, the values of our learning formula $L(i)$ indicated a decreasing trend. These results indicated that players played better in the second attempt than in the first attempt, but their performances were similar in the second and third attempts. It can be concluded that the player's learning speed gradually slows down in three attempts, which is also in line with the learning curve theory (Speelman and Kirsner, 2005). We suggest there may be three reasons for there being no significant difference in SGDs. Firstly, although most players made progress, they had not successfully completed the game tasks, which may affect their SGD. Secondly, the complexity of the game task remains unchanged, so some players may tend to evaluate the difficulty over the three attempts as close. Finally, based on the attitude theory of ego defense (Katz, 1960), people tend to stay consistent with their previous attitudes, so players may tend to maintain their evaluation of difficulty. In short, SGD is more complex for evaluation and stays stable over these attempts. Nevertheless, the results of our proposed learning function still demonstrate the player learning process well and are consistent with current learning theory.

OGD and Research For game difficulty research, our method provides more possibilities for understanding the relationship between OGD and other research objects. For example, by measuring real-time OGD, it becomes possible to investigate how players' physiological states and game difficulty relate during the interaction process. In comparison, the failure rate method cannot provide insights into the interaction process. As a reference, Chanel et al. (2008, 2011) assessed the player's emotional state during game play to distinguish game difficulty. However, difficulty in their study was quantified by game complexity levels. Therefore, our method supports further study regarding this relationship by providing a more effective real-time OGD quantification approach. Measuring the overall OGD is also valuable for studying the relationships of OGD with other factors, e.g., self-efficacy (Power et al., 2020). In addition, some serious game studies also require OGD measurement and evaluation to achieve their serious goals better. For example, Anguera et al. (2013) used a serious game to train the cognitive ability of the elderly, however, they still adopted the failure rate method to design the adaptive difficulty mechanism. Therefore, a precise and effective OGD measurement can undoubtedly support the better achievement of the serious goal in research.

5.5.4 Steps to Measure OGD Using Our Method

We have developed and validated a new OGD measuring method. However, it is still necessary to clarify how to use this method to conduct the overall and real-time OGD measurements. We suggest that the measurement be conducted in three steps: (I) Splitting the game tasks, (II) Defining the factors, and (III) Measuring & Computing.

Step I. Splitting the game tasks. According to our investigation, game tasks can be split and there are seven basic types. Therefore, to measure OGD, first, split the tasks in the game. For example, a shooting game's classical task is a simple mixed input task: target and track the enemy (continuous input) and shoot it when targeting (discrete input). For the Go game, the task can be split into the same subtask:

choose the best position for the piece in each round, which is a simple discrete input task. However, some games may be complicated containing a lot of tasks that may be hard to split. In this case, we suggest focusing on the core game task and considering the contribution of each task required to complete the game. For example, MOBA games usually contain the core game task of destroying the enemy's primary building (Mora-Cantallops and Sicilia, 2018). This core task can be split into three subtasks: 1) strengthening your in-game character, 2) winning the local battle in the game, and 3) destroying the enemy's buildings. On this basis, these subtasks can be further split by the proposed basic game tasks. Additionally, the computational weight for each basic task should also be considered. In this example, the third subtask plays a more critical role in winning the whole MOBA game. Therefore, the computational weight of this subtask in the OGD measurement needs to be increased.

Step II. Defining the factors. We have identified input incorrectness and time factors in measuring OGD. However, these two factors should be further defined for the specific game when measuring. To determine the input incorrectness factor should first define what correctness means in the game. It is nearly impossible to directly define the impact of a single input action for completing the game since many games are very complicated (and factors/tasks are often interdependent). Therefore, in the first step, we suggest researchers and designers split the game tasks into basic tasks and weigh each of them by computation. There is no standard for assigning computational weights, but rather it depends on research or design needs.

The input correctness for the basic tasks can be determined based on the input forms. For discrete input, correctness usually refers to the correct choices being made (e.g., choosing better positions for pieces) or to the correct timing of the input (e.g., reacting to the signal). For continuous input, correctness commonly refers to the controlled object's motion state, such as speed or position. In racing games, players need to avoid "driving" the vehicle off the road. In platform games, jumping between platforms requires players to control the character's trajectory to the target position with the necessary speed before jumping.

Regarding the time factor, it is not as difficult to determine the input time as it is for the time frame of game completion. Although some games have a time limit for task completion, many games allow players to challenge freely, without time restrictions. For example, Street Fighter V is a famous action game that limits the time of each playing round to 99 seconds (Capcom, 2016), while most strategy games (non-real-time strategy) do not have time limits for player decision-making (Caldwell, 2004). Therefore, we suggest that in the games without time limits, the time frame should be scoped until the player completes or gives up.

Step III. Measuring & Computing. After the two steps, the OGD of the game can be measured by measuring the two factors and applying our proposed computational formulas. The real-time OGD can be measured by a second or any other time interval, while the overall OGD is the OGD measured in the final time. By applying our method, the OGD can be measured by the game system automatically and thus more easily.

Chapter 6

SGD Measurement - Scale Development

In this chapter, we developed the subjective game difficulty scale (SGDS), by three steps of item generation, scale development, and scale testing (Fang et al., 2013; Moore and Benbasat, 1991). The developed scale shows good reliability and validity in SGD measurement. We also introduce the usage methods of this scale.

6.1 Stage i: Item Generation

We created the items for each dimension based on current scales and literature (see Table 6.1 and Table 6.2). We noticed that the two dimensions belonging to the game task part can be measured before or after playing, while the other dimensions can only be measured after playing. This is because players can form their perceptions of game complexity and their completion probability based on observation of the game or others' playing. However, this estimation which is achieved without actually playing the game is more like a first impression of the game, which may not be accurate and would likely be revised after playing. In comparison, other dimensions measure players' feelings produced in and after the specific playing process and thus cannot be measured before this process. In addition, the player competence dimension asks the player, based on their playing, to evaluate their present competence in the game. Therefore, we utilized the present tense to create the items of the three dimensions: game complexity, game completion difficulty, and player competence. For the other three dimensions, i.e., game-playing difficulty, player pressure, and player effort, we used the past tense to create the items.

For the dimension of game complexity (GC for short), we utilize the key elements of game complexity to create the items based on previous research (Liu and Li, 2012; Maynard and Hakel, 1997). A game is considered complex if it has an excessive number of game elements and element types (items GC0001 and GC0002), intricate relationships between elements (item GC0003), an overwhelming amount of information (item GC0004), or if the information provided is too vague (item GC0005). Similarly, if the game rules are complex (item GC0006) or difficult to understand (item GC0007), or if the goals of the game are unclear (item GC0008) or too numerous (item GC0009), the game may also appear complex. Finally, we include an item that describes the overall perception of game complexity (item GC0010).

Table 6.1 Initial items of the dimensions of game complexity, game completion difficulty, and game-playing difficulty.

Item NO.	Dimensions	Key Elements	Initial items
GC0001	Game	Elements	I think the number of game elements is large.
GC0002	Complexity	Elements	I think the types of game elements are too many.
GC0003		Elements	I think relationships among game elements are complex.
GC0004		Information	I think the information provided by this game is too much.
GC0005		Information	I think the information on this game is very vague.
GC0006		Rules	I think the rules of this game are complex.
GC0007		Rules	I think the rules of this game are hard to understand.
GC0008		Goals	I think the goal of this game is unclear.
GC0009		Goals	I think this game contains too many goals.
GC0010		Complexity	This game is very complex.
GCD0001	Game	Challenge	This is a very challenging game.
GCD0002	Completion	Understanding	This game is very difficult to understand.
GCD0003	Difficulty	Master	This game is very difficult to master.
GCD0004		Completion	This game is very difficult to complete.
GCD0005		Time	Completing this game in a demanding time is impossible.
GCD0006		Time	Completing this game needs to take too much time.
GCD0007		Success	This game looks impossible to win.
GCD0008		Success	The goal of this game is unachievable.
GCD0009		Demands	This game is highly mentally demanding.
GCD0010		Demands	This game is highly physically demanding.
GPD0001	Game-playing	Perception	I had to observe very carefully when playing this game.
GPD0002	Difficulty	Perception	I had to identify different things carefully in this game playing.
GPD0003		Memory	l had to memorize a lot of different things when playing this game.
GPD0004		Memory	I had to manage a lot of things at the same time when playing this game.
GPD0005		Thinking	l had to measure each decision carefully I made in this game.
GPD0006		Thinking	l had to think carefully about how to win this game.
GPD0007		Thinking	Thinking fast was an important part of playing this game.
GPD0008		Thinking	Thinking about time was an important part of playing this game.
GPD0009		Action	I had to act quickly when playing this game.
GPD0010		Action	Playing this game demanded precision in my actions.

For the dimension of game completion difficulty (GCD for short), we created the

Table 6.2 Initial items of the dimensions of player competence, player pressure, and player effort.

Item NO.	Dimensions	Key Elements	Initial items
PC0001	Player	Competent	I feel competent in this game.
PC0002	Competence	Competent	I have no confidence in this game.
PC0003		Competent	I feel very capable and effective in this game.
PC0004		Skill	My game ability is well-matched with this game's challenges.
PC0005		Skill	I think I am pretty good at this game.
PC0006		Skill	I am pretty skilled in this game.
PC0007		Performance	I am satisfied with my performance at this game.
PC0008		Performance	I think I was very successful in accomplishing this game.
PC0009		Comparison	I think I did pretty well in this game, compared to other players.
PC0010		Comparison	I am better than average in this game.
PP0001	Player	Nervous	I felt very tense while playing this game.
PP0002	Pressure	Nervous	The actions demanded in this game made me nervous.
PP0003		Exhausted	I feel very exhausted after playing this game.
PP0004		Exhausted	This game made me feel too fatigued to continue.
PP0005		Anxious	This game made me anxious about the time.
PP0006		Anxious	The possible loss of this game made me anxious.
PP0007		Pressure	The stress of this game was beyond my scope.
PP0008		Pressure	I felt very pressured while playing this game.
PP0009		Frustrated	I felt very frustrated while playing this game.
PP0010		Frustrated	Playing this game made me very discouraged.
PE0001	Player	Attention	I was very focused on playing this game.
PE0002	Effort	Attention	This game kept me on my toes.
PE0003		Effort	Playing this game required me to put great effort.
PE0004		Effort	I put much effort into this game.
PE0005		Energy	I invested much energy into this game.
PE0006		Energy	Playing this game required me to spend a lot of energy.
PE0007		Attempt	I tried very hard to make my actions correct in this game.
PE0008		Attempt	I tried very hard on this game.
PE0009		Attempt	To win this game, I performed my best.
PE0010		Attempt	I tried to give my best performance in this game.

items that describe players' general feelings of difficulty about the game's challenges (item GCD0001), understanding (item GCD0002), mastering (item GCD0003), and completion (item GCD0004). We also created time-related items: GCD0004 describes the game as highly time-demanding, and GCD0005 describes a game that requires the player to spend much time to complete. GCD0006 and GCD0007 are about the difficulty of winning or achieving the game's preset goal. We revised the items from the NASA

TaskLoad scale (Hart, 2006) to create the items GCD0009 and GCD0010 that describe the mental and physical demands of the game.

For the dimension of game-playing difficulty (GPD for short), most items were adapted from the CORGIS by Denisova et al. (2020). We further determined four key difficulty-related elements of game-playing: perception, memory, thinking, and action, based on the human cognitive and behavior process (Neisser, 2014). Regarding perception, the difficulty is in observing (item GPD0001) and identifying things (item GPD0002) in the game. Regarding memory, players may face challenges when they are required to remember many things (item GPD0003). Managing things in the game simultaneously relies on the ability of working memory, and it is also difficult to do (item GPD0004). In addition, decision-making (item GPD0005), thinking carefully (item GPD0006), thinking speed (item GPD107), and time-thinking (item GPD0008) in the game are all related to the difficulty of thinking. Regarding action, it is challenging for players to act quickly (item GPD0009) and precisely (item GPD0010) in the game.

For the dimension of player competence (PC for short), items were partially adapted from the PENS scale (Rigby and Ryan, 2007), the IMI scale (Ryan et al., 2006), and by referring to the self-efficacy theory (Bandura and Wessels, 1994). Players have competence feelings in the game (items PC0001, PC0002, and PC0003), based on assessing: (1) their skills in this game (PC0004, PC0005, and PC0006), (2) their performance (items PC0007 and PC0008), and (3) their performance compared to other players (items PC0009 and PC0010).

For the dimensions of player pressure (PP for short) and player effort (PE for short), items were partially adapted from the NASA TaskLoad scale (Hart, 2006), the IMI (Ryan et al., 2006), and the CORGIS (Denisova et al., 2020). Referring to the emotion category (Lewis et al., 2010), we classified the difficulty-related negative emotions into nervousness (items PP0001 and PP0002), exhaustion (items PP0003 and PP0004), anxiousness (items PP0005 and PP0006), pressure (items PP0007 and PP0008), and frustration (items PP0009 and PP0010) to create related items. The items in the player effort dimension were created or adapted by considering four aspects: when playing, whether the player (1) pays high attention (items PE0001 and PE0002), (2) invests much effort (items PE0003 and PE0004), (3) puts much energy (items PE0005 and PE0006), and (4) tries their best (items PE0007, PE0008, PE0009, and PE0010).

6.2 Stage ii: Scale Development

We conducted an investigation to verify the construct validity and exclude the ambiguous items. We first developed the 60 initial items in the Chinese and Japanese language versions based on the English version. The three versions were developed to improve our SGD measure's universal applicability; professional workers handled the translation between the three languages. Then, we recruited experienced game players for each version to conduct the card-sorting investigation (Moore and Benbasat, 1991). The card-sorting method asks different judges to sort items into preset categories, which can be used to determine the ambiguous items by the level of agreement.

The participants were asked to classify the 60 initial items into six dimensions: (a) game complexity, (b) game completion difficulty, (c) game-playing difficulty, (d) player competence, (e) player pressure, and (f) player effort. The items were presented in random order. Each item could be assigned to only one of the six dimension. An introduction to SGD and definitions of the six dimensions were provided before the formal investigation. A detailed example was also provided to illustrate the sorting procedure (see Appendix A.1).

We calculated the Fleiss's Kappa (Fleiss et al., 1979) of each dimension by SPSSAU*[1], which is the online access version of SPSS. Fleiss's Kappa can be used to replace Cohen's Kappa for more than two judges. We also calculated the hit ratios of the 60 initial items; this is based on how many participants select the dimension that matches our preset dimension (Moore and Benbasat, 1991). Table 6.3 shows the Kappa values of the six dimensions and hit ratios of the 60 items.

We carefully reviewed the classification results of the items that had low hit rates (lower than 0.9) and removed those items that confused participants. For the dimension of game complexity, we deleted the items GC0001, GC0005, GC0007, GC0008, and GC0010. For the dimension of game completion difficulty, we removed the GCD0001, GCD0002, GCD0003, GCD0005, GCD0009, and GCD0010. Interestingly, all the participants classified GCD0002 into the game complexity dimension, so we moved this item into that dimension. Items GPD0003, GPD0004, GPD0005, GPD0006, and GPD0008 in the game-playing difficulty dimension were also removed. For the dimension of player competence, we deleted items PC02, PC04, PC05, and PC08. Items PP0002, PP0003, PP0004, and PP0009 in the player pressure dimension and items PE0001, PE0002,

*[1] https://spssau.com/

Table 6.3 Fleiss's Kappa of the six dimensions and hit ratios of 60 initial items.

Dimensions	Fleiss's Kappa	Items	Hit ratios	Items	Hit ratios
Game Complexity	0.654	GC0001	0.8	GC0006	1
		GC0002	1	GC0007	0.8
		GC0003	1	GC0008	0.87
		GC0004	1	GC0009	0.93
		GC0005	0.73	GC0010	0.87
Game Completion	0.454	GCD0001	0.27	GCD0006	0.8
		GCD0002	0	GCD0007	0.87
		GCD0003	0.13	GCD0008	0.87
		GCD0004	0.87	GCD0009	0.07
		GCD0005	0.67	GCD0010	0
Game-playing	0.398	GPD0001	1	GPD0006	0.67
		GPD0002	0.93	GPD0007	0.8
		GPD0003	0.67	GPD0008	0.73
		GPD0004	0.67	GPD0009	0.87
		GPD0005	0.53	GPD0010	1
Player Competence	0.703	PC0001	1	PC0006	1
		PC0002	0.4	PC0007	1
		PC0003	1	PC0008	0.73
		PC0004	0.93	PC0009	1
		PC0005	0.93	PC0010	1
Player Pressure	0.580	PP0001	1	PP0006	1
		PP0002	0.87	PP0007	0.93
		PP0003	0.8	PP0008	1
		PP0004	0.87	PP0009	0.87
		PP0005	0.93	PP0010	0.93
Player Effort	0.626	PE0001	0.67	PE0006	0.6
		PE0002	0.13	PE0007	0.6
		PE0003	0.87	PE0008	1
		PE0004	1	PE0009	0.87
		PE0005	1	PE0010	0.87

PE0006, and PE0007 in the player effort dimension were also removed for the low level of agreement between judges. We recalculated the Fleiss's Kappa of the six dimensions and listed the first version of our SGD scale in Table 6.4. After removing these items, all the dimensions have Kappa values larger than 0.8. The game-playing difficulty dimension, with the value of 0.787, was considered close enough to 0.8 and was acceptable. This means that all the dimensions achieve high consistency among the judges. Therefore, we believe that the instrument reached a reliable quality for testing in the next stage.

Table 6.4 First version of our SGD scale and Fleiss's Kappa of the six dimensions

Old Item NO.	New Item NO.	Dimensions	Fleiss's Kappa	Items
GC0002	GC1001	Game Complexity	0.946	I think the types of game elements are too many.
GC0003	GC1002			I think relationships among game elements are complex.
GC0004	GC1003			I think the information provided by this game is too much.
GC0006	GC1004			I think the rules of this game are complex.
GC0009	GC1005			I think this game contains too many goals.
GCD0002	GC1006			This game is very difficult to understand.
GCD0004	GCD1001	Game Completion Difficulty	0.822	This game is very difficult to complete.
GCD0006	GCD1002			Completing this game needs to take too much time.
GCD0007	GCD1003			This game looks impossible to win.
GCD0008	GCD1004			The goal of this game is unachievable.
GPD0001	GPD1001	Game-playing Difficulty	0.787	I had to observe very carefully when playing this game.
GPD0002	GPD1002			I had to identify different things carefully in this game playing.
GPD0007	GPD1003			Thinking fast was an important part of playing this game.
GPD0009	GPD1004			I had to act quickly when playing this game.
GPD0010	GPD1005			Playing this game demanded precision in my actions.
PC0001	PC1001	Player Competence	0.904	I feel competent in this game.
PC0003	PC1002			I feel very capable and effective in this game.
PC0006	PC1003			I am pretty skilled in this game.
PC0007	PC1004			I am satisfied with my performance at this game.
PC0009	PC1005			I think I did pretty well in this game, compared to other players.
PC0010	PC1006			I am better than average in this game.
PP0001	PP1001	Player Pressure	0.906	I felt very tense while playing this game.
PP0005	PP1002			This game made me anxious about the time.
PP0006	PP1003			The possible loss of this game made me anxious.
PP0007	PP1004			The stress of this game was beyond my scope.
PP0008	PP1005			I felt very pressured while playing this game.
PP0010	PP1006			Playing this game made me very discouraged.
PE0003	PE1001	Player Effort	0.882	Playing this game required me to put great effort.
PE0004	PE1002			I put much effort into this game.
PE0005	PE1003			I invested much energy into this game.
PE0008	PE1004			I tried very hard on this game.
PE0009	PE1005			To win this game, I performed my best.
PE0010	PE1006			I tried to give my best performance in this game.

6.3 Stage iii: Scale Testing

This stage aimed to test the reliability and validity of our scale through an online survey. Because our SGD scale should be measured based on players' experience with a specific game, the testing in this stage could only focus on the players' SGD in one game. Therefore, to test our developed scale effectively, we selected the testing game by these criteria: (1) this game is popular with many players, (2) to avoid extreme test results, this game should not be too difficult (or too complex) or too easy, and (3) this game demands focus on the action, thinking and playing time. Finally, we chose the *Plants vs. Zombies* game (Games, 2009), which perfectly satisfies our criteria. *Plants vs. Zombies* is a casual game but still presents enough challenge. The gameplay of this game requires players to collect sunlight resources in order to set plants to fight against zombies, and it involves thinking, action, and time pressure. This game is very popular and has sold over 30 million copies.

To ensure our participants have played the game, we set five questions (see Appendix A.1.2) before the formal survey, which they must correctly answer to prove they are familiar with the game. Only those participants who correctly answered all five questions could attend the formal survey. The formal survey contained two parts: the first part collected the basic information about participants, and the second part was a 35-item Likert questionnaire (with the first version of the 33-item SGD scale and two polygraph questions). The questions in this questionnaire were shown in random order, and responses ranged from 1 (strongly disagree) to 7 (strongly agree). The survey had three language versions, namely, English, Chinese, and Japanese, and professional workers handled the translation between the various languages.

All the participants were recruited online through Amazon Mechanical Turk[*2], CloudWorks[*3], and Wenjuanxing[*4]. 565 responses were collected, but only 342 participants correctly answered the five preset questions and completed the formal survey. Sixteen participants were removed for answering the polygraph questions incorrectly or not providing adequate demographic information. Finally, the data from 326 participants (175 males and 151 females) were used in the analysis. Our participants were aged from 18 to 60 (M = 26.23, SD = 7.29), their nationality were the United States of America (134 of 326), China (111 of 326), and Japan (81 of 326); they use English,

[*2] https://www.mturk.com/
[*3] https://crowdworks.jp/
[*4] https://www.wjx.cn/

Chinese, or Japanese as their first language.

The game experience of our participants ranged from 1 to 45 years ($M = 11.25$, $SD = 7.94$). For playing frequency and game skills, most of our participants had played games within a few days (203 of 326) or a few days ago (59 of 326) and rated their game skills as ordinary (114 of 326) or skillful (99 of 326). All the participants have played the genre of casual games before, and the other game genres they most preferred were Role-playing games (RPG) (169 of 326) and action games (186 of 326). In addition, most of our participants had a game time of *Plants vs. Zombies* of 1-5 hours (137 of 326) or within 1 hour (134 of 326), and rated their game skills in this game as ordinary (127 of 326) or skillful (85 of 326).

6.3.1 Reliability Testing

We employed Cronbach's α Tavakol and Dennick (2011) to measure the reliability, and the analysis was conducted by SPSSAU. The Cronbach's α was calculated for the six dimensions using all the collected data. The results show the six dimensions all have Cronbach's α larger than 0.8, see Table 6.5. Therefore, the internal consistency of each dimension is acceptable.

Table 6.5 Internal consistency (Cronbach's α) of our SGD scale.

Dimensions	Cronbach's α
Game Complexity	0.912
Game Completion Difficulty	0.852
Game-playing Difficulty	0.830
Player Competence	0.896
Player Pressure	0.902
Player Effort	0.806

6.3.2 Validity Testing

A confirmatory factor analysis (CFA) was done on the survey data using the 6-factor structure based on the six dimensions for validity testing. CFA is commonly used to analyze the efficacy of measuring models and test the scale's validity (Harrington, 2009). The stand factor loading by CFA is shown in Table 6.6. Based on the literature (Harrington, 2009), the estimated value of factor loading should be larger than 0.6, or it means the correlation between this item and the factor is not strong enough.

Table 6.6 CFA of our first version SGD scale

NO.	Dimensions (factors)	Items	Std. factor loading
GC1001	Game	I think the types of game elements are too many.	0.802
GC1002	Complexity	I think relationships among game elements are complex.	0.797
GC1003		I think the information provided by this game is too much.	0.818
GC1004		I think the rules of this game are complex.	0.791
GC1005		I think this game contains too many goals.	0.777
GC1006		This game is very difficult to understand.	0.792
GCD1001	Game	This game is very difficult to complete.	0.796
GCD1002	Completion	Completing this game needs to take too much time.	0.599
GCD1003	Difficulty	This game looks impossible to win.	0.826
GCD1004		The goal of this game is unachievable.	0.879
GPD1001	Game-playing	I had to observe very carefully when playing this game.	0.775
GPD1002	Difficulty	I had to identify different things carefully in this game playing.	0.733
GPD1003		Thinking fast was an important part of playing this game.	0.611
GPD1004		I had to act quickly when playing this game.	0.632
GPD1005		Playing this game demanded precision in my actions.	0.748
PC1001	Player	I feel competent in this game.	0.819
PC1002	Competence	I feel very capable and effective in this game.	0.737
PC1003		I am pretty skilled in this game.	0.798
PC1004		I am satisfied with my performance at this game.	0.700
PC1005		I think I did pretty well in this game, compared to other players.	0.829
PC1006		I am better than average in this game.	0.742
PP1001	Player	I felt very tense while playing this game.	0.831
PP1002	Pressure	This game made me anxious about the time.	0.668
PP1003		The possible loss of this game made me anxious.	0.719
PP1004		The stress of this game was beyond my scope.	0.839
PP1005		I felt very pressured while playing this game.	0.822
PP1006		Playing this game made me very discouraged.	0.789
PE1001	Player Effort	Playing this game required me to put great effort.	0.753
PE1002		I put much effort into this game.	0.820
PE1003		I invested much energy into this game.	0.768
PE1004		I tried very hard on this game.	0.446
PE1005		To win this game, I performed my best.	0.552
PE1006		I tried to give my best performance in this game.	0.422

We carefully checked all the items with a value less than 0.8 and removed items GC1001, GC1005, GCD1002, PP1002, PP1003, PE1004, PE1005, and PE1006 to improve the quality of the measurement model. Items GPD1003 (factor loading = 0.611) and GPD1004 (factor loading = 0.623) were retained even with low loading values because this dimension measures different aspects of game playing, and we thought that values were still acceptable. After removing these items, a second CFA was conducted; see Tables 6.7 and 6.8. According to the literature Hair (2009), the chi-squared statistic and other model fit indices (χ^2 is significant and $\chi^2/df < 3$; CFI > 0.9, TLI > 0.9, RNI > 0.9, SRMR < 0.1, RMSEA < 0.08) should be used for confirming whether the model fit is acceptable. As the result shown, the chi-squared statistic ($\chi^2 = 643.34$, df $= 260$, p < 0.001; $\chi^2/df = 2.474$) and the model fit indices (CFI = 0.930, TLI = 0.919, RNI = 0.921, SRMR = 0.065, RMSEA = 0.067) demonstrate that the CFA model is acceptable. We found that all the items had a value of factor loading that was larger than 0.6 and showed good correlation relationships between the items and dimensions.

Table 6.7 Second CFA ($N = 326$) after removing some items of our first version SGDS (Total number of items = 25). Results of the standard factor loading

NO.	Dimensions	Std. factor loading	NO.	Dimensions	Std. factor loading
GC1002	Game	0.791	PC1001	Player	0.818
GC1003	Complexity	0.796	PC1002	Competence	0.736
GC1004		0.796	PC1003		0.798
GC1006		0.828	PC1004		0.698
GCD1001	Game	0.791	PC1005		0.830
GCD1003	Completion	0.838	PC1006		0.743
GCD1004	Difficulty	0.889	PP1001	Player	0.836
GPD1001	Game-playing	0.775	PP1004	Pressure	0.865
GPD1002	Difficulty	0.732	PP1005		0.805
GPD1003		0.611	PP1006		0.782
GPD1004		0.629	PE1001	Player Effort	0.775
GPD1005		0.749	PE1002		0.819
			PE1003		0.790

Table 6.8 Second CFA ($N = 326$) after removing some items of our first version SGDS (Total number of items = 25). Results of the model fit index. Note: ***Denotes significant at the 0.001 levels (2-tailed).

χ^2	df	p	CFI (> 0.9)	TLI (> 0.9)	RNI (> 0.9)	SRMR (< 0.1)	RMSEA (< 0.08)
643.34***	260	< 0.001	0.930	0.919	0.921	0.065	0.067

Based on the result of the second CFA, we further analyzed the validity of this scale by calculating the Composite Reliability (CR) and Average Variance Extracted(AVE). According to the literature Hair (2009), the convergent validity of the measurement will be good when AVE values are larger than 0.5, and CR values are larger than 0.7. The result in Table 6.9 shows that all the values of AVE and CR fit the standard except the AVE of game-playing difficulty, which had a value of 0.493. The reason might be that we retained the items GPD03 and GPD04. However, the value was very close to 0.5, which we believed was still acceptable.

Table 6.9 Composite Reliability (CR) and Average Variance Extracted(AVE) the six dimensions.

Dimensions	Average Variance Extracted (AVE)	Composite Reliability (CR)
Game Complexity	0.645	0.879
Game Completion Difficulty	0.706	0.878
Game-playing Difficulty	0.493	0.828
Player Competence	0.596	0.898
Player Pressure	0.676	0.893
Player Effort	0.632	0.838

We also conducted a Pearson correlation analysis to calculate the coefficient r between each pair of dimensions, see Table 6.10. According to the literature Hair (2009), if the square root of AVE is larger than the dimension's r with other dimensions, the discriminant validity of that dimension is good. The result showed that dimensions of GC (0.803 > 0.802), GCD (0.840 > 0.796), GPD (0.702 > 0.657), PC (0.772 > 0.366), PP (0.822 > 0.802), and PE (0.795 > 0.672) had their AVE's square root values larger than the related r values with other dimensions. Therefore, the discriminant validity of SGDS with these six dimensions could also be rated as good.

6.3.3 Invariance Testing

Because we conducted the survey in three language versions, it is necessary to test the invariance between different versions. Based on research by Fischer and Karl Fischer and Karl (2019), we conducted a multi-group CFA in R and calculated the values of Tucker's ϕ and the correlation coefficient between the three language groups. The two values that are closer to 1 represent these groups are more similar to each other. The

Table 6.10 Pearson's correlation r between the dimension and square root of AVE.

	GC	GCD	GPD	PC	PP	PE	Square root of AVE
GC	1						0.803
GCD	0.796	1					0.840
GPD	0.508	0.405	1				0.702
PC	0.174	-0.006	0.360	1			0.772
PP	0.802	0.785	0.481	0.089	1		0.822
PE	0.672	0.559	0.657	0.368	0.657	1	0.795

result showed a minor but acceptable invariance between these three language groups, see Table 6.11.

Table 6.11 Invariance testing

	Tucker's $\phi 1$	Tucker's $\phi 2$	correlation 1	correlation 2
English	0.96	0.92	0.97	0.96
Chinese	0.96	0.95	0.96	0.98
Japanese	0.99	0.97	0.91	0.9

6.4 Discussion

In this chapter, we developed a new scale, SGDS, to measure the subjective game difficulty (SGD) of players. This scale has been verified for its reliability and validity. We compared this scale with other SGD measurements and provided an introduction to the usage of this instrument. Overall, we provide a useful scale for measuring SGD in video games.

6.4.1 SGDS Compared to the Other SGD Measuring Methods

As mentioned in Chapter 2, there are currently three main SGD measuring methods: simple self-report, structured self-report, and physiological measurement. Our developed scale is a new instrument that belongs to the structured self-report method. We compared our scale with current methods and instruments to discuss how it improves or supports current SGD measurement.

Compared to the simple self-report method, our instrument supports a more detailed and comprehensive measurement of SGD. More specifically, the simple self-report method asks the player to rate whether the game is difficult, usually with scores. However, as we summarized, the difficulty of the game has different aspects, but this method merely measures players' overall perception of the game's difficulty. Instead, the developed SGDS uses six dimensions, namely, game complexity, game completion difficulty, game-playing difficulty, player competence, player pressure, and player effort to measure the SGD of players. Through structured questions in these dimensions, this instrument can help players evaluate the different aspects of game difficulty in more detail and comprehensively.

As a new instrument of the structured self-report method, we compared it with the existing scale, CORGIS (Denisova et al., 2020). There are three main differences between SGDS and CORGIS. 1) Core concept: our scale relies on the game difficulty concept, which is theoretically more precise than the challenge concept for measuring. 2) Measuring objects: CORGIS measures the challenge experience of the game because the scale's developers defined the perceived challenge as an experience. In comparison, we defined SGD as the players' structured perception, i.e., the personal processing of the experience. 3) Measuring dimensions: CORGIS classifies challenges into four types of measure; we suggest that such measuring is more about the difficulty in the interaction part, which is one of our dimensions. In comparison, our scale also considers the players' perceptions about the difficulty related to the games' attributes, completion, and players' playing state. In short, our dimensions are more reasonable and comprehensive in the structure. In summary, SGDS improves the CORGIS in the three aspects as a new structured self-report method.

Physiological measurements have become popular to measure players' feelings in recent years (Caroux et al., 2015). This type of method is more objective and operational than subjective measuring methods. This method provides information about players' physiological state, e.g., emotion, pressure, etc., rather than their thoughts and views. Therefore, although the state of effort and pressure can be measured more directly, other dimensions in our scale still cannot. More importantly, the state of high effort or pressure does not equal a high player SGD perception. To better interpret the physiological state, a precise measurement of the players' SGD perception is still necessary. Therefore, this method cannot replace the self-report method. SGDS, as a better instrument for precisely measuring SGD, can support the future development of the physiological measurement method.

6.4.2 How to Use This Scale

We provide the final version of the SGD scale in Appendix B (English version in B.1, Chinese version in B.2, and Japanese version in B.3). This scale is a 25-item Likert questionnaire, and we suggest using the range 1 (strongly disagree) to 7 (strongly agree) for measurement. The data in the five subscales, namely game complexity, game completion difficulty, game-playing difficulty, player pressure, and player effort, is positively related to the level of SGD, while the subscale of player competence is negatively related. Therefore, the higher the average score of each subscale (the lower the average score of the player competence subscale), the more difficult the player perceives that SGD dimension to be.

To use this scale in the SGD measurement, it is necessary to first determine the measuring goals and timings. Although it is common to measure SGD after play, our scale also supports the measurement before and during play. For example, when the goal of measuring is to identify the players' rough impression of the game's difficulty (without playing) or to assess difficulty perception changes during play, the after-test is inapplicable.

In the stage before playing, the players may just have a rough idea after observing the game. Due to the lack of play, players have no idea about the four dimensions' SGD in the interaction part and player part. However, they can still assess the two dimensions of game complexity and game completion difficulty roughly based on their first impression of the game. One of the values of SGD measurement in this stage is identifying players' first impressions of the game's difficulty. This impression may affect their initial willingness to play this game and their subsequent play experience (Huang et al., 2024). Players may choose not to begin a game that looks too complex and they have no confidence to complete.

Player SGD may change over time during gameplay, making real-time measurement necessary in some research. Measuring SGD during gameplay is also valuable for designers because they need to draw players' difficulty curves for game difficulty and to design the various levels of the game appropriately (Guo et al., 2024). Therefore, we suggest using the game-playing difficulty dimension to measure the SGD during the process of playing the game. This dimension was created to describe the difficulty that players face while playing. In addition, the whole SGD scale with six dimensions can also be applied in this stage. However, interrupting issues should be considered because answering so many items may affect the fluency of the player's experience. For this reason, eliminating some items based on the research requirement is also an option.

In the after-playing stage, the whole SGD scale is applicable for measuring, but researchers can also choose some of the dimensions based on their requirements. For example, if the researchers care more about how the players make an effort to overcome the difficulty during the playing, game-playing difficulty and player effort may be more important in the measurement. In addition, the SGD scale can be used for game difficulty comparison between games or player groups. This scale can also be combined with other scales (e.g., player experience scale; Ryan et al., 2006) to study the interactive relationships between SGD and other factors.

In summary, SGDS can be used in different forms and situations based on the needs of researchers and designers.

Chapter 7

Rethinking Dynamic Difficulty Adjustment

This chapter rethinks the theoretical fundamentals of DDA through four crucial questions, in which novel insights into DDA's plight, definition, scope, value, and design are provided. Based on the definitions and interaction model proposed in Chapter 3, we further redefine DDA from the interactive perspective and discussed its scope and value. Finally, we present a goal-based DDA framework and proposed a 6-step DDA design process as a practical design approach.

7.1 Introduction

Dynamic Difficulty Adjustment (DDA), as a game difficulty mechanism, has emerged with the development of computer science and psychology theory, bringing high expectations for game design. DDA has been proposed for nearly 20 years with the goal of creating a better game experience through adaptive challenge balancing (Hunicke, 2005). More specifically, this goal in most DDA studies is creating the game experience of the Flow State (Flow; Zohaib, 2018). Flow has become a foundational concept in DDA as it occurs in various DDA definitions (Zohaib, 2018). Until recently, most DDA studies still cite Flow (e.g., Seyderhelm et al., 2019; Pfau et al., 2020; Yildirim et al., 2021; Knorr and Vaz de Carvalho, 2021; Moon et al., 2022; Sepulveda et al., 2019) or the Flow-related balanced challenge description of "not too easy, nor too hard" (e.g., Moniaga et al., 2018; Purnama et al., 2018; Shakhova and Zagarskikh, 2019).

However, despite the ongoing development of various DDA mechanisms, DDA has not yet fully achieved its promised success. There are few commercial successes of DDA (Adams, 2014; Schell, 2019) and its effectiveness in enhancing player experience is questionable (Smeddinck et al., 2016; Ang and Mitchell, 2017; Wang et al., 2016;

Spiel et al., 2019; Hind and Harvey, 2022). Interestingly, studies on serious games with DDA indicate that it is more effective in achieving serious goals but shows no difference in player experience compared to other non-DDA approaches (Sampayo-Vargas et al., 2013; Kitakoshi et al., 2020; Hooshyar et al., 2021; Jemmali et al., 2022; Valencia et al., 2018).

We argue that the lack of reflection in DDA design research is the main reason for current issues. There is little systematic research on the root, concept, significance, and design framework of DDA; instead, brief descriptions of DDA are scattered in various design or application studies (He et al., 2010; Hawkins et al., 2012; Sutoyo et al., 2015; Lach, 2017; Demediuk et al., 2019). This lack of reflection is concerning, as designers and researchers keep searching for better methods or techniques without considering the theoretical fundamentals of DDA. To improve DDA design, we propose four essential questions as threads for rethinking DDA:

Q1: What limits the effectiveness of current DDA?

Q2: What is the definition and scope of DDA?

Q3: Why is DDA valuable for video game design?

Q4: How to design DDA for video games?

To address these four research questions, we conducted a literature review through two rounds of paper-searching, in which the first round considered the DDA concept and the second round considered the DDA's theoretical foundation. In the first round, we searched for the following terms: game AND (difficulty OR challenge) AND ("dynamic difficulty adjustment" OR "adaptive" OR "DDA") in the ScienceDirect, ACM Digital Library, and IEEE Xplore Advanced Search, with the search area limited to Titles, Abstracts, and Keywords of the papers published in the past 20 years (from May 2003 to May 2023). A total of 546 papers were searched. After excluding (1) non-English literature, (2) non-research articles, (3) non-video game papers, and (4) other papers less relevant to the DDA topic, 231 papers were finally retained. The terms: game AND (difficulty OR challenge) AND ("definition" OR "Flow") were used as our keywords to search for DDA's theoretical background in the second round. To carefully contain DDA's theoretical foundation papers related to the first round of searching, we investigated the references of the papers found in the first round to supplement topic-relevant papers (including papers from other databases) before our second round of formal paper-searching. In the second round of searching, we applied the same criteria as the first round to search and exclude papers and further excluded those papers less relevant to the game difficulty, game challenge, or Flow theory. A total of 740 were found in the second round combining these two sources (references and databases) and 256 papers were retained. After these two rounds of searching and primary screening, we further screened papers to exclude (1) short papers and (2) papers with unclear or limited contributions. 134 papers from our search were included and cited in this work.

In summary, we adopted a loose but effective literature research method, aiming to understand the theoretical foundation, definition, value, and design practices of DDA comprehensively. In more detail, Section 7.2 addresses Q1 by reviewing relevant work on DDA and Flow theory. We critically examine the dependence of DDA on Flow and explain why it has led to current issues. Section 7.3 answers Q2 and Q3 by defining and scoping DDA and highlighting DDA's value based on the interactive perspective of game difficulty. Section 7.4 first explores the relationship between game difficulty and design goals and then confirms the goal-based evaluation criteria of DDA by analyzing design theories. A goal-based DDA design framework and a 6-step design process are proposed as the final answer to Q4.

7.2 What Limits DDA: The Relationship of Flow Theory to DDA

This section aims to reveal the intrinsic relationship between DDA and Flow theory. Flow serves as the foundation for current DDA, but it also restricts DDA's definition and design goal. Unfortunately, there have been few studies on this relationship. We combine the related work of DDA and Flow theory to clarify this relationship and the limitations incurred.

7.2.1 The Fundamental of DDA: Flow Theory

DDA was born after 2000 with the bloom of video games. The first well-known literature on DDA is from Hunicke (2005) and his Helmet system. This system changes the in-game parameters (e.g., weapon damage) to adapt to the player's health points. Before that, Flow theory had already become popular and was applied in many areas such as sports, HCI, and games (Csikszentmihalyi and Csikzentmihaly, 1990; Jackson and Csikszentmihalyi, 1999; Zhang and Dillon, 2003; Juul, 2004). Flow theory mainly described the nine dimensions of Flow, which contain its reaching conditions (e.g., challenges appropriate to one's skill) and typical subjective experience (e.g., absorption, immersion, and autotelic experience) (Nakamura and Csikszentmihalyi, 2014). Flow theory seems to provide proper guidance on designing challenges to realize the ideal experience, which is what game designers are looking for. Chen (2007) explained how to apply Flow theory in designing game difficulty by creating the "challenge-skill balance", also known as the Flow Channel (Hunicke, 2005; Schell, 2019; Corrêa et al., 2022). This model soon became widespread and has been utilized in many game studies (Fong et al., 2015).

Currently, Flow has been regarded as the fundamental of DDA's definition and components. According to Soderman (2021), citing Flow theory or Flow for DDA research has nearly become necessary. Zohaib (2018) defined DDA in his review paper as *"a method of automatically modifying a game's features, behaviors, and scenarios in real-time, depending on the player's skill, so that the player, when the game is very simple, does not feel bored or frustrated, when it is very difficult"*. This definition regards the challenge-skill balance as the primary goal of DDA and similar definitions are also commonly found in other DDA studies (e.g., Hunicke, 2005; He et al., 2010; Baldwin et al., 2013; Alexander et al., 2013; Karpinskyj et al., 2014; Denisova and Cairns, 2015; Silva et al., 2015; Lach, 2017; Pfau et al., 2020).

Based on the literature review in Section 2.5, most researchers and designers agree that DDA aims to avoid unbalanced game difficulty with two components: (1) a player evaluation mechanism for measuring player performance and (2) a difficulty adjustment mechanism to change the level of game difficulty (Adams, 2014; Yin et al., 2015; Demediuk et al., 2017). These two components of DDA naturally correspond to skill and challenge. In comparison, difficulty adjustment mechanisms gain more attention from researchers than player evaluation mechanisms. Considering the relationship between DDA and Flow theory, we believe the reason is that the evaluation criteria of the player evaluation mechanisms are restricted to "whether the challenge is balanced", which is even simplified to a 50% success rate in some studies (e.g., Demediuk et al., 2017; Anguera et al., 2013; Sampayo-Vargas et al., 2013; Kitakoshi et al., 2020).

Flow has also been regarded as the goal for most DDA designs. DDA researchers commonly mention Flow or the challenge-skill balance in the background and then introduce how they attempt to achieve this goal through AI agents (Tan et al., 2011; Ishihara et al., 2018), algorithm (Sorenson et al., 2011; Wheat et al., 2016) or system design (Hawkins et al., 2012; Tsai, 2016). Flow State Scale is also used to verify whether this design goal has been reached (Ang and Mitchell, 2017; Kaplan et al., 2018). Flow has become "*an almost unquestioned reference*", as Fassone (2017) mentioned. In short, it can be seen from the existing literature that DDA's definition, components, and design goal are Flow-related.

7.2.2 Why this Relationship Restricts the Effectiveness of DDA Design

It can be concluded that Flow has been regarded as the fundamental of most DDA designs. Many DDA methods have been designed and developed based on the challenge-skill balance from Flow theory. However, there is a lack of critical thinking regarding the application of this theory in DDA design, resulting in the narrow perspective of current design practice.

It seems that Flow theory does not live up to its promise of enhancing player experience when designers apply it in the DDA designs. Many studies indicate that neither entertainment nor serious games create a better player experience with DDA (Smeddinck et al., 2016; Ang and Mitchell, 2017; Kitakoshi et al., 2020; Hooshyar et al., 2021). Soderman (2021) is critical, stating that, based on Flow theory, the current DDA systems are only concerned about the duration while they ignore the possible exhausting side of the game play. Flow-based DDA is also questioned by some famous game designers who suggest keeping caution on using DDA in game design (Adams, 2014; Schell, 2019). For example, Schell (2019) in *The Art of Game Design*, listed some issues with DDA and emphasized that the abuse of DDA may ruin a healthy game-playing process.

According to the literature review, the relationship between DDA and Flow restricts the effectiveness of DDA design, and the restriction mainly comes from three aspects:

(1) Flow theory itself — Some researchers argue that the ambiguity of Flow theory is a fatal problem (Swann et al., 2018; Jalife et al., 2021; Norsworthy et al., 2021; Burns and Tulip, 2017), which causes the Flow state described by this theory to be measured only by subjective scales but with no reliable objective detection methods (Norsworthy et al., 2021). The controversy also appears in the contribution of challenge-skill balance in Flow. Fong et al. (2015) suggested they have a moderate relationship through a meta-analysis, but other studies challenged this view (Løvoll and Vittersø, 2014; Cutting et al., 2022; Nacke and Lindley, 2008; Klarkowski et al., 2015; Martin and Magerko, 2020; Scheepers and Keller, 2022).

(2) The way DDA applies Flow — Many DDA designs applied challenge-skill balance by adjusting the game difficulty to a specific success rate. However, some researchers have questioned the 50% success rate and what balance really means (Allart et al., 2017; Lomas et al., 2017; Cutting et al., 2022; Masanobu et al., 2017). This rough application approach hinders DDA from ensuring a good experience for players. However, the more fundamental reason is the ambiguity of Flow theory, because it is hard to quantify the balanced challenge in DDA design.

(3) The guidance of Flow — We acknowledge the value of the Flow channel model, which is intuitive and has garnered substantial evidence for preventing extreme negative experiences (Schell, 2019). However, the complexity of the player experience makes us wary. Flow theory looks limited in guidance when facing various design goals. Some proofs from game research have shown that players prefer low game difficulty levels over a balanced challenge (Alexander et al., 2013; Lomas et al., 2017; Allart et al., 2017; Koskinen et al., 2023; Lemmens and von Münchhausen, 2023). On the other hand, Stammer et al. (2015) found that experienced players were opposed to challenge removal. It cannot be ignored that casual play supports pleasure and relaxation (Alexander et al., 2013; Tyack and Mekler, 2021), while failure makes in-game growth meaningful (Juul, 2009; Laffan et al., 2016). In addition, serious goals are more varied and require relative theoretical support rather than any possibly naive application of Flow theory.

We found that positive changes have been made in DDA design with the development of various theories and techniques. Affective games, which come from the framework of Affective Computing (Picard, 2000), provide a different perspective for applying emotional design in games (Gilleade et al., 2005; Hudlicka, 2009). Therefore, some studies have applied emotion detection methods in designing DDA for evaluating player responses (Liu et al., 2009; Kivikangas et al., 2011; Bontchev, 2016; Khalili-Mahani et al., 2020; Darzi et al., 2021). However, the innovation and introduction of such methods does not solve the fundamental problems, more theoretical developments are still required to support DDA and its design.

7.3 What is DDA & Why DDA: Definition, Scope, and Value

Current DDA is heavily influenced by Flow theory and thus it lacks independent definition. Fortunately, the very term DDA, "Dynamic Difficulty Adjustment", provides more detail regarding what it is. However, most works on DDA seem to focus on the "dynamic" rather than the "difficulty". This section clarifies the definition and scope of DDA by game difficulty definitions proposed in Chapter 3. The value of DDA is then discussed based on our refined DDA definition.

7.3.1 Game Difficulty and DDA

Based on our proposed model and definitions in Chapter 3, game difficulty occurs in the interaction between players (each with a specific skill level) and game tasks (each task with a particular level of complexity). Fig. 7.1 shows how game difficulty, consisting of OGD and SGD, occurs in the game interaction. In this interaction process, OGD is that "the dynamic meeting of the player's skill to the game task demand"; while SGD is that "the player's subjective evaluation of game difficulty based on their structured perceptions".

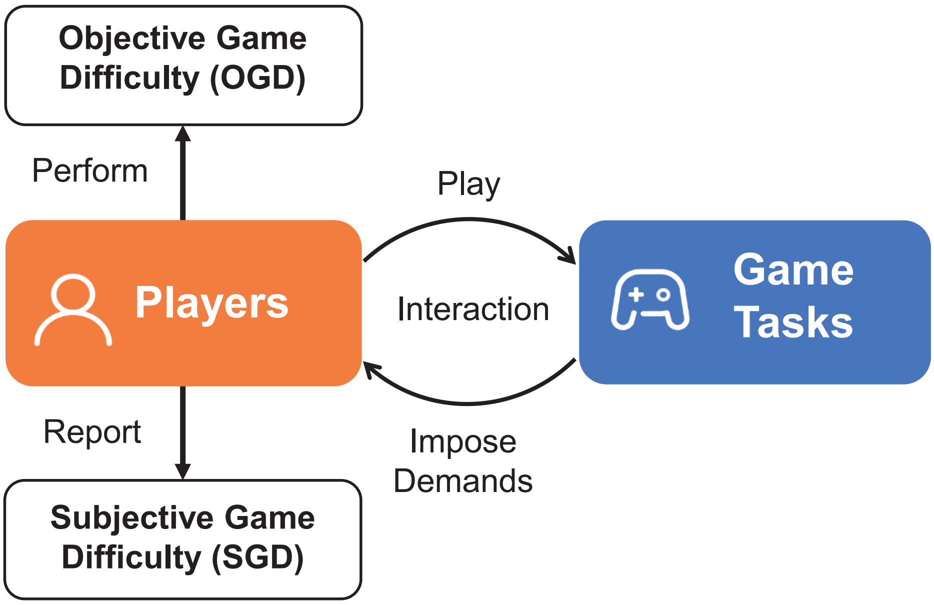

Fig. 7.1 Game difficulty can be divided into objective game difficulty (OGD) and subjective game difficulty (SGD). These two kinds of difficulties occur in the interaction between players and game tasks.

Therefore, if game difficulty refers to the interaction result at every single moment, DDA means catching these results dynamically and then adjusting the following results. We integrate DDA into the player-game interaction in Fig. 7.1 to draw Fig. 7.2. Fig. 7.2 shows that DDA consists of a player evaluation mechanism and a difficulty adjustment mechanism, and its player evaluation mechanism needs to evaluate feedback of OGD

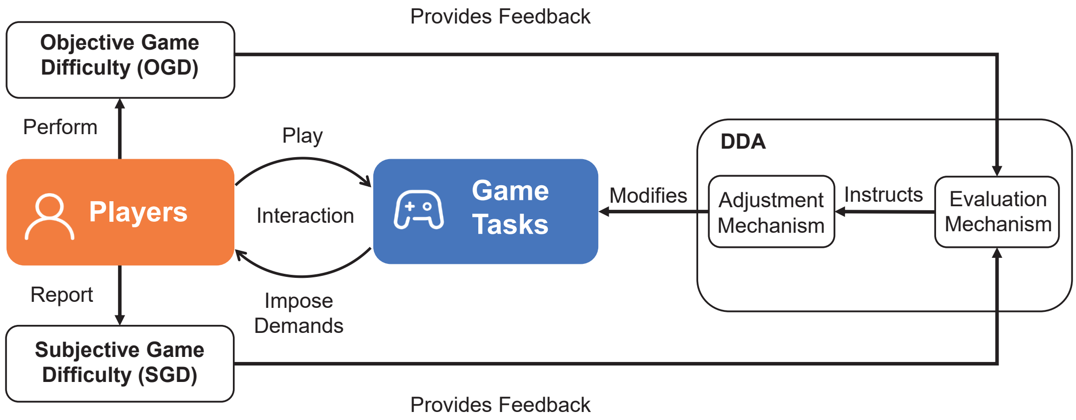

Fig. 7.2 We add the DDA component in Fig. 7.1 according to the literature review (e.g., Alexander et al., 2013; Xue et al., 2017; Frommel et al., 2018; Darzi et al., 2021).

and SGD to instruct the difficulty adjustment mechanism to modify the game tasks. Such separation of game difficulty into OGD and SGD is important for supporting DDA design, which we explain in Section 7.4. It should be noted that although it is called Dynamic Difficulty Adjustment, DDA does not adjust the difficulty directly but adjusts it by modifying the game tasks.

We thus define Dynamic Difficulty Adjustment as **"a game difficulty control mechanism that aims to control the difficulty automatically in game interaction by evaluating objective and subjective game difficulty data and modifying game tasks."** This new DDA definition preserves the core connotations of dynamic, difficulty, and adjustment. It enriches the concept of DDA from the perspective of interaction and game difficulty but detaches from any essential relationship or description of Flow.

7.3.2 The Scope of DDA

Even though we have defined DDA, the scope of DDA still requires further clarification compared to other similar mechanisms and concepts. This would provide the precise basis for applying DDA in game design to avoid misuse or confusion.

According to our definition, DDA is an automatic game difficulty control mechanism and thus it differs from Manual Difficulty Adjustment (MDA) or Static Difficulty (SD). MDA means that players can change the game tasks manually (Smeddinck et al., 2016), while SD refers to the game tasks remaining unchanged or changing in a fixed mode

(Alexander et al., 2013). In the interaction process, DDA is more flexible and sensitive to specific design goals compared to these two mechanisms. Another difficulty mechanism, the Player-Matching mechanism, is also discussed as the multiplayer DDA (Baldwin et al., 2013, 2016; Tsai, 2016). It is a mechanism that finds the most appropriate human opponents for players based on the player's previous performances. Strictly speaking, this mechanism does not work in an interaction process but before the beginning of each round of games. Therefore, it does not belong to DDA according to our definition but is another automatic difficulty adjustment mechanism. However, if we adopt a broad perspective to regard the interaction as a long-term relationship between players and games (Hornbæk and Oulasvirta, 2017), this mechanism can also be included in DDA.

Besides these difficulty mechanisms, Personalization and Adaptation are similar concepts related to DDA. According to Karpinskyj et al. (2014), personalization refers to the "*automatic customization of content and services based on a prediction of what the user wants*" and DDA can be regarded as player performance personalization (Bakkes et al., 2014; Blom et al., 2019). However, the vague use of the term "wants" prevents personalization from fully encompassing DDA. For example, in some serious games that use DDA, difficulty adjustment relies on the researcher's research requirements and serious theories rather than the player's needs (Anguera et al., 2013; Ozkul et al., 2019; Bakkes et al., 2012). According to Dörner et al. (2016), adaptation refers to the "*continuous adjustment of the game based on the actions and performance of a user and the current state of the game towards a desired state.*" This concept with a larger scope includes difficulty, game content, narrative, etc. DDA can be regarded as the difficulty adaptation (Carvalho et al., 2022) which is also commonly used in current research (Chanel et al., 2008, 2011; Yin et al., 2015; Nagle et al., 2016). However, we disagree with Dörner et al. (2016) that difficulty adaptation equals creating a balanced challenge to achieve Flow. DDA is more about a difficulty control tool and can serve many more design goals than just Flow. DDA can and should be flexible, useful for various goals.

Player characteristics, including player profiles, models, and preferences, have also been utilized in some DDA research (Monterrat et al., 2015; Charles and Black, 2004; Hocine et al., 2011; Dias and Martinho, 2011). Player characteristics are studied to better understand the players and their needs, which thus belong to user and personalization research (Paraschos and Koulouriotis, 2023; Karpinskyj et al., 2014). According to our definition, DDA is only a difficulty control mechanism for the interaction process.

Therefore, the content of player characteristics cannot be one of the essential parts of DDA. In addition, most studies applied player characteristics in DDA by establishing player profiles before game interaction (e.g., Monterrat et al., 2015), which does not match our DDA scope. The content of player characteristics can also be used in video games to support other non-difficulty adaptions, e.g., music adaption (Rossoff et al., 2010). We agree that the content of player characteristics can support a better understanding of player enjoyment and predict needs satisfaction, and combining it with DDA in game design will be more comprehensive. Therefore, to avoid confusion or imprecise scope expansion, we exclude player characteristics from the DDA definition or scope but further discuss how to apply it in supporting DDA design in Section 7.4.

In summary, we scope DDA as an adaptive game difficulty mechanism that only takes effect in game interaction. We support combining DDA with other mechanisms and components in video game design, but precise distinctions will better support such a combination.

7.3.3 The Value of DDA

Since we provided the new DDA definition and scope, it is necessary to rediscuss the value of DDA. Currently, other researchers believe the value of DDA is that it can be used for achieving a balanced challenge in video games. However, we defined DDA as a general game difficulty control mechanism and proposed that, more than creating a balanced challenge, it can control game difficulty more precisely. Such precise control makes DDA a valuable tool for designers.

The reason is that the game difficulty is generated within interactions and varies for different players. Such interactions, being dynamic and changeable, are hard to control. A well-known issue is that, ideally, player skills improve as the game progresses (Huniche and Chapman, 2005; Jennings-Teats et al., 2010), which constructs the basis of Flow-based DDA. More than skill, however, other OGD-related and SGD-related factors (e.g., player motivation, real-time task complexity, etc.) also change as the game interaction progresses. In this case, DDA can control the process of game difficulty by thoroughly collecting and evaluating the OGD and SGD data and modifying game tasks.

More importantly, different games should consider different goals (e.g., entertainment, training, education, etc.), and game difficulty has a great impact on achieving these design goals (Klimmt et al., 2009; Sarkar and Cooper, 2019; Dörner et al., 2016). Until recently, designers still needed to invest much time and effort in designing proper game difficulty when facing various goals (Adams, 2014; Fullerton, 2014). In this case, DDA as a general difficulty control mechanism can be valuable for designers to adjust the difficulty to achieve these goals.

In summary, DDA should be a game difficulty control tool that can guide game difficulty progress more precisely to support the difficulty design for various goals, which makes it valuable for designers.

7.4 How to Design DDA: Goals, Theories, Framework and Design Process

Designers are concerned with how DDA can affect specific design goal achievement. Generally, most video games can be separated into entertainment video games or serious video games (Dörner et al., 2016). Entertainment video games aim to create an enjoyable playing experience, namely entertainment goals. Serious video games are designed to achieve serious goals of education (Sampayo-Vargas et al., 2013), rehabilitation (Nirme et al., 2011), etc., and selectable entertainment goals (Dörner et al., 2016). DDA should be designed to control the game difficulty towards achieving these goals.

Fig. 7.3 shows DDA can help designers track the player curve (dashed line) and revise it to achieve the design curve (solid line). Difficulty curves represent how difficulty changes as the game progresses (Adrian and Luisa, 2013; Adams, 2014; Nagle et al., 2016; Aponte et al., 2011b). More generally, to design the progress of game difficulty is to design the objective or subjective difficulty curve and then validate whether the player curve of difficulty is similar to the design (Corrêa et al., 2022). In this sense, DDA can be a beneficial process-control tool that allows the designer to guide the progress of game difficulty better (Ouellette et al., 2019). Because our paper focuses on DDA design, we will not discuss how to design ideal difficulty curves but focus on how DDA can be designed to control the process of difficulty.

Therefore, it is important to establish reliable fundamentals for DDA before design practice begins. All designs need to be fit for their purpose (Buchanan, 1992; Cooper et al., 2014). We are concerned to know how DDA supports the realization of game design goals. Flow is no longer the primary goal of DDA and the evaluation criteria of “not too hard or not too easy” is also no longer applicable. Therefore, unlike studies that attempt to develop adjustment mechanisms of DDA, we propose a new DDA design methodology (i.e., a goal-based DDA design framework and a 6-step DDA design

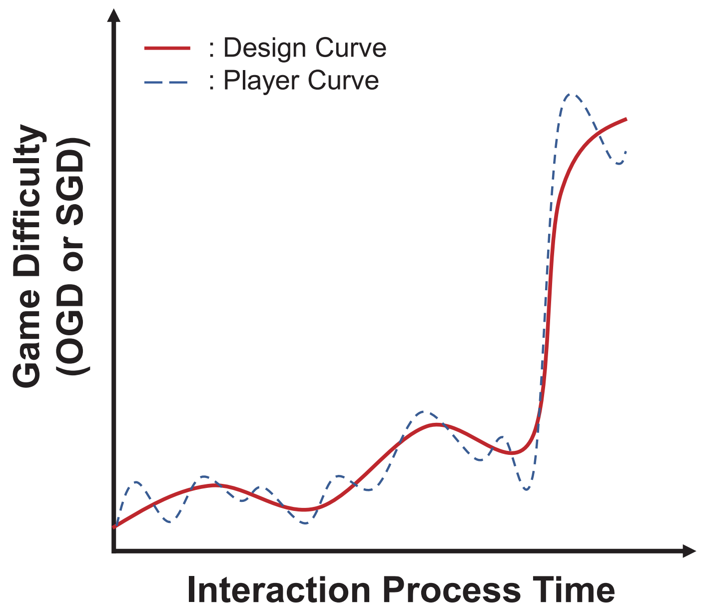

Fig. 7.3 DDA can be used to control the difficulty process in the player-game interaction. We redraw the figure according to Nagle et al. (2016) and Aponte et al. (2011b).

process, Section 7.4.3) by discussing these two issues: 1) how game difficulty should be adjusted to achieve goals (i.e., the relationship between difficulty and design goals, Section 7.4.1); 2) how DDA can evaluate based on goals (i.e., how goal-related theories serve the evaluation mechanism design, Section 7.4.2).

7.4.1 Game Difficulty and Design Goals

DDA design aims to control game difficulty better. However, how "better" should be defined depends on the primary goal of game design. It is necessary to clarify the relationship between OGD, SGD, entertainment goals, and serious goals to support designing DDA in making proper adjustments. Based on the findings of Chapter 4, these two difficulties are partially matched and should be designed and evaluated separately to support the achievement of design goals.

The design goal is a broad concept in video game design. Gibson (2014) divided design goals into designer-centric and player-centric dimensions. Designer-centric design goals are the personal goals of the designer, such as personal fortune or personal expression. By contrast, player-centric design goals, such as players' enjoyment and engagement, require designers to prioritize what the game design can do for players.

Here, we use the design goal to describe the ultimate purpose of game design and adopt it as a player-centric perspective.

We argue that design goals can be divided into entertainment and serious goals according to general game types. According to Schell (2019), the ultimate goal of an entertaining game is to deliver a great experience to the player. According to Dörner et al. (2016), serious games typically have two goals of entertainment and seriousness. For these two types of goals, entertainment goals are supportive, while serious goals are primary and can be specifically classified into characterizing goals (e.g., educational goals, rehabilitation goals, persuasion goals, etc.). Dörner et al. (2016) further proposed that characterizing goals should match the player's competence domains (e.g., cognitive and perceptual skills, sensory-motor skills, social skills, etc.).

Based on our findings in Chapter 4, SGD has a more direct and important impact on entertainment goals. SGD affects player experience, motivation to engage, and self-efficacy more directly. OGD has the potential to directly affect the player experience, but more through its impact on the SGD indirectly. In our opinion, players' self-explanation process relates OGD to SGD and thus becomes "*the source of game enjoyment*" (Juul, 2009). In conclusion, such influence from OGD to SGD and entertainment goals is complex.

Regarding serious goals, OGD seems to play a more critical role because serious characterizing goals and OGD are both related to the player's skills. According to existing literature, OGD has a direct impact on the effects of learning (Wu et al., 2012; Maertens et al., 2014), training (Ulmer et al., 2022), persuasion (Klisch et al., 2012; Bowman, 2018), and rehabilitation (Nirme et al., 2011; Andrade et al., 2016; Hocine et al., 2015; Pezzera and Borghese, 2020). The impact of SGD on serious goals depends on how player experience works on these goals. Usually, SGD contributes to this goal through its impacts on entertainment (Fu et al., 2009; Callies et al., 2015; Hooshyar et al., 2021), which still needs to be based on the hypothesis that good entertainment will improve engagement to better achieve serious goals (Laamarti et al., 2014; Dörner et al., 2016). However, in some cases, SGD can also work directly as an experience, depending on whether it is necessary to make players feel that the game is difficult. An example is the game *My Cotton Picking Life* (Rawlings, 2012), in which players need to pick cotton by repeating click operations constantly for 6 hours to complete the task. This game is not fun, but researchers found that such a torturous process (called Procedural Rhetoric in their paper) made players feel a high level of SGD and thus strengthened persuasive effects (i.e., affects players' attitudes to child slave labor; Jacobs et al., 2020). Therefore, OGD directly supports serious goals, but SGD's impact is more indirect and depends on the design scenario.

We summarize the relationships of design goals and game difficulty in Fig. 7.4:

1. Subjective game difficulty (SGD) and objective game difficulty (OGD) interact with each other.
2. In entertainment games, entertainment goals are essential and are affected directly by SGD but also potentially by OGD.
3. In serious games, serious goals are primary and are affected by OGD and entertainment goals directly; OGD potentially affects entertainment goals, and SGD potentially affects serious goals.

These relationships indicate that game difficulty adjustment should be given more attention to achieve the specific design goal.

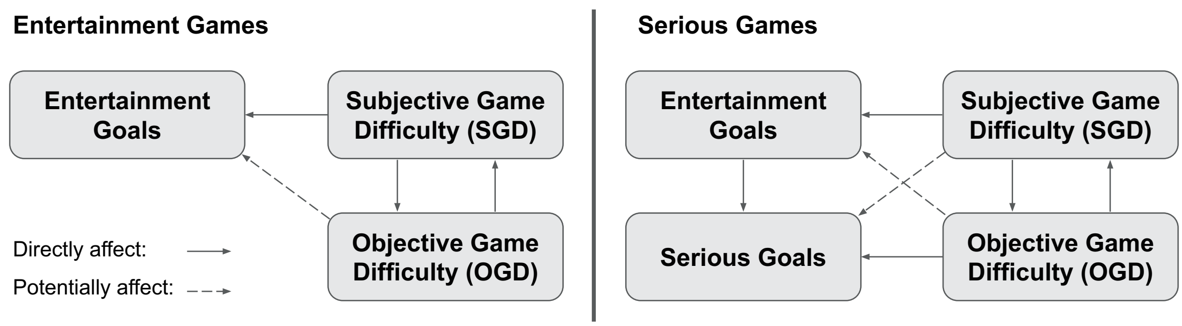

Fig. 7.4 The interactive relationships between subjective game difficulty (SGD), objective game difficulty (OGD), and design goals in entertainment games and serious games.

7.4.2 Goal-related Theories that Serve DDA Design

There is still a lack of clear distinction between the two difficulties in DDA design research (Dziedzic and Włodarczyk, 2018). This gap in understanding causes current DDA design issues, which can now be explained. Flow-based DDA mainly adjusts game difficulty by evaluating the player's performance and thus contributes more to serious goals but not to entertainment goals. In this case, we might expect more research that reports the effects of DDA in achieving serious goals and fewer reports on entertainment.

The issue of "*player overconfidence*" (Constant and Levieux, 2019), which is caused by DDA might be because DDA creates a mismatch of these two kinds of difficulty by producing lower failure and thus misleading players regarding estimations of SGD. Therefore, it is necessary to clarify how to design DDA to adjust OGD and SGD towards specific goals.

Generally, design goals instruct the directions for difficulty adjustment. Such instructions should be provided by the evaluation mechanism in DDA. Thus, goal-related theories can be used in the evaluation criteria design of the evaluation mechanism of DDA. Fig. 7.5 shows how these two goals and related theories provide evaluation criteria and guide the game difficulty to be adjusted. There follows a detailed discussion.

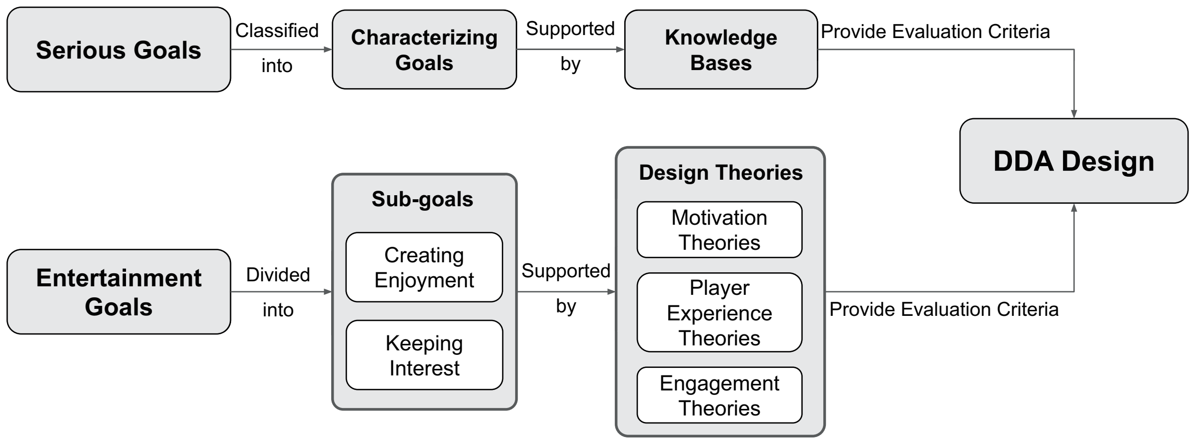

Fig. 7.5 Entertainment Design Goals and Serious Design Goals provide evaluation criteria for DDA design through related theories.

As we mentioned, the entertainment goal can be thought of as delivering a great experience for players. To be more precise, the heart of entertainment is enjoyment (Mekler et al., 2014) and according to Vorderer et al. (2004) and Tan (2008), the key factors shaping enjoyment of the entertainment experience are interest and imagination. We here identify entertainment goals as two sub-goals: 1) creating enjoyment for players and 2) keeping them interested in the game. Various theories are currently available as design support to achieve these two sub-goals. These theories can be roughly divided into three categories: theories of *motivation*, *player experience*, and *engagement*, which we have introduced in Section 2.4.

Therefore, motivation and experience-related theories identify the enjoyment of entertainment and introduce how to satisfy it and measure it subjectively to support

DDA design for the first sub-goal of entertainment. Engagement theories support the second sub-goal by answering the question of whether players' interests are sustained. These three types of theory provide guidance for the design of DDA, and explain how to establish the evaluation criteria in designing the evaluation mechanism of DDA. Specifically, DDA for entertainment goals should establish evaluation criteria based on these three categories of theories. For the motivation part, creating criteria to adjust OGD and SGD to most motivate players by fulfilling challenge and relaxation needs that adapt to player characteristics. For the player experience part, creating criteria to evaluate whether the player experience is under supervision and kept healthy. One example is that Moon et al. (2022) reported the application of player experience by modeling players in their DDA evaluation mechanism design. For the engagement part, the objective state measuring of cognitive, behavioral, and emotional aspects can be used to evaluate players to sustain good engagement and has now been applied in some DDA research (Buttussi et al., 2007; Liu et al., 2009; Hocine et al., 2011; Cruz and Uresti, 2017); replay and churn rate can also be used to establish evaluation criteria for DDA in a long-term engagement perspective.

In short, evaluation criteria of entertainment goal-based DDA should be designed according to entertainment-related theories. To serve entertainment goals, the evaluation criteria can be considered: 1) player motivation, 2) subjective player experience, and 3) objective engagement state. These three criteria need not be utilized simultaneously, depending on design requirements.

Turning to serious goals, general theories have limited guidance due to the differences in specific serious characterizing goals. According to Dörner et al. (2016) and Wiemeyer and Hardy (2013), there are six general competence domains, and various serious characterizing goals can be classified accordingly. However, due to characterizing goals varying, difficulty design cannot be guided by general theories, but related references are scattered throughout the various empirical studies (Rawlings, 2012; Kickmeier-Rust and Albert, 2010; Sampayo-Vargas et al., 2013). Dörner et al. (2016) then proposed that the corresponding knowledge base should be used not only to design the serious tasks but also to evaluate whether the game can gain the serious effect of the characterizing goal (also see in Hussaan and Sehaba, 2013). For example, Sampayo-Vargas et al. (2013) designed a serious game with the educational goal of learning about Spanish cognates. They created the game difficulty and DDA mechanism based on the Victorian Curriculum and Assessment Authority, an authorized educational knowledge base.

Therefore, DDA design for serious goals needs to rely on corresponding knowledge bases to create the evaluation criteria.

To summarize, first, serious goals can be classified into different characterizing goals and should be supported by corresponding knowledge bases separately, which provide evaluation criteria for DDA design; next, entertainment goals can be divided into two sub-goals and supported by three categories of design theories, which provide evaluation criteria for DDA design not only for entertainment games but serious games.

7.4.3 Goal-based DDA Design Framework and a 6-step Design Process

Referring to the literature (Lopes and Bidarra, 2011), to realize the proper difficulty control according to our DDA definition, DDA should be designed considering three issues: 1) what the current difficulty is, 2) what the following difficulty should be, and 3) how to adjust the difficulty. Therefore, we agree that DDA contains two essential components: *an evaluation mechanism* that solves the first and second issues and *an adjustment mechanism* that solves the third issue. To support the design of these two components of DDA, we proposed a DDA design methodology, including a DDA design framework and a 6-step design process.

We argue that DDA should be designed for entertainment games or serious games. For entertainment games, the general design goals are entertainment goals, and designers need to focus on realizing their specific entertainment goal (e.g., challenging in process or interesting for replay). For serious games, both serious goals and entertainment goals should be considered. The reason is even though the serious goals are primary, entertainment in serious games impacts the serious goals in most cases. Designers need to consider how entertainment affects the serious effects carefully in different design scenarios. For example, for educational goals, knowledge popularization games (e.g., knowledge about toxic chemicals; Klisch et al., 2012) are more likely one-time games that focus on player motivation and experience in one playing process, while skill training games (e.g., Sampayo-Vargas et al., 2013) are long games that need to focus more on player experience, state and how player's interests can be sustained for a long-term engagement. Fig. 7.6 shows the framework of goal-based DDA design. Such design is not only about the two components of DDA, but the whole consideration of design goals, OGD, SGD, and how they support the design of evaluation and adjustment processes of DDA.

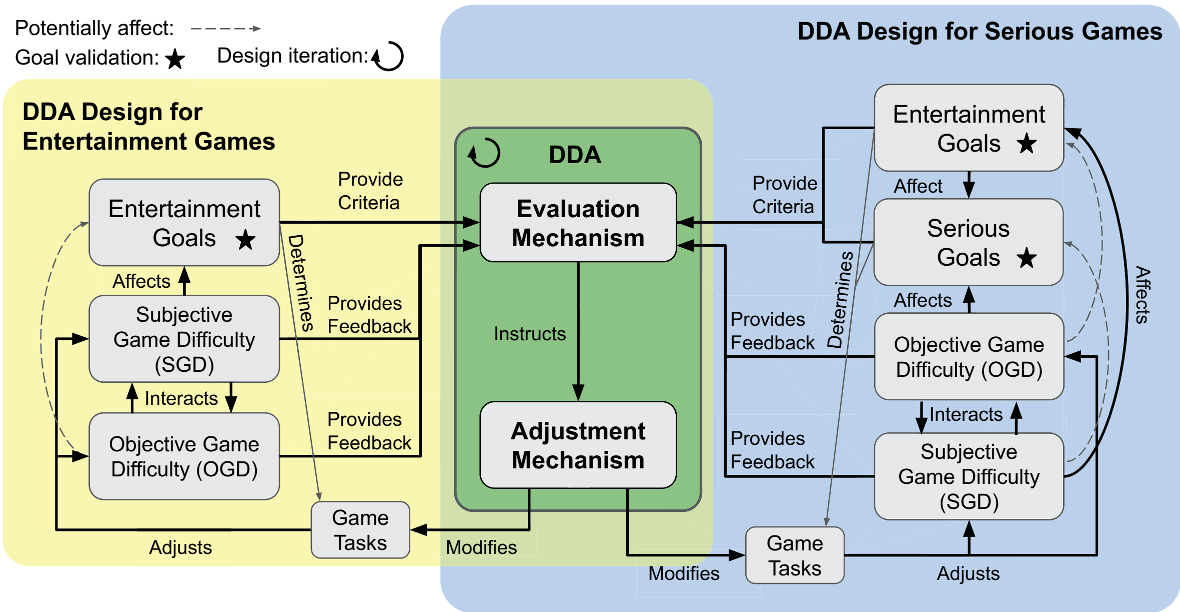

Fig. 7.6 The goal-based DDA design framework for entertainment games and serious games (partially integrating Fig. 7.1 and Fig. 7.2). This framework provides comprehensive design insights that are a significant improvement over the narrow focus on Flow theory in DDA design.

The evaluation mechanisms should be designed in three parts: 1) database of evaluation, 2) evaluation criteria, and 3) difficulty adjustment instructions. The database is designed to store and evaluate real-time SGD and OGD data to support practical evaluation. The data category is difficulty-related and can include player performance, player behavior, player difficulty experience, player pressure & emotion, etc., which rely on specific design requirements. The evaluation criteria should be highly based on the design goals and related theories. In addition, the content of player characteristics can be used in evaluation criteria design in two ways: 1) as criteria for those DDA mechanisms that adapt to player preference and 2) as a model for explaining OGD data to support evaluation. Finally, difficulty adjustment instructions are about how the evaluation results indicate the following difficulty should be. The instruction should be designed to guide the OGD and SGD adjustment properly and purposely.

The adjustment mechanism is designed for modifying the game task to adjust OGD and SGD separately (Sakaue et al., 2023; Yanase and Narumi, 2016). It should be noted that by modifying game tasks, OGD can be adjusted more directly but SGD is usually affected through OGD adjustment. Research in adjusting the SGD directly is scant; one example is Zhang (2021) who tried to control SGD by adding fake bullets (that cause no damage or cannot hit the player) in a shooting game. We argue that time pressure (that may increase nervousness; Masanobu et al., 2017; Denisova and Cairns, 2015), tips (that may provide help), task perception (e.g., the enemy looks invincible), and other factors have the potential to affect SGD and can be applied in the adjustment mechanism design. However, due to the complex interactive relationship between these two difficulties, such adjustment should be subtle and rely more on the practical testing process.

Based on our design framework and inspired by the work of Hendrix et al. (2018), we propose a 6-step design process for DDA as a general design approach.

Step 1. Determining design goals. Designers need first to determine the goal of the game and then design the DDA accordingly. The primary goal should be confirmed as entertainment or serious. The design goals of game difficulty in entertainment games include the enjoyment of challenge or relaxation, and keeping players interested in the game, etc. The design goals of game difficulty in serious games include serious goals like persuasion, learning, rehabilitation and supportive entertainment goals like making players' engagement state sustained. All subsequent difficulty and DDA designs should consider direct and indirect impacts on these design goals.

Step 2. Determining the game task. The game task is what the player interacts with and is the basis on which to produce the game difficulty. Game tasks should be decided based on specific goals and game genres. For example, a math educational game has the serious goal of math education and the entertainment goal is to make this learning process fun. The basic game tasks should be determined as math learning tasks like calculation, number reasoning, etc. Task complexity is tightly related to game difficulty and should be appropriately designed. For example, for the calculation task in this game, task complexity can be confirmed by how many numbers should be calculated and how many types of calculations.

Step 3. Identifying game difficulty. Designers need first to find the relationship between difficulty and game task. OGD and SGD are the interaction results between players and game tasks; these should be identified by design investigations. Games are mainly designed for a specific group of players, especially some types of serious games. It is necessary to clarify how this group of players' OGD and SGD relate to particular tasks. Because the OGD and SGD interact with each other, we recommend that designers distinguish all elements into three groups: affects OGD, affects SGD, and affects both. For example, in a shooting game, the damage value of the enemy may affect OGD, the number of tips and time pressure may affect SGD, and the number of enemies may affect both. Next, designers need to design the game difficulty curves. Designed curves of SGD and OGD are the baseline references for DDA to track; they must be designed before DDA.

Step 4. Designing the difficulty adjustment mechanism. The difficulty adjustment mechanism should be designed to affect the OGD and SGD effectively. First, designers need to select different task elements from those three groups (i.e., affects OGD, affects SGD, and affects both) to change (lower or increase) OGD or SGD. To avoid unforeseen circumstances caused by the complex interaction effect, fewer elements are better for efficient adjustment. Second, designers need to design the change interval for each element, making sure that OGD and SGD are adjusted smoothly to corresponding levels. We recommend adopting novel and effective adjustment techniques only when they are essential for adjustment effects. Finally, the difficulty adjustment mechanism should organize all element adjustments to modify the game task, and this modification should follow instructions from the evaluation mechanism. To support game difficulty design based on the proposed OGD and SGD, best practices and experiences related to the above difficulty adjustment mechanisms can be formulated into templates that can be shared with game designers.

Step 5. Designing the player evaluation mechanism. The player evaluation mechanism should be designed to confirm the current situation and instruct the adjustment direction. Designers should decide the evaluation contents and criteria to trigger adjustment. Data collection methods should avoid interruptions to ensure a smooth gameplay experience for players. For example, forcing players to rate the game's difficulty while playing can be annoying. Evaluation contents and criteria should be based on goals and theories. For example, an evaluation mechanism for serious educational games can evaluate four items: 1) the player learning performance in games (OGD); 2) the perceived difficulty feeling of players (SGD); 3) whether the players are motivated (entertainment goal-related criteria), which can be measured by physiological measurements (Gergelyfi et al., 2015; Sideridis et al., 2014) and experience scales (IJsselsteijn et al., 2013); 4) whether the players make progress in learning (serious goal-related criteria), which can be measured by comparing player performance with educational standards. The evaluation mechanism should provide specific adjustment instructions by comparing these contents with the evaluation criteria and baseline (difficulty curves). In this example, the SGD should be easier if player motivation is lower; or OGD should be more difficult if players meet the educational standards well but still need more training.

Step 6. Validating the design goal and iterating the design. Designers need to validate whether their DDA design is effective for achieving the design goals and iterate their design. Designers should validate the realization of their design goals by confirming that (1) appropriate and effective difficulty changes have been created by DDA and (2) specific goals (e.g., enjoyment of relaxation, interest sustainment, and learning) have been realized. Effect comparisons with other designs are optional but valuable, for example, comparing DDA with other difficulty mechanisms in commercial games on the entertainment effect, or DDA in serious educational games with traditional education methods on the learning effect. Designers need to iterate the DDA design by checking Steps 2-5 and adjusting or redesigning relative content until the intended goals are achieved.

7.5 Discussion

This section summarizes our general answers to our four questions and then it further discusses how our work can resolve current DDA design issues.

7.5.1 The Answers to Q1-Q4

We asked four questions in Section 7.1 and attempted to answer them through theoretical exploration in Sections 7.2-7.4. Here we summarize general answers to each question.

Q1: What limits the effectiveness of current DDA?

A1: DDA arose with the impact of Flow theory and is restricted by this relationship. Neither Flow nor challenge-skill balance is strong support for flexible DDA design, and this causes DDA design issues in many studies and commercial applications. Our work provides a critical perspective on DDA's foundations for designers. We argue that DDA is a game difficulty mechanism and should not depend on or be limited by Flow-based definitions and goals but on the analysis of game difficulty in player-game interaction.

Q2: What is the definition and scope of DDA?

A2: Our definition is that *DDA refers to a game difficulty control mechanism that aims to control the difficulty automatically in game interaction by evaluating objective* *and subjective game difficulty data and modifying game tasks*; it provides a novel view on DDA and extends its connotation. The scope of DDA becomes clear based on clarification of DDA's attributes, i.e. *DDA is an adaptive game difficulty mechanism that only takes effect in game interaction*, which can help designers to distinguish DDA from other mechanisms. In summary, DDA is not designed to pursue a balanced challenge but to control the process of game difficulty during interactions.

Q3: Why is DDA valuable for video game design?

A3: We argue that DDA can be used to control game difficulty processes more precisely and effectively because the game difficulty is dynamic and changeable in the interaction process. Based on this, DDA can support designers in designing the whole game difficulty control process to achieve specific goals. This is the advantage of DDA, which makes it valuable. However, DDA's value is currently underestimated. One reason is that Flow theory over-simplifies the DDA design goal and prevents designers from understanding and applying DDA's deeper potential. On the other hand, designers seem to find it hard to imagine how DDA can be applied in their design without balancing challenges. We argue that liberated from Flow, DDA can be more flexible and potent in affecting interaction and supporting better entertainment and serious game design.

Q4: How to design DDA for video games?

A4: We first propose a goal-based DDA design framework to summarize our literature review findings. A 6-step design process is then presented as a general DDA design approach to support design research and practice: (1) determining design goals; (2) determining the game task; (3) identifying game difficulty; (4) designing the difficulty adjustment mechanism; (5) designing the player evaluation mechanism; (6) validating the design goal and iterating the design. As Flow is no longer the primary goal of DDA, our work fills this gap by suggesting how to design goal-based DDA.

7.5.2 Comparison of the Proposed Method with Two Existing Methods

We first introduce two design methods from Gibson (2014) and Miyake (2015). The method proposed by Gibson is the Layered Tetrad framework and related design guidelines, which organically combines three famous game design frameworks of MDA (i.e., Mechanics, Dynamics, and Aesthetics, by Hunicke et al., 2004), FDD (i.e., Formal, Dramatic, and Dynamic Elements, by Fullerton, 2014) and the Elemental Tetrad (i.e., Mechanics, Aesthetics, Technology, and Story, by Schell, 2019). This method is about video game design and is comprehensive enough to be a good design reference (O'Shea and Freeman, 2019). The game AI design method from Miyake (2015) provides the framework and design guidance for game AI and is also related to DDA design.

The Layered Tetrad framework introduces four elements: mechanics, aesthetics, technology, and narrative, as well as three layers of inscribed, dynamic, and cultural. Each layer contains those four elements. Designers design games on the inscribed layer, and players play games to interact with the inscribed layer to form the dynamic layer. Gibson mentioned that the designer is responsible for the experience at the dynamic layer through game system design, but it is very difficult. We argue that DDA should belong to the element of "mechanics" and be more related to the inscribed and dynamic layers. Therefore, DDA can be very valuable in solving the game system design issue that Gibson (2014) proposed. Based on our framework and design process, by designing DDA in the inscribed layer, designers can control the game difficulty in the dynamic layer by DDA. This allows the designer's influence to extend into the player's game interaction process. Specifically, designers need to consider what DDA should be designed based on our framework, and then confirm what difficulty will happen to players in design step 3 and design the control process in dynamic interaction in design steps 4 and 5. Therefore, the proposed method contributes to expanding and improving the difficulty control part of Gibson's method.

The game AI design method from Miyake (2015) introduces a framework containing three main kinds of game AI: Character AI, Meta AI, and Navigation AI, and their design ways. Character AI is the control mechanism of a non-player control game character. Meta AI is the game system's AI that controls game sequences by generating enemies, terrain, and other game content. Navigation AI supports pathfinding and other related functions of character AI and meta AI by providing information about the level (a part of the game world). It can be seen that the character AI and the meta AI are more related to DDA and should be considered in DDA design. Miyake discussed how to make these AIs and mentioned the game example of "*Left 4 Dead*" (Valve, 2008), which applies DDA as part of meta AI in its game system. However, information about designing DDA to support game AI is still limited in this method, and clear design guidance for game AI-contained DDA is not mentioned by the author.

Designing DDA to support game AI can be a specific design goal, which means using DDA to make game AI more adaptive. Due to the varying games, a game can contain different game AI types. For example, the Go game only needs a character AI as the player's opponent, while a shooting game usually needs the character AI to control the enemy and the meta AI to generate the enemy. Designers need to determine the game task and difficulty in design steps 2 and 3. For this shooting game, for example, the task can be to beat five strong enemies. Based on our literature review, the character AI in DDA design can refer to the adaptive game AI, and the meta AI is related to the adaptive content generation and adaptive content adjustment, which all belong to the difficulty adjustment mechanisms. Therefore, in design step 4, designers can design the difficulty adjustment by applying these techniques. Continuing with this example, designers can adjust the attacking desire of the enemy's character AI towards players, making it adaptive. They can also use meta AI to change the number of enemies and adjust the weapon damage of players to modify the game difficulty. How and when the game AI should be activated to adjust difficulty depends on the evaluation mechanism design in design step 5, based on this specific design goal (support game AI) and on the primary goal (for entertainment or seriousness). Therefore, the proposed method provides clear DDA design guidance to complement Miyake's game AI design method.

In summary, our goal-based DDA design framework and 6-step design process can fill the gap in current game design methods and can be used to support better game design along with these two design frameworks. We look forward to more researchers verifying this DDA design methodology through specific design practices.

7.5.3 How Our Framework Dispels Two Concerns from DDA Designers

Even though interest in DDA research keeps growing, most game designers treat DDA cautiously when designing video games (Adams, 2014; Schell, 2019). Researchers indicated that Flow-based DDA is automatic in adjusting difficulty but needs to be invisible to players (Hunicke, 2005). This view causes designers to be concerned that DDA may have adverse effects on their game designs. The two main concerns are 1) automatic DDA may take control over players from designers (Kristensen et al., 2022); 2) it is difficult to make DDA completely invisible (Schell, 2019; Constant and Levieux, 2019). To dispel these two concerns, we provide brief discussions on how our definition and design framework of DDA solve these issues.

Does DDA take designers' control over players? Some researchers mentioned designers' control over game difficulty. Kristensen et al. (2022) implied that designers worry that DDA will take their control over the game difficulty, and thus designers cannot guide the player experience. Suppose the goal of DDA is only to avoid difficulty levels that are too hard or too easy, it is reasonable for designers to worry that game difficulty can be adjusted by computer solely with no need to design. However, their concerns only related to Flow-based DDA. As mentioned in Section 7.3, our DDA is a difficulty control mechanism and such control on game difficulty should be based on difficulty curves and goal-based evaluation criteria. Designers can freely design their difficulty curves and evaluation criteria toward specific goals. In this sense, DDA supports more precise control of designers rather than taking the designers' control over difficulty away.

Can DDA be visible? The visibility of DDA has gained strong concern from researchers and designers since its birth. Hunicke (2005) indicated that DDA should be invisible or players may feel deceived by games. Schell (2019) proposed that DDA should be used cautiously, or if players discover it, this mechanism could be used for cheating, e.g., by purposely losing, DDA will lower the difficulty for the player. However, as Hunicke (2005) admitted, it is almost impossible to keep DDA invisible to players when the game task is modified frequently. This has led many designers to stay away from DDA, because they believe it would be a disaster if DDA were exposed. In addition, researchers have found that invisible DDA has other issues, for it misleads players into over-evaluating their performance (Alexander et al., 2013) and makes them overconfident (Constant and Levieux, 2019).

According to our work, telling players honestly that game difficulty is adaptive might not be a poor choice. Firstly, our DDA is not only based on player performance (OGD) to balance challenges as Flow-based DDA, so the value of cheating for players is relatively lower. More specifically, our DDA is designed to rely more on goal-based evaluation criteria, e.g., setting enough trigger conditions combining performance, playing time, and player emotion. Therefore, even though players perform poorly on purpose, such DDA would not lower the difficulty directly but first confirm and evaluate the whole player situation under evaluation criteria. Secondly, making DDA visible can remove barriers to feedback collection from players, while the invisibility of DDA may cause unknown results from the interaction. Finally, DDA is designed for specific goals, not only for controlling players. Even though invisible DDA may support better control, it should be kept in mind that DDA is intended to better achieve design goals (e.g., good player experience, learning). Therefore, finding the best way to combine visible DDA in game design is better than hiding it awkwardly. For example, designers can transfer the control of DDA to players in case they prefer to turn this mechanism off. Recent studies support our views (Denisova and Cairns, 2019; Masanobu et al., 2017) and show that using visible DDA can also create a positive experience if players have a positive attitude toward game adaptation. We believe this concern can be dispelled gradually as more designers accept and apply visible DDA in their game designs.

Chapter 8

Case Study - A Cognitive Training Game for the Elderly

We have proposed a new DDA design methodology in Chapter 7. This chapter tests this approach in guiding the DDA design through a case study. In this case, we design a cognitive training VR exergame with a DDA mechanism for the elderly and test whether the designed game achieved serious and entertainment goals. The experiment results show that our methodology has guided the DDA design well in this case.

8.1 Background

Elderly people (over 60 years old) commonly suffer from cognitive decline problems, e.g., in memory and attention (Harada et al., 2013). Research has shown that the brains of the elderly are still plastic (Hübener and Bonhoeffer, 2014), and the cognitive abilities of the elderly can be improved through cognitive training (Kueider et al., 2012). Cognitive training is a systematic method developed for improving specific cognitive abilities through repetitive training of the brain. For example, a study by Wenger et al. (2012) found that the elderly group who received navigation function cognitive training effectively improved their way recognition. In addition, cognitive training can also improve the mood of the elderly in daily life (Smith et al., 2009), and enhance the self-efficacy of the elderly to a certain extent (Cavallini et al., 2003). In short, cognitive training is beneficial for the improvement of abilities and the quality of life of the elderly.

Recently, video games have been used for cognitive training. This approach has been shown to be effective in improving the cognitive abilities of the elderly (Anguera et al., 2013; Anguera and Gazzaley, 2015; Bainbridge and Mayer, 2018; Hardy et al., 2015). Cognitive training was reported as more likely to cause cognitive fatigue for its boring repetition, which compromised the training effects (Anguera and Gazzaley, 2015). In comparison, elderly people may feel more enjoyable when training by games.

Commercial games were first used for cognitive training, but such games are not specifically designed to train certain types of cognitive abilities, which seems to make the effect of ability improvement vague and uncertain (Spence and Feng, 2010; Toril et al., 2016). In addition, although the effect of exercise in improving cognitive abilities has been confirmed (Colcombe and Kramer, 2003), commercial exergames (i.e., exercise games) have also been found to be vague in the effects (Ordnung et al., 2017). Therefore, attention has been drawn to specially designed serious games (Dörner et al., 2016).

To design the cognitive training games, a promising approach is based on the multitasking paradigm (referring to conducting two or more tasks at the same time), which has proven effective in cognitive ability improvement (Dahlin et al., 2008; Anguera et al., 2013; Kayama et al., 2014). Multitasking paradigms involve attention, working memory, and executive functions of cognition; these abilities decline rapidly as the elderly age (Verhaeghen et al., 2003). Cognitive exergames were also developed in studies and showed similar enhancement effects with the group both conducting cognitive and physical training (Schättin et al., 2016; Guimarães et al., 2018). Virtual Reality (VR) technology has also recently been applied to cognitive training game design. According to research, VR technology stimulated the four different cognitive abilities of the elderly in training, including attention, memory, language, and visuospatial processing (Zajkac-Lamparska et al., 2019). Another study using a cognitive training VR exergame (CTVR exergame for short) for training has also shown a good effect on improving the working memory of the elderly (Li et al., 2020).

However, designing an effective cognitive training game that combines these complex factors is still a challenge. Especially, because elderly people's abilities and training processes are various, it is difficult to design the optimal game difficulty by considering all these factors. Game difficulty is a critical factor affecting player experience (PX) and training effects, which should be carefully designed in cognitive training games (Dörner et al., 2016). Therefore, it is a good choice to design a Dynamic Difficulty Adjustment (DDA) mechanism that can adaptively adjust the difficulty of the game. Currently, studies on designing DDA in cognitive training games mainly rely on the Flow theory (Hendrix et al., 2018; Hocine et al., 2015; Kitakoshi et al., 2020) and showed mixed results.

Based on the DDA design methodology we proposed in Chapter 7, we design a CTVR exergame for training the elderly in this chapter. We determined the goal of the game design is cognitive training (Step 1) and adopted multi-tasking (Step 2) in design. The OGD was designed by identifying four factors of quantity, types, input, and time of gameplay (Step 3). Subsequently, we designed the adjustment mechanism (Step 4) by changing the complexity level of the factors and the evaluation mechanism (Step 5) assessed players' performance to guide the adjustment. We conducted a pretest to validate our game design (Step 6). After the pretest, we experimented to validate the training effects and PX of the designed DDA mechanism and Manual Difficulty Adjustment (MDA) mechanism (Salehzadeh Niksirat et al., 2017). The results proved that the elderly who played the game with the DDA achieved the same training effect as MDA but had a better experience.

8.2 Game Design

Following the design steps proposed in Chapter 7, for the first step, we set our game design goals as (1) improving the cognitive abilities of the elderly and (2) providing a good PX. Subsequently, it is necessary to determine the game task (the second step). As we mentioned in Section 8.1, it would be effective to combine the factors of multitasking, exercise, and VR in the game design. Multitasking can enrich the fun of the game without compromising the training effect, which can be the core task of the game. More specifically, we decided to adopt the "Whack-a-Mole" task for design, which is a typical multi-task and adopted in studies (Urakami et al., 2021; Li et al., 2020; Salehzadeh Niksirat et al., 2017). The task of "Whack-a-Mole" is to hit multiple animals that appear from the holes at the same time. We changed it to a feeding form to avoid the elderly feeling uncomfortable about the violent hitting actions (Salehzadeh Niksirat et al., 2017).

We designed and developed the game, *Zoo Feeder*, which runs in a VR environment. The main game scene is a zoo set on a small island, see Fig. 8.1. Players need to wear a VR helmet and use the VR controllers to play the game. In the game, players need to play a feeder to feed the animals in the zoo. As a multitasking, players are required to feed multiple animals with corresponding foods quickly and correctly, or they will disappear. For the design of the exercise element, players are required to do more movements in the game. Therefore, to feed the animals, players need to use the left hand to get food and the right hand to feed them, both by the controllers. Additionally, food is designed in places far away from animals, and players need to complete feeding tasks through moving, turning, and arm raising. Imitating the content of the game

"Whack-a-Mole", we set up 12 bowls for feeding in the game. Suspended food for feeding will be on the left side of the player's field of vision. After the game starts, different types of animals will appear and move to the back of the bowl. Players must immediately grab the corresponding food (left hand) and feed the animal by selecting the correct bowl (right hand). The auxiliary aiming line and aiming cursor will assist the player when feeding, see Fig. 8.2.

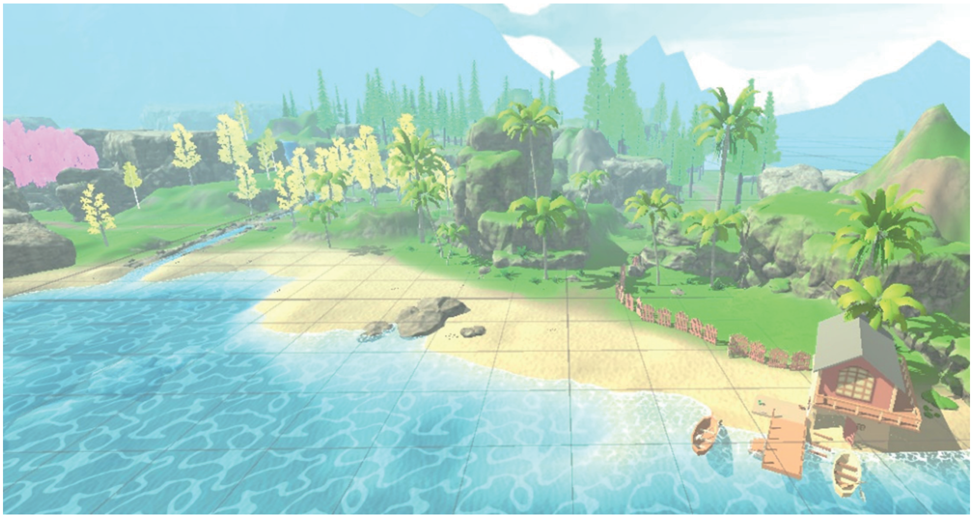

Fig. 8.1 The main scene used in this game is a zoo set on a small island.

Fig. 8.2 The gameplay is that players grab the corresponding food with the left hand and feed the animal by selecting the correct bowl with the right hand. The auxiliary aiming line and aiming cursor will assist the player when feeding.

When playing, the game automatically calculates and records the player's feeding performance. There are two situations of feeding failure: feeding is not completed within the specified time, and the food fed does not match the animal. In other situations, e.g., not selecting the correct bowl, players are allowed to keep trying.

To design the game's difficulty, we further designed the content and complexity of the tasks. Players need to perform multitasking in the game, and the content of the multitasking determines the complexity and thus affects the OGD. Therefore, from the two aspects of multitasking and OGD, we designed four dimensions of task complexity. More specifically, for multitasking, the number and difference of tasks are related to complexity. We accordingly designed the numbers and types of animals in the game. For OGD, input correctness and time are two essential factors. We accordingly designed the bowl size and the time limits. To distinguish these four dimensions, we titled them: quantity complexity, type complexity, input complexity, and time complexity. All four dimensions have three levels of complexity. See Table 8.1.

Table 8.1 Complexity levels setting of the four dimensions of quantity, type, input, and time

Dimensions	Level 1	Level 2	Level 3
Quantity Complexity (animal number)	1	2	3
Type Complexity (animal type)	1	2	3
Input Complexity (target radius)	4	3	2
Time Complexity (time limit)	15 seconds	12 seconds	9 seconds

The game is developed using the Unity3D engine and testing environment is an AMD 3700X PC equipped with a Windows 10 system and HTC VIVE platform. All the elements and scenes in the game were constructed by combining free and paid assets in the Unity Asset Store[*1].

8.3 Game Pretest

Based on our DDA design methodology, the third step requires identifying the relationship between task and game difficulty through design investigations. Therefore, we conducted a pretest on our game prototype to test our game and difficulty designs. The goal of the pretest is to test whether: (1) the game was well-developed for playing and the data could be auto-collected correctly and (2) all the four complexity dimensions can produce expected different levels of OGD and SGD.

[*1] https://www.assetstore.unity.com/

8.3.1 Participants

The participants recruited for this study were healthy elderly people from Fujian, China. A total of 22 people were recruited, including 9 males and 13 females. The ages of participants ranged from 62-79 years old (M= 69.81, SD = 4.69). Participants with underlying medical conditions, depression, and anxiety were excluded from the test. Moreover, to ensure consistency with formal cognitive training, we only selected elderly people with healthy cognitive function status. The Mini-mental State Examination (MMSE; Tombaugh and McIntyre, 1992), Self-rating Anxiety Scale (SAS; Zung, 1971) and Self-rating Depression Scale (SDS; Zung, 1965) used to test their cognitive and health status. Participants with MMSE scores below 24 points (primary school education below 20 points), SAS scores above 50 points, and SDS scores above 53 points were excluded. Participants who passed the screening all signed the informed consent form and fulfilled the form of demographic information collection. All 22 participants participated in the pretest and the interview after the pretest.

8.3.2 Materials and Apparatus

The game in this study runs on a PC equipped with Windows 10 system. The computer's CPU is a 3.6 GHz AMD 3700x processor. The VR device uses the HTC VIVE Cosmos VR kit (including a helmet, two controllers, and two laser positioning base stations). The experiment was conducted in an elderly activity room in a community in China. The room space is enough for VR game playing, see Fig. 8.3. The data on player performance was collected in-game automatically, and the data analysis utilized IBM SPSS 26. After playing the game, participants must attend the structured interview to provide their opinions about the difficulty. We set two questions: (1) Did you feel the difficulty change? (2) How do you feel about the overall difficulty? These two questions were set for investigating the SGD.

8.3.3 Procedure

The participants were randomly divided into four groups, and each group corresponds to one dimension. The four groups were the quantity group with 5 people (M = 66.40, SD = 2.97) and the type group with 6 people (M = 71.83, SD = 3.13), 6 people in the input group (M = 68.33, SD = 5.47), and 5 people in the time group (M = 72.60, SD = 4.83). The participants were given sufficient time to practice in the guiding level

Fig. 8.3 The room space for VR game playing in the game pretest and the following experiment.

before the formal test. After participants confirmed that they had learned the basic gameplay, they could start the formal test at any time. Participants were required to play all three complexity levels of their group's dimension while the other dimensions were kept at the first level. Participants played the three levels randomly to balance the order effect. Each complexity level had 30 trials, and the success rate of operations was recorded automatically as the OGD indicator. After the game, players took the interview. The entire pretest process for each lasted approximately 30 minutes.

8.3.4 Results and Discussion

The game ran well, and all the data was collected correctly. Using these data, we conducted a repeated-measures ANOVA and Bonferroni post hoc test, see Table 8.2. Results showed that there were significant differences in the three levels of the quantity group ($F(2, 4) = 22.935$, $p < 0.001$) and the input difficulty group ($F(2, 5) = 30.089$, $p < 0.001$). However, there were no differences in the three levels of type and time group. According to the post hoc test, for the quantity group, the first level was significantly easier than the second level ($p = 0.008$) and the third level ($p = 0.005$); but there were no differences between the second and the third level. For the input group, the first level was significantly easier than the second level ($p = 0.041$) and the third level ($p = 0.002$); the second level was also significantly easier than the third level ($p = 0.028$).

Table 8.2 The results of repeated-measures ANOVA and Bonferroni post hoc test on the four dimensions of quantity, type, input, and time (each with three levels). Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

Dimensions	Levels	M	SD	F	p	Bonferroni
Quantity Complexity	Level 1	26.20	2.864	22.935***	$p < 0.001$	1<2, $p = 0.008$**
($N = 5$)	Level 2	23.20	3.421			1<3, $p = 0.005$**
	Level 3	19.80	1.304			2<3, $p = 0.130$
Type Complexity	Level 1	21.33	6.743	1.217	$p = 0.336$	1<2, $p = 0.877$
($N = 6$)	Level 2	23.83	3.061			1<3, $p = 1.000$
	Level 3	23.00	4.472			2<3, $p = 0.776$
Input Complexity	Level 1	28.00	1.673	30.089***	$p < 0.001$	1<2, $p = 0.041$*
($N = 6$)	Level 2	25.33	2.733			1<3, $p = 0.002$**
	Level 3	21.00	3.406			2<3, $p = 0.028$
Time Complexity	Level 1	20.60	5.771	1.146	$p = 0.365$	1<2, $p = 1.000$
($N = 5$)	Level 2	19.40	7.503			1<3, $p = 0.240$
	Level 3	18.20	6.573			2<3, $p = 1.000$

All 22 participants answered the two questions. For the first question, most of them (20 of 22) reported a change in difficulty, but only participants from quantity and input groups found the changes were from complexity. In more detail, participants in the quantity group mentioned that the number of animals was increasing, and the input group's participants mentioned the size of the bowl changed when playing. In contrast, participants from other groups believed the difficulty change was caused by their skill improvement. Therefore, this result of SGD is consistency with the OGD results. For the second question, 15 participants thought it was moderately difficult, 6 participants thought it was easy, and 1 subject thought it was slightly difficult. In short, the SGD is moderate and well-designed.

The results indicate that no differences in OGD and SGD of the two groups are caused by two reasons. One is that the complex levels of type and time dimensions are too low to be distinguished. In addition, because the number dimension was set at the lowest complexity level, the participants needed to select only one type of food to feed one animal. Thus, different OGD and SGD cannot be produced. Another reason is the visibility of task complexity. We did not design a visible countdown or different speed of disappearing for participants so they would hardly notice the differences in the different time limits. Therefore, these factors should be considered in the DDA design in the next step.

8.4 Game Mechanism Design

The fourth and fifth steps are to design the difficulty adjustment mechanism and player evaluation mechanism of DDA. Based on the pretest results, we first adjusted the complexity of the four dimensions. For the time dimension, we reduced the time limits of all levels by 3 seconds. In addition, we add the countdown above the head of each animal as a reminder. However, we did not adjust the other dimensions' complex level but expanded the complexity level to four levels for the number and type dimensions, see Table 8.3.

Table 8.3 The expanded version of complexity levels setting of the four dimensions of quantity, type, input, and time, in our formal training game.

Dimensions	Level 1	Level 2	Level 3	Level 4
Quantity Complexity (animal number)	1	2	3	4
Type Complexity (animal type)	1	2	3	4
Input Complexity (target radius)	4	3	2	None
Time Complexity (time limit)	12 seconds	9 seconds	6 seconds	None

8.4.1 Level Design

According to the four complexity dimensions, we designed nine game levels, and each has three sublevels (27 sublevels in total). The dimensions of quantity, time, and input with different complexity were used to design different levels. In addition, sublevels have different complexity of the number dimension. The nine levels were set in different places of the island scene to create better PX, see Fig. 8.4. The whole complexity of game levels increased as the level order. The complexity of sublevels also increased as the order. See Table 8.4.

Fig. 8.4 The nine levels in our games. These levels were set in different places of the island scene.

8.4.2 DDA and MDA Modes Design

Based on our level design, we first designed the difficulty adjustment mechanism. Our goal is to adjust OGD for a good cognitive training effect and adjust SGD for a good PX. Therefore, we realized our difficulty adjustment mechanism in three design aspects: (1) When the player passes the sublevel, the system automatically progresses to the next sublevel. If the player passes all three sublevels of a game level, the system automatically progresses to the next game level. (2) If players can not pass the sublevel in a specific number of trials, the system automatically rolls back to the previous sublevel. However, even if the player cannot pass, the system will not roll back to the previous game level but stay at the first sublevel of the game level. (3) any changes in the level will not provide a reminder. The first two aspects are designed to provide optimal OGD. In addition, not rolling back to the previous game level is for a better training effect and sustains players' self-efficacy. The final aspect is designed to avoid SGD changes and keep players in the Flow state.

For the player evaluation mechanism, we evaluated player performance based on the real-time success rate of the feeding. More specifically, after feeding 10 trials, if the real-time success rate is equal to or higher than 80%, the player will pass the sublevel. If the real-time success rate is lower than 80% after 30 trials, the player will be judged

Table 8.4 Game levels design. There are 27 sublevels and 9 game levels. In the table, GLevel represents the game level, and CLevel represents the complexity level.

Game Levels	Sublevels	Type Dimension	Input Dimension	Time Dimension	Quantity Dimension
GLevel 1	1-1	CLevel 1	CLevel 1	CLevel 1	CLevel 1
	1-2				CLevel 2
	1-3				CLevel 3
GLevel 2	2-1	CLevel 2	CLevel 1	CLevel 1	CLevel 1
	2-2				CLevel 2
	2-3				CLevel 3
GLevel 3	3-1	CLevel 2	CLevel 2	CLevel 1	CLevel 1
	3-2				CLevel 2
	3-3				CLevel 3
GLevel 4	4-1	CLevel 2	CLevel 2	CLevel 2	CLevel 1
	4-2				CLevel 2
	4-3				CLevel 3
GLevel 5	5-1	CLevel 3	CLevel 2	CLevel 2	CLevel 1
	5-2				CLevel 2
	5-3				CLevel 3
GLevel 6	6-1	CLevel 3	CLevel 3	CLevel 2	CLevel 1
	6-2				CLevel 2
	6-3				CLevel 3
GLevel 7	7-1	CLevel 3	CLevel 3	CLevel 3	CLevel 1
	7-2				CLevel 2
	7-3				CLevel 3
GLevel 8	8-1	CLevel 4	CLevel 3	CLevel 3	CLevel 1
	8-2				CLevel 2
	8-3				CLevel 3
GLevel 9	9-1	CLevel 4	CLevel 3	CLevel 3	CLevel 2
	9-2				CLevel 3
	9-3				CLevel 4

as failing to pass the sublevel.

We designed two game modes, DDA and MDA, for effect comparison. In DDA mode, players can only begin from the first game level and cannot select or change the game level during play. Instead, the DDA mechanism automatically changes the game level. In MDA mode, players can freely choose any level to play, whether from the beginning or at any time of playing. After passing the level, they are also required to choose the level for play again. In both modes, players can pause the game at any time to rest or quit, but only in the MDA mode, they can change the game level. In addition, in the formal training process, players are required to continue on the game level and the sublevel of the previous training.

8.5 Experiment

To test our CTVR exergame, we conducted a two-month training experiment. We adopted a two-factor mixed design; the between-subjects variable is the game difficulty mode (DDA or MDA), the within-subjects variable is time, and the dependent variables are the cognitive abilities and PX of participants. The goal of the experiment is to test whether: (1) the game improves the cognitive ability of the elderly, (2) the participants have a good PX after playing, and (3) DDA has a better effect on these two aspects than MDA.

8.5.1 Participants

The participants recruited for this study were healthy elderly people from Fujian, China. A total of 24 people were recruited, including 9 males and 15 females. The ages of participants ranged from 62-79 years old (M = 69.81, SD = 4.69). Participants with underlying medical conditions, depression, and anxiety were excluded from the test. Moreover, to ensure consistency with formal cognitive training, we only selected elderly people with healthy cognitive function status. The same as in the game pretest, MMSE, SAS, and SDS were used to test their cognitive and health status, and the same screening standard was employed. Participants who passed the screening all signed the informed consent form and fulfilled the form of demographic information collection. Participants fully understood that the experiment's period was two months and the purpose was for cognitive training. Their cognitive abilities were measured as the baseline before the training. All 24 participants underwent the 2-month game training; after finishing the training, they were measured for their cognitive abilities as a post-test and filled out the PX questionnaire.

8.5.2 Materials and Apparatus

The following cognitive measuring tools were used in the pretest and post-test.

(1) Simple Reaction Time Test: Measuring the subject's simple reaction time. This

test requires subjects to respond to stimuli as quickly as possible, and it is widely used in various types of cognitive training to measure the ability of reaction (Ordnung et al., 2017; Maillot et al., 2012).

(2) Number Comparison Test: Measuring the subject's number comparison reaction time (Ackerman and Cianciolo, 2000). This test requires subjects to choose one of two numbers as quickly as possible. It can be used to calculate subjects' processing speed in conjunction with the results of simple response times (Maillot et al., 2012).

(3) Spatial Span Task: Measuring the subject's visuospatial working memory (Smyth and Scholey, 1992). This test requires subjects to remember blocks that appear in sequence on a screen and can be used to test the working memory span of the subject (Wiechmann et al., 2010).

(4) Attention Network Task (ANT): Measuring the subject's attention. This task asks the subject to respond to arrows appearing in different areas. In this study, we mainly focused on subjects' no-cue attention responses. This test is often used in measuring the effects of cognitive training (Schoene et al., 2015; Ordnung et al., 2017).

(5) Go No-Go: Measuring the subject's inhibitory control ability of executive function. The subject needs to restrain himself from making wrong choices. We will judge the subject's ability through the accuracy rate. This test is also widely used to measure the cognitive training effect of subjects (Schättin et al., 2016; Li et al., 2020).

(6) Raven's Standard Progressive Matrices (SPM): Measuring the subject's reasoning ability. The test will present different patterns to the subject and require them to choose an image to match the logical pattern. It is widely used to measure problem-solving and reasoning abilities (Raven and Court, 1998; Shute et al., 2015; Li et al., 2020).

The game in this experiment also ran on a PC equipped with Windows 10 system. The computer's CPU is a 3.6 GHz AMD 3700x processor. The VR device uses the HTC VIVE Cosmos VR kit. The experiment was conducted in the same room as our game pretest. After finishing the 2-month training, participants must fill out a 51-item PX questionnaire (see Appendix C). Each item in the questionnaire is scored on a 7-point scale, ranging from 1 (strongly disagree) to 7 (strongly agree). We combined the following three scales to create the questionnaire.

(1) Player Experience of Need Satisfaction (PENS). This study used three subscales: Autonomy, Competence, and Intuitive control, with 9 items in total (Rigby and Ryan, 2007).

(2) Intrinsic Motivation Inventory (IMI). The study used three subscales: Interest/Enjoyment, Effort/Importance, and Value/Usefulness, with 19 items in total. We replaced the "activity" on the original scale with "games" to make it more suiTable (Ryan et al., 2006).

(3) Instrument for Measuring Flow Experience in Computer Game Play. The study used all the six subscales in this scale: Challenging, Clear goal and feedback, Concentration, Control, Immersion, and Autotelic experience, with 23 items in total (Fang et al., 2013).

8.5.3 Procedure

The participants were randomly divided into two groups: DDA mode with 12 people ($M = 68.83$, $SD = 3.881$) and MDA mode with 12 people ($M = 67.00$, $SD = 4.936$). The participants were given sufficient time to practice in the guiding level before the formal training. After participants confirmed that they had learned the basic gameplay, they could start at any time. Each subject completed 24 times of 0.5 hours each, for a total of 12 hours of training. The interval between each training is 2-3 days to avoid fatigue and consolidate training effects. If participants feel tired during the training, they can rest freely. However, participants still needed to complete a 0.5-hour training session each time. Experiment workers stay closely to ensure the safety of the elderly participants while playing. Before the formal training, each participant underwent the cognitive ability test. After the 2-month training, all the participants took the cognitive test again and filled out the PX questionnaire.

8.5.4 Results

Cognitive Abilities We conducted a repeated-measures ANOVA. The results showed that the abilities of no-cue attention ($F(1, 11) = 36.667$, $p < 0.001$), working memory ($F(1, 11) = 25.869$, $p < 0.001$), processing speed ($F(1, 11) = 4.879$, $p = 0.049$), simple reaction time ($F(1, 11) = 6.738$, $p = 0.025$), and numerical comparison reaction time ($F(1, 11) = 7.528$, $p = 0.019$) were significantly improved after training. However, there was no significant improvement in the abilities of inhibitory control ($p = 0.065$) and reasoning ($p = 0.834$). There were no differences between groups and no significant interaction effects. In summary, the participants' most of cognitive abilities were improved in both groups. See Table 8.5.

Table 8.5 The results of repeated-measures ANOVA of the cognitive abilities in the variables of time and group (N =24). Notes: *Denotes significant at the 0.05 level (2-tailed); **Denotes significant at the 0.01 levels (2-tailed); ***Denotes significant at the 0.001 levels (2-tailed).

Cognitive	Time				Group		Time*Group	
Abilities	M	SD	F	p	F	p	F	p
No-Cue	Before, 784.69	95.23	36.667***	<0.001	0.006	0.939	0.000	0.998
Attention	After, 725.40	69.36						
Working	Before, 3.28	0.64	25.869***	<0.001	2.881	0.118	0.705	0.419
Memory	After,3.80	0.79						
Processing	Before, 341.79	98.10	4.879*	0.049	0.448	0.517	1.997	0.185
Speed	After, 302.05	48.57						
Simple	Before, 312.96	33.24	6.738*	0.025	0.387	0.547	1.295	0.279
Reaction Time	After, 301.65	24.67						
Number	Before, 654.75	110.75	7.528*	0.019	0.581	0.462	2.459	0.145
Comparing Time	After, 603.71	62.08						
Inhibitory	Before, 64.60	56.02	3.936	0.073	0.084	0.778	2.930	0.115
Control	After, 43.21	45.44						
Reasoning	Before, 35.63	10.30	0.041	0.844	0.352	0.565	1.032	0.331
Ability	After, 35.25	8.34						

Player Experience We conducted an independent samples t-test to compare the after-training PX scores between the two groups. The results showed a significant difference between the groups (t = 2.976, p = 0.007), and the DDA group (M = 6.27, SD = 0.46) had a higher PX than the MDA group (M = 5.75, SD = 0.40). In short, participants in the DDA group provided a higher rating of their experience of game training.

8.6 Discussion

In this chapter, we design a CTVR exergame with a DDA mechanism to enhance the cognitive ability of the elderly. The DDA design methodology proposed in Chapter 7 was utilized, and the results of the experiment showed good effects in improving the elderly's cognitive abilities and PX by the designed game. We further discussed how the two difficulty mechanisms affect the improvement of cognitive abilities and PX and the functions of our design methodology in DDA design.

8.6.1 Difficulty Modes and Cognitive Abilities

Our results showed that most elderly people's cognitive abilities improved in both MDA and DDA groups. Therefore, our game design, which combines the elements of multitasking, exercise, and VR, is successful. However, there was no difference between the two groups. Combining our observation of the experiment, we found that participants tend to find the optimal challenges for themselves. In more detail, if they could not overcome a level, they would keep trying on this level. In addition, the cognitive abilities of both groups of subjects improved in four aspects: working memory, reaction ability, processing speed, and attention, which is consistent with the findings of current game training research (Guimarães et al., 2018; Li et al., 2020). However, no significant differences were found in inhibitory control and reasoning abilities. This may be because there are no specialized elements to train these two abilities in the game. This also shows that training through games is limited to realize the transfer effect on ability enhancement, which is also consistent with other current research results (Toril et al., 2016). Therefore, specific designs of the game task are necessary to train corresponding abilities.

8.6.2 Difficulty Modes and Player Experience

Our results showed that the game with the DDA mechanism provides a better PX than that with the MDA. These results are different from previous studies (Smeddinck et al., 2016; Salehzadeh Niksirat et al., 2017), which found MDA was equal to or even better than DDA in PX. We believe it is because we adopted the proposed DDA design methodology in the game design. Based on our observation, some participants chose the level with too high OGD and could not pass for a long time. However, they did not re-select an easier level but kept trying, which may compromise their self-efficacy and cause frustration. In contrast, participants in DDA mode have not many choices, but they seem immersed in playing more. Considering if the training game has a poor experience that may affect players' motivation and engagement (Li et al., 2020), we found applying the DDA mechanism is promising in the cognitive training game design. However, the prerequisite to ensure the effect is still excellent DDA design. Therefore, the DDA design methodology we proposed in Chapter 7 is valuable support for the design of this type of game and even most serious games.

8.6.3 Our Design Methodology and DDA design

We proposed the DDA design methodology in Chapter 7 including a framework and a 6-step design process. The results of the case study indicated that our design methodology well guided the DDA design in achieving the serious goal and entertainment goal. It should be noted that the design methodology merely provides limited information on how to use different knowledge bases and theories in the design. Therefore, it is more a reminder of how to design DDA step by step, but not a detailed template to guide the design. Applying the methodology still requires designers to master adequate skills in game design and knowledge of the destined field.

8.7 Conclusion

This study designed a cognitive training VR exercise with a DDA mechanism to enhance the cognitive ability of the elderly. Based on the DDA design methodology proposed in Chapter 7, we designed two game difficulty mechanisms for effect comparison. The experiment showed that the designed DDA had a similar effect in improving the elderly's cognitive abilities with MDA but better improved the player experience.

Chapter 9

General Discussion

In this chapter, we discuss the implications and applications of our findings in the previous chapters for the concepts, measurement, and design of game difficulty in video games.

9.1 Concepts of Game Difficulty

We introduced the development of game difficulty and related concepts in Chapter 2. Based on the work in this dissertation, we further enriched the connotation of the game difficulty concepts. Firstly, we redefined OGD as **"during the interaction process between players and game tasks, the dynamic meeting of the player's skill to the game task demand"**, and redefined SGD as **"the player's subjective evaluation of game difficulty based on their perceptions of the game task, game-playing, and themselves."** in Chapter 3. These two definitions adopt the interaction perspective and describe these two concepts of their dynamic characteristics more precisely. In Chapter 4, we found that by the present measuring method, OGD and SGD only partially match each other. We argue that the reason is their structure differences both in concepts and measuring methods. More specifically, OGD is more narrowly about players' skill and task demands. However, the SGD involves more dimensions when it is evaluated by players. Therefore, we discuss the two aspects of game difficulty in the following subsections: (1) what the dynamic means for the game difficulty concept, and (2) what the ideal relationship between OGD and SGD is.

9.1.1 How to Understand Dynamic in the Game Difficulty Concept

We have pointed out that game difficulty is an interaction-related concept. Due to the dynamic of interaction, it is natural that game difficulty is correspondingly dynamic. For OGD, the dynamic means the changes in the task's demand on players. From a narrow sense, OGD occurs in the specific interaction process, the changes are more related to this dynamic process, i.e., changing with each input and output. From a broad sense, the interaction can be regarded as a long-term relationship (Hornbæk and Oulasvirta, 2017), and the changes are because of the skill improvement of players and differences in each playing attempt. For SGD, the perception of players is also dynamic in three aspects: (1) no matter short-term or long-term, the perception process of a player is dynamic, (2) the contents the player perceives are various and dynamic (e.g., OGD), and (3) players themselves are dynamic, with different and developing evaluation patterns of game difficulty.

Therefore, with such a dynamic in the game difficulty concept, designing them accordingly seems very difficult. In this case, precise measuring methods are crucial to game difficulty design. Therefore, our proposed OGD and SGD measuring methods can support better measurement. In addition, as we mentioned in Chapter 7, the DDA mechanism is a valuable approach to adapt to the dynamic process of game difficulty in the interaction. Another approach is to determine the more predictable factors (including static factors) in the dynamic process. For example, the complexity of the game task is usually static. However, to make players' perception of this complexity more predictable, clear presentation and feedback are necessary, e.g., the countdown was added in the game case in Chapter 8. Another example is that the average simple reaction time of humans is about 0.3 seconds (Wilkinson and Allison, 1989); therefore, designing task demands for reaction time nearly to this value can better predict the OGD and SGD.

9.1.2 The Ideal Relationship between OGD and SGD

Chapter 4 has pointed out that OGD and SGD match partially if we use the current concept scopes, which is because their concepts have different structures. We also briefly discussed the the goal of OGD measuring in Chapter 5. We here further discuss that whether we should expect an ideal matching relationship between them. As also indicated in Chapter 4, changing their concepts' scope accordingly can result in a better match. In more detail, if SGD narrows to only relate to players' perception of their competence, or if OGD expands to contain the dimensions of game complexity and completion difficulty, they may match better.

However, we suggest pursuing this ideal relationship between OGD and SGD, which seems to lack necessity. There are three reasons. Firstly, OGD and SGD are both dynamic. Therefore, they can not always keep a matching relationship. Instead, partial matching may be a more common situation. Secondly, assuming they match each other is convenient for game design. However, after understanding these two concepts fully, designers can also design them separately to achieve a good game design. For designers, it is not critical whether they match in the playing process, but what they are and how to push them to design expectations. Finally, considering the connotations of these two concepts, we think it is reasonable for their partial match. Human perception is subjective and depends on the processing of information in cognition. Considering the information is from the dynamic interaction process and the possibility of misinterpretation and misunderstanding of the information, we can not expect this perception process to objectively and precisely reflect reality.

In conclusion, we suggest that an ideal relationship between OGD and SGD is not a complete match. Instead, we need to respect their nature and acknowledge that there will be a partial match between OGD and SGD in the interaction of players and games.

9.2 Measurement of Game Difficulty

Based on the understanding of the game difficulty concepts, OGD and SGD are both human-related, and the measurement of these two concepts should depend on measurable psychological and physical variables. However, as introduced in Chapter 2, the clarification of the gap between concepts and measurement is lacking, and the present measuring methods of OGD and SGD have flaws and need to be improved. Therefore, we first proposed and validated a new OGD measuring method in Chapter 5. This method includes an operational definition of OGD, **an integral ratio of the player's input incorrectness to the game task's required input correctness within a given time frame**, and some quantifying formulas. In Chapter 6, we developed and tested a new self-report scale for SGD measuring based on the six SGD dimensions summarized in Chapter 2. These two studies are valuable in supporting the measurement of game difficulty in game research and game design.

It should be noted that our measuring methods aim to measure one player's difficulty in both single-player and multiplayer games. Therefore, applying these two methods depends on the measurement needs. We further illustrate how to conduct the measurement of OGD and SGD in the three main situations: (1) validating the difficulty design, (2) conducting the difficulty research, and (3) adjusting the difficulty during players' playing.

To validate the game difficulty design, designers need to ensure the OGD and SGD are consistent with their design goals. Therefore, OGD and SGD should be measured accordingly in the validation process and after the process. In more detail, the real-time OGD in the process should be investigated to validate the relationship between OGD and the game task complexity. The OGD and SGD measured after the process are more related to the core design goal, e.g., towards the entertainment goal, whether a good player experience is created.

For game research, how to conduct OGD and SGD measurements relies more on the research goal. For example, if researchers aim to investigate how game difficulty affects players' emotional state (Bontchev, 2016), it is necessary to first divide the state into subjective emotional state (measured by self-report; Schaefer et al., 2010) and objective emotional state (measured by physiological indicators; Caroux et al., 2015). In this case, OGD measurement can also be conducted during game playing to correspond to the objective emotional state measurement. SGD should be measured together with the subjective emotional state. Therefore, by comparing these two difficulties with the two types of emotional states, researchers are able to get valuable and comprehensive findings.

To adjust the difficulty in real time, it is necessary to determine the game's difficulty during play. As we mentioned in Chapter 7, the DDA mechanism must determine the OGD and SGD before adjusting them. Therefore, precise measuring of these two difficulties is fundamental for DDA to conduct the adjustment, and to design the DDA should add the OGD and SGD measurement in the player evaluation mechanism of DDA. Such a measuring method can be regarded as a pre-set but automatically conducting measurement.

9.3 Implications for Game Difficulty Design

In Chapter 7, we elaborated how to design a DDA mechanism based on the concepts and measurements of OGD and SGD, and other critical player factors. We proposed a new DDA definition, **"a game difficulty control mechanism that aims to control the difficulty automatically in game interaction by evaluating objective and subjective game difficulty data and modifying game tasks"**; and a new DDA design methodology, including a DDA design framework and a 6-step design process.

We further validated this new design methodology by a case study of designing the cognitive training game in Chapter 8. However, it is still necessary to discuss how to design OGD and SGD based on our findings.

Designers should consider designing OGD and SGD separately rather than the game difficulty in general. For example, as a useful design tool, difficulty curves represent how difficulty changes as the game progresses (Adrian and Luisa, 2013; Adams, 2014; Nagle et al., 2016; Aponte et al., 2011b). Based on our findings, there should be two separate curves of OGD and SGD, one for the change in OGD and another for the player's subjective experience of difficulty. Secondly, determine the effect of OGD on SGD. Due to OGD and SGD matches partially, it is necessary to determine whether the designed OGD can produce the expected SGD. Finally, avoid extremely low OGD (players can always succeed) or extremely high SGD (players regard success as impossible) in design. According to our findings, these two situations are considered negative to the player experience and engagement.

Designers also should design OGD and SGD towards specific design goals accordingly. For the goals of player experience, SGD plays a more important role in affecting players' playing processes. This finding inspires designers to focus more on players' subjective difficult feelings rather than only designing the failure rate. For example, it may be useful to remind players of their effort, luck, and skill enhancement in the process of creating a positive difficult experience. For player engagement, meaningful failures (Juul, 2009) are required to promote their future replay. In more detail, designers need to realize players do not hate failures, but failures without any value, e.g., failures due to game bugs, even though not a design issue, are very frustrating (Miller and Mandryk, 2016) and will stop players' replay. While failures that reveal future possibility of success may be popular by players. For player self-efficacy, this factor is an important design consideration in some serious games, such as cognitive training games for the elderly (Anguera et al., 2013; Khalili-Mahani et al., 2020). We found that there was no correlation between self-efficacy and player experience, but self-efficacy is negatively related to engagement. Therefore, we recommend player engagement design (Xue et al., 2017; Nuutila et al., 2021) should consider creating a good balance between self-efficacy and engagement motivation by difficulty design.

9.3.1 Implications for OGD Design

Our method of measuring the real-time OGD and the overall OGD is practical for application in the design and research. Generally speaking, the evaluation of real-time OGD can support designers in adjusting the OGD according to changes as the game proceeds, while the overall OGD evaluation is important for researchers to confirm the relationship between OGD and other factors.

As a useful difficulty design tool, the difficulty curve reflects how game difficulty changes as the game progresses and is usually used for game level difficulty design (Adrian and Luisa, 2013; Adams, 2014; Nagle et al., 2016; Aponte et al., 2011b). Changing difficulty during the playing process is real-time OGD. Usually, designers preset a difficulty curve to design the content of the game level and to evaluate their design by measuring real-time OGD. Current methods include relying on designer experience (Larsen, 2010), estimating the difficulty level by algorithm (Adrian and Luisa, 2013), calculating the failure rate of certain nodes of the level by player testing, etc. These methods are not effective for evaluation or they are costly. Our method is valuable for measuring the real-time OGD to support level design. Designers can use our method to get the real-time OGD values to draw the difficulty curve, which can be based on a single player or multiple players. The curve based on a single player is suitable for quick design validation in the early stages. The difficulty curve based on multiple players can be used as the average difficulty curve of this level. However, it should be noted that difficulty curves may differ across different groups of players, and they should be drawn based on specific groups (e.g., beginner players).

Our OGD measuring method can also support personalized difficulty design and Dynamic Difficulty Adjustment (DDA) mechanism design. Personalization refers to providing game content adapted to the specific needs of individual players (Karpinskyj et al., 2014). Due to the differences between players, researchers still attempt to develop more effective techniques to provide appropriate game content (Streicher and Smeddinck, 2016; Orji et al., 2017; Zhu and Ontañón, 2020; Kristensen et al., 2022). Difficulty personalization is meant to provide game content with appropriate difficulty based on the needs of different players. However, players usually have a bias in their game skills, which makes providing the most appropriate difficulty level a challenge (Constant and Levieux, 2019; Huang et al., 2024). Our OGD method can realize personalization more quickly. In addition, our proposed learning function is also helpful to better realize difficulty personalization by quantifying players' learning process.

Different from focusing on providing personalized difficulty content, DDA focuses more on subtle adjustments to difficulty in real-time (Guo et al., 2024). Currently, adjustments in some DDA mechanisms are based on evaluating the player's failure rate (Demediuk et al., 2017; He et al., 2010; Anguera et al., 2013). This makes the adjustment by the mechanism inflexible. Our method supports DDA in its real-time evaluation of OGD. With the introduction of our OGD method, DDA works even on the first play. Specifically, game designers can design to achieve the ideal OGD curve of a game level and apply DDA to guide players to follow this ideal curve by real-time OGD evaluation and adjustment (Guo et al., 2024). We also suggest this adjustment can be realized by modifying the real-time error tolerance $e(t)$. For example, in a shooting game, by slightly adjusting the size of the target area and the critical hit rate of the weapon, real-time OGD will be accordingly changed with the changes in error tolerance. In summary, through our method, DDA can be more precise and can adapt more quickly to different players while achieving the expected design goals.

9.3.2 Implications for SGD Design

Effective SGD evaluation is required for validating SGD design. Currently, SGD can be measured by direct difficulty rating or subjective scales (Miller and Mandryk, 2016; Cox et al., 2012; Wheat et al., 2016; Wehbe et al., 2017; Demediuk et al., 2019; Denisova et al., 2020; Rigby and Ryan, 2007). However, the difficulty rating is too general and lacks details, and the scales fail to consider the multidimensional structure of SGD but only involves the competence dimension (Rigby and Ryan, 2007; Ryan et al., 2006).

Therefore, our proposed six-dimensional SGD measuring tool (Game Complexity, Game Completion Difficulty, Game-playing Difficulty, Player Competence, Player Pressure, and Player Effort) is a preliminary but more comprehensive SGD design evaluation tool. In addition, these dimensions also represent the accordingly subjective difficulty that can be designed. Referring to the example we mentioned in Chapter 7, the serious game *My Cotton Picking Life* (Rawlings, 2012) demands players to pick cotton by repeating click operations constantly for 6 hours, and such a torturous process strengthened persuasive effects (i.e., affects players' attitudes to child slave labor; Jacobs et al., 2020). This example is to make players feel a high level of SGD by only designing the effort dimension. In short, we suggest designers employ these dimensions in their SGD design and validation to support a better SGD design.

Currently, simply considering OGD for adjusting in DDA design will no longer be reliable because the skill-challenge balance proposed by flow theory is related to SGD. Designers need to think more about how to design a proper balance by influencing SGD. Some researchers have noticed this point and adopted SGD in their DDA design (Chanel et al., 2011; Wheat et al., 2016; Frommel et al., 2018). Furthermore, the partial matching relationship between OGD and SGD allows designers to adjust the OGD finely without influencing SGD. Results of some DDA research supported this view and implied it is possible to adjust OGD but keep SGD stable (Denisova and Cairns, 2015; Khajah et al., 2016). Designers can also guide players to a benign mismatch between OGD and SGD because players will not perceive the game as difficult merely based on a failure. For example, by reducing the frustration of player failure and the difficulty of completion, players may evaluate the game's failure as more acceptable and decrease the abandonment.

9.3.3 HEC and Game Difficulty Design

HEC theory provides the possibility to reshape the values of games (Ren et al., 2019). Most games are designed for players to experience, and this experience is generally positive. It is widely agreed that the pursuit of relaxation and pleasure is the main motivation of players (Malaby, 2007). Game designers mostly focus on "how to create games that can bring enjoyable experiences to players" (Bernhaupt, 2010). In short, the value of the game is strongly bound to the pleasure. However, we found the value of games diverse due to the development of games. The seriousness in games becomes increasingly important. Researchers and designers have recognized the serious functions of games. For example, games are also used as models in economic gaming theory or designed for commercial goals by gamification (Krath et al., 2021). However, it seems hard for players to accept the seriousness of games currently.

Based on HEC theory, players' perceptions of games can be shaped by synergized interaction, which can be deliberately created through design. Dewey (2018) emphasized the significant role of experience in shaping cognition. Therefore, when players engage in gameplay, the interactive experience itself can have a profound impact, surpassing the initial purpose, motivation, and pre-existing perceptions. We argue that it is the enjoyable game interaction itself that gradually establishes in players the perception that games are inherently pleasurable activities. Consequently, we believe that involving seriousness in players' experience activity is essential to improve the current situation.

Games designed based on the HEC theory, which can enhance human abilities and create positive mindfulness, are promising to reshape the players' perception of games and their values.

This dissertation provides valuable implications for game difficulty design to realize the HEC game design. As we introduced in Chapter 1, one of the goals of HEC theory is to develop engaging computers. The theme of game difficulty in this dissertation plays an important role in designing engaging games. SGD, as part of experiences, supports deep engagement in playing the game, while OGD affects the player's capability in their voluntary learning process. We have introduced how our research provides the design implications for these two difficulties. Furthermore, we have provided a case to show how DDA mechanisms support better cognitive ability enhancement for elderly people. These are all highly related to the vision of HEC.

Furthermore, DDA is a promising mechanism for realizing the synergized interaction in games. We have introduced the design methodology of DDA and emphasized how to design it based on the design goal. Toward the goal of synergized interaction, DDA should evaluate players' engagement states in real-time to adjust the game difficulty. Therefore, the evaluation mechanism of DDA should make corresponding criteria to determine the engagement state of players, and how the OGD and SGD should be adjusted accordingly.

In summary, HEC provides a new perspective to inspire the seriousness of games through synergized interaction. Our work on game difficulty design can support the realization of HEC visions.

Chapter 10

Conclusion

This chapter provides the limitations and future work, concludes our work by chapter, and illustrates the contributions of this research.

10.1 Limitations & Future Work

This dissertation investigated the theory, measurement, and design of game difficulty and provided valuable definitions, measuring methods and a validated DDA design framework. We further discuss the limitations and future directions in this section.

For the study exploring the relationship between SGD and OGD, we chose the failure rate and operationalized it as dichotomous to measure OGD. Although it was a commonly used player performance factor, it may not totally and exactly represent the connotation of OGD. Due to the uncontrollable game results, we could not fully control the game's complexity factors. Therefore, our comparison was limited to assessing the OGD and SGD of participants under two varying game levels with different complexities. Additionally, our study examined casual games with a simple form, and the relationship between OGD and SGD in other more complex game types could require further investigation. We also controlled the skill factor in this study, which we believed could be explored further after the development of better OGD measuring tools.

For the study of the OGD measuring method, the games used in our experiment were relatively simple. Further validation of this method in more complex games, e.g., serious games and popular commercial games, is necessary. Due to the lack of standard methods, the present validation of real-time OGD measuring validity is not adequate. We plan to conduct design case studies that apply our method to provide practical and empirical support for our method. In addition, this method is well applied in measuring the game in which the player can take a certain ideal action, but how to apply it to a complicated game with many optimal actions should be studied in the future.

For the study of the SGD measuring method, data from 326 participants is still limited. Therefore, we plan to collect more data from the players responding in the three languages. International comparisons on this issue, considering the differences in culture and personality among countries, have become a new challenge for future research. Further validation of this instrument is expected to solve the issues of minor invariance between the three language versions of the scale. In addition, broader validation in more genres of games is still necessary because this study merely used a casual game to test the scale. In the CORGIS, the type of emotional challenge is included in the measurement. We plan to clarify this concept in the future to to support the improvement of our SGD scale. In short, the SGDS developed in this work still needs iteration and improvement to enhance its quality.

Regarding the study of DDA and its design case, we design a serious game for the DDA design methodology validation. However, further research is necessary to validate the DDA design in entertainment games. We adopted a discrete difficulty design based on the four complexity dimensions and presented them as game levels. However, more validation should be conducted on more complex and continuous types of difficulty in DDA design. In addition, there is no specific VR-targeted difficulty design in this case.

We further proposed the following four research directions:

(1) Research on game difficulty concept. We proposed the new definitions of OGD and SGD. Further empirical and design research is necessary to validate and utilize these definitions in more types of games. In addition, further research is also required to deepen the understanding of the interaction perspective for game difficulty. The dynamic of OGD and SGD needs more elaboration and validation. The findings in these research directions will support future improvements in the OGD and SGD definitions.

(2) Research on game difficulty measurement. Future work is necessary to validate our proposed method to measure OGD and improve it for different research requirements. Therefore, based on our method, the development of more specific OGD measuring methods for different game genres is promising. Our SGD scale is still in its initial state and needs iteration and improvement. We plan to develop more related items to expand the current version of the scale. It is also promising to apply this scale with other SGD measurements to conduct cross-validation. Additionally, this scale can be used to support the future development of physiological measurements. This work can also be a reference for task difficulty measurement in other fields, e.g., HCI, and psychology.

(3) Research on game difficulty design. Task and player interaction determine game

difficulty, but how task complexity elements (e.g., goals, input, process, etc) affect game difficulty remains unclear. Systematic work is still needed on how these task elements can be modified during the game process to influence OGD and SGD effectively. In addition, empirical design research is still necessary to provide more evidence of the relationship between OGD and SGD and apply this relationship to game design. Future work needs more design practices on this topic, which can offer valuable practical perspectives and facts of game difficulty design.

(4) Research on DDA design. For DDA design, entertainment and serious goals can be further refined into various specific design goals. Currently, except for Flow, which has specific DDA evaluation criteria (i.e., whether it is too difficult or too easy), there is a lack of other evaluation criteria for specific goals of entertainment (i.e., relaxation but not boring for the casual goal). Additionally, although there are DDA design case studies toward different serious characterizing goals, providing generic evaluation criteria fit for specific categories of serious goals (e.g., education, persuasion, etc.) is essential. Research on adjusting subjective game difficulty directly: There is still little research on adjusting SGD directly, but some researchers have noticed this issue (Zhang, 2021). These directions are promising to improve the DDA design practically. Further research is expected on adaptive DDA, which can dynamically change the DDA mechanism to achieve optimal or synergistic mapping between SGD and OGD by applying HEC synergistic interaction theory to DDA design.

10.2 Conclusion

Game difficulty is an essential component in video game design and is significant to the realization of HEC's vision to develop engaging computers. However, we found three challenges in current game difficulty research: there is a lack of (1) clarification of the relationship between objective game difficulty (OGD) and subjective game difficulty (SGD), (2) effective measuring methods of OGD and SGD, and (3) sufficient design research on game difficulty and DDA based on understanding of OGD and SGD.

Therefore, this dissertation aims to provide solutions to these three challenges in concept, measurement, and design. To reframe game difficulty, a player-game interaction perspective was adopted, and theoretical (Chapters 3 and 7), exploratory (Chapter 4), quantifying (Chapters 5-6), and empirical (Chapter 8) studies were accordingly conducted. We proposed new definitions of SGD and OGD and an interpretive interaction model. The partial matching relationship between SGD and OGD was determined, and the effective measuring methods of them were provided. We also proposed and validated a new Dynamic Difficulty Adjustment (DDA) definition and the design methodology. Further insights and implications to game difficulty were finally discussed.

We summarize our findings by chapter as follows.

Chapter 2 first introduced how the concepts of game difficulty are divided into OGD and SGD. However, clear definitions of them from an interaction perspective were lacking. We also found that the relationship between OGD and SGD needs to be clarified. The current measuring methods for these two difficulties also need to be improved. We further introduced the multidimensional structure of SGD and indicated this summary could be the basis for developing a new SGD measuring method. For game difficulty design, we introduced the current research on how game difficulty impacts players, but we found the separate impacts of OGD and SGD need to be clarified. Research on the DDA mechanism was then introduced, but we found rethinking the theoretical fundamentals of DDA is urgent for design.

Chapter 3 first introduced the three components of player-game interaction, then built a model to illustrate how OGD and SGD occur in the game-playing process. We further redefined OGD and SGD based on this interaction perspective.

Chapter 4 explored the relationship between SGD and OGD by experiment. We found that OGD and SGD only partially match each other, and we argue that the reason is their structure differences. Our findings support that SGD mediates the OGD's effect on player experience, engagement, and self-efficacy and indicate that SGD has an indispensable role in influencing players. These findings provide empirical support to the partial matching relationship between OGD and SGD and valuable insights into game difficulty design.

Chapter 5 provided a new OGD measuring method with quantifying factors, operational defintion, and computational formulas. We first provided a conceptual definition of OGD based on a comprehensive literature review. Second, we identified seven basic game tasks by a game investigation and determined that, for each task, OGD can be quantified through two factors: input incorrectness and input time. Finally, we proposed an operational definition and corresponding computational formulas of OGD as the measuring method. To validate the proposed method, a game incorporating seven basic task types was developed for experiment. We compared our method with the failure rate and incompletion rate methods in the experiment. The results showed that our method is effective in measuring OGD and has better validity than the other two methods.

Chapter 6 developed and validated an SGD scale with six dimensions of game complexity, game completion difficulty, game-playing difficulty, player competence, player pressure, and player effort. We adopted the three stages of item generation, scale development, and scale testing to complete this scale. The results showed that our final version of the SGD scale with 33 items had good reliability and validity, and thus promising in the future measurement of SGD.

Chapter 7 rethinked the fundamentals of DDA mechanisms to improve its design theory. We have addressed the four crucial questions regarding DDA's issues, definition & scope, value, and design through a literature review and discussion. This rethinking offers new insights into DDA and its design: DDA should not depend on Flow theory but should be defined based on game difficulty and it should be designed toward specific design goals. We further proposed a new DDA design methodology, including a design framework and a 6-step design process. This work is promising to improve DDA through theoretical exploration.

Chapter 8 designed a cognitive training VR exergame with a DDA mechanism to enhance the cognitive ability of the elderly. The DDA design methodology proposed in Chapter 7 was utilized, and the game with DDA showed good effects in improving the elderly's cognitive abilities and player experience. This case study validated the effectiveness of our DDA design methodology and provided valuable insights into game difficulty design.

The main contributions of this dissertation are three-fold: (1) Enhancing the theoretical fundamentals of game difficulty by clarifying the concepts' connotations and relationships. (2) Clarifying the relationship between concept to measurement and proposing effective measuring methods for game difficulty. (3) Exploring the game difficulty's impacts on players and rethinking the DDA mechanism to provide practical design methodology and implications for game difficulty to support the HEC game design. The other specific contributions of this dissertation include: (i) Proposing new definitions of subjective game difficulty (SGD) and objective game difficulty (OGD) and an interaction model to interpret these two difficulties. (ii) Determining the partial matching relationship between SGD and OGD. (iii) Proposing a new Dynamic Difficulty Adjustment (DDA) definition and design methodology and validate them with a case study.

Acknowledgement

I would like to express my heartfelt gratitude to all those who have supported and contributed to the completion of this dissertation. First and foremost, I would like to extend my deepest appreciation to my supervisor, Prof. Xiangshi Ren, for his guidance, invaluable insights, and unwavering support throughout my PhD studying journey.

I would like to express my appreciation to my co-supervisors, Prof. Hiroki Nishino and Prof. Yukinobu Hoshino, for their guidance and support in enhancing the quality of this research. I am also grateful to the other two members of the dissertation committee, Prof. Shinichi Yoshida and Prof. Zixue Cheng, for their helpful comments, valuable feedback, and constructive criticism. Their scholarly support and help have immensely enriched my work. I would like to express my thanks to Prof. Xinxin Ma for her valuable guidance on the study in Chapter 6 and Prof. Ruck Thawonmas for his continuous support and guidance on the study in Chapter 7. I also would like to thank Mr. John Cahill, Prof. Huawei Tu, and Prof. Jeffrey Bardzell for their advice on this work.

My sincere thanks go to the staff and faculty of Kochi University of Technology, especially the International Relations Section, for providing resources, facilities, dormitory, and unreserved help to my research and living in Kami, Kochi. I am also thankful to the members of the Center for Human-Engaged Computing, especially Ms. Hatakenaka Kyoko, for their always kind help and support during my research process.

I would like to acknowledge the support and encouragement from my colleagues and fellow researchers, especially Peng Tan, Chen Wang, Yibin Jia, and Yujia Liu, who have shared their expertise, engaged in fruitful discussions, and provided technical support throughout this challenging process. Furthermore, I am grateful to the participants who willingly participated in my studies. Their valuable contributions and cooperation have significantly enriched the empirical findings of this research.

Finally, I would like to express my deepest gratitude to my family, especially my parents, my wife, and my son, for their unconditional love, encouragement, and belief in me. I also would like to express my sincerest gratitude to my friends, especially Tian Tian, Xiao Liu, Hongtao Zhang for their support during this hard process.

This dissertation would not have been possible without the collective efforts, support, and encouragement of all those mentioned above and many others. Thank you all again for your invaluable contributions to this exciting academic journey.

Publications

Many of the ideas and figures contained in this dissertation have been previously published in or are scheduled to submit to the following peer-reviewed publications:

1. Papers published in or submitted to academic journals

- **Guo, Z.**, Thawonmas, R., & Ren, X. (2024). Rethinking dynamic difficulty adjustment for video game design. *Entertainment Computing*, 100663. (**JCR Q2 Level, Chapter 7**)
- **Guo, Z.** & Ren, X. (2024). An Exploratory Study of the Relationship Between Objective Game Difficulty and Subjective Game Difficulty. This paper has been submitted to the academic journal *International Journal of Human-Computer Studies* July 15, 2024 (**JCR Q1 Level, Chapter 4**)
- **Guo, Z.** & Ren, X. (2024). Measuring Objective Game Difficulty in Video Games: Concept, Operational Definition, and Computational Formulas. This paper has been submitted to the academic journal *ACM Transactions on Computer-Human Interaction* August 18, 2024. (**JCR Q1, Chapter 5**)
- **Guo, Z.**, Ren, X., & Ma, X. (2024). Measuring Perceived Difficulty in Video Games: Development & Validation of the Subjective Game Difficulty Scale (SGDS). This paper has been submitted to the academic journal *Entertainment Computing* June 27, 2024. (**JCR Q2, Chapter 6**)

2. Presentations in international conferences.

- Liu, Y., Wang, J., **Guo, Z.**, & Ren, X. (2023, September). Can We Understand Other Drivers? An Investigation on Empathic Triggers through Voice Information Transmission in Driving. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 42-47).

3. Presentations in domestic conferences.

- **Guo, Z.** & Ren, X. (2023). Rethinking the Definition of Dynamic Difficulty Adjustment in Video Games. *Shikoku-section Joint Convention of the Institutes of Electrical and related Engineers.*

- Jia, Y., **Guo, Z.**, & Ren, X. (2023). Influencing Perceived Difficulty in Video Games: The Role of Self-Efficacy as a Controlling Factor. *Shikoku-section Joint Convention of the Institutes of Electrical and related Engineers.*
- Kubo, T., **Guo, Z.**, & Ren, X. (2023). The Effects of Difficulty Adjustment on PX in VR games. *Shikoku-section Joint Convention of the Institutes of Electrical and related Engineers.*
- Shinki, G., **Guo, Z.**, & Ren, X. (2023). The impact of game control devices on user engagement. *Shikoku-section Joint Convention of the Institutes of Electrical and related Engineers.*

References

Ackerman, P. L. and Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6(4):259.

Adams, E. (2014). *Fundamentals of game design.* Pearson Education.

Adrian, D.-F. H. and Luisa, S.-G. C. A. (2013). An approach to level design using procedural content generation and difficulty curves. In *2013 IEEE Conference on Computational Inteligence in Games (CIG)*, pages 1–8. IEEE.

Akbar, M. T., Ilmi, M. N., Rumayar, I. V., Moniaga, J., Chen, T.-K., and Chowanda, A. (2019). Enhancing game experience with facial expression recognition as dynamic balancing. *Procedia Computer Science*, 157:388–395.

Alan, A. K., Kabadayi, E. T., and Aksoy, N. C. (2022). Replaying online games for flow experience and outcome expectations: An exploratory study for the moderating role of external locus of control based on turkish gamers' evaluations. *Entertainment Computing*, 40:100460.

Alexander, J. T., Sear, J., and Oikonomou, A. (2013). An investigation of the effects of game difficulty on player enjoyment. *Entertainment computing*, 4(1):53–62.

Allart, T., Levieux, G., Pierfitte, M., Guilloux, A., and Natkin, S. (2017). Difficulty influence on motivation over time in video games using survival analysis. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, pages 1–6.

Altimira, D., Mueller, F., Lee, G., Clarke, J., and Billinghurst, M. (2014). Towards understanding balancing in exertion games. In *Proceedings of the 11th conference on advances in computer entertainment technology*, pages 1–8.

Andrade, K. d. O., Pasqual, T. B., Caurin, G. A., and Crocomo, M. K. (2016). Dynamic difficulty adjustment with evolutionary algorithm in games for rehabilitation robotics. In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–8. IEEE.

Ang, D. and Mitchell, A. (2017). Comparing effects of dynamic difficulty adjustment systems on video game experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 317–327.

Ang, D. and Mitchell, A. (2019). Representation and frequency of player choice in player-oriented dynamic difficulty adjustment systems. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 589–600.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., et al. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501(7465):97–101.

Anguera, J. A. and Gazzaley, A. (2015). Video games, cognitive exercises, and the enhancement of cognitive abilities. *Current Opinion in Behavioral Sciences*, 4:160–165.

Aponte, M.-V., Levieux, G., and Natkin, S. (2011a). Difficulty in videogames: an experimental validation of a formal definition. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, pages 1–8.

Aponte, M.-V., Levieux, G., and Natkin, S. (2011b). Measuring the level of difficulty in single player video games. *Entertainment Computing*, 2(4):205–213.

Armstrong, R. A. (2014). When to use the b onferroni correction. *Ophthalmic and Physiological Optics*, 34(5):502–508.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, 59(1):36–47.

Bainbridge, K. and Mayer, R. E. (2018). Shining the light of research on lumosity. *Journal of Cognitive Enhancement*, 2:43–62.

Bakkes, S., Tan, C. T., and Pisan, Y. (2012). Personalised gaming: a motivation and overview of literature. In *Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System*, pages 1–10.

Bakkes, S., Whiteson, S., Li, G., Vişniuc, G. V., Charitos, E., Heijne, N., and Swellengrebel, A. (2014). Challenge balancing for personalised game spaces. In *2014 IEEE Games Media Entertainment*, pages 1–8. IEEE.

Baldwin, A., Johnson, D., and Wyeth, P. (2016). 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*, pages 326–337.

Baldwin, A., Johnson, D., Wyeth, P., and Sweetser, P. (2013). A framework of dynamic difficulty adjustment in competitive multiplayer video games. In *2013 IEEE international games innovation conference (IGIC)*, pages 16–19. IEEE.

Bandura, A., Freeman, W. H., and Lightsey, R. (1999). Self-efficacy: The exercise of control.

Bandura, A. and Wessels, S. (1994). *Self-efficacy*, volume 4. na.

Berger, V. W. and Zhou, Y. (2014). Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*.

Bernhaupt, R. (2010). *Evaluating user experience in games: Concepts and methods*. Springer Science & Business Media.

Bevana, N., Kirakowskib, J., and Maissela, J. (1991). What is usability. In *Proceedings of the 4th International Conference on HCI*, page 24. Citeseer.

Blandford, A. E. (2013). Semi-structured qualitative studies. Interaction Design Foundation.

Blom, P. M., Bakkes, S., and Spronck, P. (2019). Modeling and adjusting in-game difficulty based on facial expression analysis. *Entertainment Computing*, 31:100307.

Bontchev, B. (2016). Adaptation in affective video games: A literature review. *Cybernetics and Information Technologies*, 16(3):3–34.

Bostan, B. (2009). Player motivations: A psychological perspective. *Computers in Entertainment (CIE)*, 7(2):1–26.

Bostan, B. and Öğüt, S. (2009). In pursuit of optimal gaming experience: challenges and difficulty levels. In *Entertainment= Emotion, ed PVMT Soto. Communication présentée à l' Entertainment= Emotion Conference (Benasque: Centro de Ciencias de Benasque Pedro Pascual (CCBPP))*.

Bowman, N. D. (2018). *Video games: A medium that demands our attention*. Routledge.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Brown, M., Kehoe, A., Kirakowski, J., and Pitt, I. (2015). Beyond the gamepad: Hci and game controller design and evaluation. *Game user experience evaluation*, pages 263–285.

Buchanan, R. (1992). Wicked problems in design thinking. *Design issues*, 8(2):5–21.

Burke, J., McNeill, M., Charles, D., Morrow, P., Crosbie, J., and McDonough, S. (2010). Designing engaging, playable games for rehabilitation. In *Proceedings of the 8th international conference on disability, virtual reality & associated technologies*, pages 195–201.

Burke, J. W., McNeill, M., Charles, D. K., Morrow, P. J., Crosbie, J. H., and McDonough, S. M. (2009). Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, 25:1085–1099.

Burns, A. and Tulip, J. (2017). Detecting flow in games using facial expressions. In

2017 IEEE Conference on Computational Intelligence and Games (CIG), pages 45–52. IEEE.

Buttussi, F., Chittaro, L., Ranon, R., and Verona, A. (2007). Adaptation of graphics and gameplay in fitness games by exploiting motion and physiological sensors. In *International Symposium on Smart Graphics*, pages 85–96. Springer.

Byström, K. and Järvelin, K. (1995). Task complexity affects information seeking and use. *Information processing & management*, 31(2):191–213.

Caldwell, N. (2004). Theoretical frameworks for analysing turn-based computer strategy games. *Media International Australia*, 110(1):42–51.

Callies, S., Sola, N., Beaudry, E., and Basque, J. (2015). An empirical evaluation of a serious simulation game architecture for automatic adaptation. In *R. Munkvold & L. Kolas, Proceedings of the 9th European Conference on Games Based Learning (ECGBL 2015)*, pages 107–116.

Calvo, R. A. and Peters, D. (2014). *Positive computing: technology for wellbeing and human potential*. MIT press.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of management review*, 13(1):40–52.

Capcom (2016). Street fighter v. *Capcom Co., Ltd.*

Card, S. K. (2018). *The psychology of human-computer interaction*. Crc Press.

Caroux, L., Isbister, K., Le Bigot, L., and Vibert, N. (2015). Player–video game interaction: A systematic review of current concepts. *Computers in human behavior*, 48:366–381.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Number 1. Cambridge University Press.

Carter, M., Downs, J., Nansen, B., Harrop, M., and Gibbs, M. (2014). Paradigms of games research in hci: a review of 10 years of research at chi. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, pages 27–36.

Carvalho, C., Teran, L., Mota, M., and Pereira, R. (2022). A systematic mapping study on digital game adaptation dimensions. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*, pages 1–14.

Cavallini, E., Pagnin, A., and Vecchi, T. (2003). Aging and everyday memory: the beneficial effect of memory training. *Archives of gerontology and geriatrics*, 37(3):241–257.

Cechanowicz, J. E., Gutwin, C., Bateman, S., Mandryk, R., and Stavness, I. (2014). Improving player balancing in racing games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, pages 47–56.

Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2008). Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, pages 13–17.

Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1052–1063.

Charles, D. and Black, M. (2004). Dynamic player modeling: A framework for player-centered digital games. In *Proc. of the International Conference on Computer Games: Artificial Intelligence, Design and Education*, pages 29–35.

Charles, R. L. and Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics*, 74:221–232.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, 50(4):31–34.

Chen, S., Epps, J., Ruiz, N., and Chen, F. (2011). Eye activity as a measure of human mental effort in hci. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 315–318.

Clement, J. (2024). Video game industry - statistics and facts.

Colcombe, S. and Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychological science*, 14(2):125–130.

Constant, T. and Levieux, G. (2019). Dynamic difficulty adjustment impact on players' confidence. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12.

Constant, T., Levieux, G., Buendia, A., and Natkin, S. (2017). From objective to subjective difficulty evaluation in video games. In *IFIP Conference on Human-Computer Interaction*, pages 107–127. Springer.

Cooper, A., Reimann, R., Cronin, D., and Noessel, C. (2014). *About face: the essentials of interaction design*. John Wiley & Sons.

Corrêa, S. M., dos Santos, P. R., Cerqueira, B. B., Mossmann, J. B., and Barbosa, D. N. F. (2022). Model for automatic generation of difficulty curves in digital games. *IEEE Latin America Transactions*, 20(9):2123–2130.

Cox, A., Cairns, P., Shah, P., and Carroll, M. (2012). Not doing but thinking: the role of challenge in the gaming experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 79–88.

Cruz, C. A. and Uresti, J. A. R. (2017). Player-centered game ai from a flow perspective: Towards a better understanding of past trends and future directions. *Entertainment Computing*, 20:11–24.

Csikszentmihalyi, M. and Csikzentmihaly, M. (1990). *Flow: The psychology of optimal experience*, volume 1990. Harper & Row New York.

Cutting, J., Deterding, S., Demediuk, S., and Sephton, N. (2022). Difficulty-skill balance does not affect engagement and enjoyment: A pre-registered study using ai-controlled difficulty.

Dahlin, E., Nyberg, L., Bäckman, L., and Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychology and aging*, 23(4):720.

Darzi, A., McCrea, S. M., Novak, D., et al. (2021). User experience with dynamic difficulty adjustment methods for an affective exergame: Comparative laboratory-based study. *JMIR Serious Games*, 9(2):e25771.

De Houwer, J. (2003). The extrinsic affective simon task. *Experimental psychology*, 50(2):77.

Deci, E. L. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer.

Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268.

Demediuk, S., Tamassia, M., Li, X., and Raffe, W. L. (2019). Challenging ai: Evaluating the effect of mcts-driven dynamic difficulty adjustment on player enjoyment. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–7.

Demediuk, S., Tamassia, M., Raffe, W. L., Zambetta, F., Li, X., and Mueller, F. (2017). Monte carlo tree search based algorithms for dynamic difficulty adjustment. In *2017 IEEE conference on computational intelligence and games (CIG)*, pages 53–59. IEEE.

Denisova, A. and Cairns, P. (2015). Adaptation in digital games: the effect of challenge adjustment on player performance and experience. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 97–101.

Denisova, A. and Cairns, P. (2019). Player experience and deceptive expectations of difficulty adaptation in digital games. *Entertainment Computing*, 29:56–68.

Denisova, A., Cairns, P., Guckelsberger, C., and Zendle, D. (2020). Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (corgis). *International Journal of Human-Computer Studies*, 137:102383.

Deterding, S. (2015). The lens of intrinsic skill atoms: A method for gameful design. *Human–Computer Interaction*, 30(3-4):294–335.

Dewey, J. (2018). *Logic-The theory of inquiry*. Read Books Ltd.

Diaper, D. and Stanton, N. (2003). The handbook of task analysis for human-computer interaction.

Dias, R. and Martinho, C. (2011). Adapting content presentation and control to player personality in videogames. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, pages 1–8.

Dörner, R., Göbel, S., Effelsberg, W., and Wiemeyer, J. (2016). *Serious games*. Springer.

Dunbar, K. (2017). Problem solving. *A companion to cognitive science*, pages 289–298.

Dziedzic, D. and Włodarczyk, W. (2018). Approaches to measuring the difficulty of games in dynamic difficulty adjustment systems. *International Journal of Human–Computer Interaction*, 34(8):707–715.

Ebrahimi, A. and Akbarzadeh-T, M.-R. (2014). Dynamic difficulty adjustment in games by using an interactive self-organizing architecture. In *2014 Iranian Conference on Intelligent Systems (ICIS)*, pages 1–6. IEEE.

Fang, X., Zhang, J., and Chan, S. S. (2013). Development of an instrument for studying flow in computer game play. *International journal of human-computer interaction*, 29(7):456–470.

Fassone, R. (2017). *Every game is an island: Endings and extremities in video games*. Bloomsbury Publishing USA.

Fischer, R. and Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in r. *Frontiers in psychology*, 10:440108.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381.

Fleiss, J. L., Nee, J. C., and Landis, J. R. (1979). Large sample variance of kappa in the case of different sets of raters. *Psychological bulletin*, 86(5):974.

Fong, C. J., Zaleski, D. J., and Leach, J. K. (2015). The challenge–skill balance and antecedents of flow: A meta-analytic investigation. *The Journal of Positive Psychology*, 10(5):425–446.

Frommel, J., Fischbach, F., Rogers, K., and Weber, M. (2018). Emotion-based dynamic difficulty adjustment using parameterized difficulty and self-reports of emotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pages 163–171.

FromSoftware (2012). Dark souls: Artorias of the abyss.

FromSoftware (2014). Dark souls ii: Scholar of the first sin.

FromSoftware (2016). Dark souls iii.

Fu, F.-L., Su, R.-C., and Yu, S.-C. (2009). Egameflow: A scale to measure learners ' enjoyment of e-learning games. *Computers & Education*, 52(1):101–112.

Fulcher, G. and Reiter, R. M. (2003). Task difficulty in speaking tests. *Language testing*, 20(3):321–344.

Fullerton, T. (2014). *Game design workshop: a playcentric approach to creating innovative games.* CRC press.

Fulmer, S. M. and Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction*, 27:11–20.

Gallego-Durán, F. J., Molina-Carmona, R., and Llorens-Largo, F. (2018). Measuring the difficulty of activities for adaptive learning. *Universal access in the information society*, 17(2):335–348.

Games, P. (2009). Plants vs. zombies. *PopCap Games.*

Gergelyfi, M., Jacob, B., Olivier, E., and Zénon, A. (2015). Dissociation between mental fatigue and motivational state during prolonged mental activity. *Frontiers in behavioral neuroscience*, 9:176.

Gibson, J. (2014). *Introduction to Game Design, Prototyping, and Development: From Concept to Playable Game with Unity and C.* addison-wesley professional.

Gilleade, K., Dix, A., and Allanson, J. (2005). Affective videogames and modes of affective gaming: assist me, challenge me, emote me. *DiGRA 2005: Changing Views–Worlds in Play.*

González-Duque, M., Palm, R. B., Ha, D., and Risi, S. (2020). Finding game levels with the right difficulty in a few trials through intelligent trial-and-error. In *2020 IEEE Conference on Games (CoG)*, pages 503–510. IEEE.

González Sánchez, J. L., Padilla Zea, N., and Gutiérrez, F. L. (2009). From usability to playability: Introduction to player-centred video game development process. In *Human Centered Design: First International Conference, HCD 2009, Held as Part* *of HCI International 2009, San Diego, CA, USA, July 19-24, 2009 Proceedings 1*, pages 65–74. Springer.

Gopher, D. and Donchin, E. (1986). Workload: An examination of the concept.

Gray, C. M. (2014). Evolution of design competence in ux practice. In *Proceedings of the SIGCHI Conference on human factors in computing systems*, pages 1645–1654.

Guadagnoli, M. A. and Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of motor behavior*, 36(2):212–224.

Guimarães, A. V., Barbosa, A. R., and Meneghini, V. (2018). Active videogame-based physical activity vs. aerobic exercise and cognitive performance in older adults: a randomized controlled trial. *Journal of Physical Education and Sport*, 18(1):203–209.

Guo, Z., Thawonmas, R., and Ren, X. (2024). Rethinking dynamic difficulty adjustment for video game design. *Entertainment Computing*, page 100663.

Hair, J. F. (2009). Multivariate data analysis.

Harada, C. N., Love, M. C. N., and Triebel, K. L. (2013). Normal cognitive aging. *Clinics in geriatric medicine*, 29(4):737–752.

Hardy, S., Dutz, T., Wiemeyer, J., Göbel, S., and Steinmetz, R. (2015). Framework for personalized and adaptive game-based training programs in health sport. *Multimedia Tools and Applications*, 74:5289–5311.

Harrington, D. (2009). *Confirmatory factor analysis.* Oxford university press.

Harrison, S., Tatar, D., and Sengers, P. (2007). The three paradigms of hci. In *Alt. Chi. Session at the SIGCHI Conference on human factors in computing systems San Jose, California, USA*, pages 1–18.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier.

Hawkins, G., Nesbitt, K., and Brown, S. (2012). Dynamic difficulty balancing for cautious players and risk takers. *International Journal of Computer Games Technology*, 2012.

He, S., Wang, J., Liu, X., Huang, W., et al. (2010). Dynamic difficulty adjustment of game ai by mcts for the game pac-man. In *2010 sixth international conference on* *natural computation*, volume 8, pages 3918–3922. IEEE.

Heintz, S. and Law, E. L.-C. (2015). The game genre map: A revised game classification. In *Proceedings of the 2015 annual Symposium on computer-human Interaction in play*, pages 175–184.

Helander, M. G. (2014). *Handbook of human-computer interaction.* Elsevier.

Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., and Dunwell, I. (2018). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, 11(4):320–327.

Hind, D. and Harvey, C. (2022). A neat approach to wave generation in tower defense games. In *2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET)*, pages 1–8. IEEE.

Hocine, N., Gouaïch, A., Cerri, S. A., Mottet, D., Froger, J., and Laffont, I. (2015). Adaptation in serious games for upper-limb rehabilitation: an approach to improve training outcomes. *User Modeling and User-Adapted Interaction*, 25(1):65–98.

Hocine, N., Gouaich, A., Di Loreto, I., and Joab, M. (2011). Motivation based difficulty adaptation for therapeutic games. In *2011 IEEE 1st International Conference on serious games and applications for health (SeGAH)*, pages 1–8. IEEE.

Hooshyar, D., Malva, L., Yang, Y., Pedaste, M., Wang, M., and Lim, H. (2021). An adaptive educational computer game: Effects on students' knowledge and learning attitude in computational thinking. *Computers in Human Behavior*, 114:106575.

Hornbæk, K. and Oulasvirta, A. (2017). What is interaction? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5040–5052.

Hsu, S. H., Wen, M.-H., and Wu, M.-C. (2007). Exploring design features for enhancing players' challenge in strategy games. *CyberPsychology & Behavior*, 10(3):393–397.

Huang, T.-L., Liao, G.-Y., Cheng, T., Chen, W.-X., and Teng, C.-I. (2024). Tomorrow will be better: Gamers' expectation and game usage. *Computers in Human Behavior*, 151:108021.

Hübener, M. and Bonhoeffer, T. (2014). Neuronal plasticity: beyond the critical period. *Cell*, 159(4):727–737.

Hudlicka, E. (2009). Affective game engines: motivation and requirements. In *Proceedings of the 4th international conference on foundations of digital games*, pages 299–306.

Huizinga, J. (2014). *Homo ludens ils 86.* Routledge.

Hultsch, D. F., MacDonald, S. W., and Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 57(2):P101–P115.

Hung, C.-M., Huang, I., and Hwang, G.-J. (2014). Effects of digital game-based learning on students ’ self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, 1:151–166.

Huniche, R. and Chapman, V. (2005). Ai for dynamic difficulty adjustment in games. *Valencia, Spain.*

Hunicke, R. (2005). The case for dynamic difficulty adjustment in games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 429–433.

Hunicke, R., LeBlanc, M., Zubek, R., et al. (2004). Mda: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, volume 4, page 1722. San Jose, CA.

Hussaan, A. M. and Sehaba, K. (2013). Adaptive serious game for rehabilitation of persons with cognitive disabilities. In *2013 IEEE 13th International Conference on Advanced Learning Technologies*, pages 65–69. IEEE.

IJsselsteijn, W. A., De Kort, Y. A., and Poels, K. (2013). The game experience questionnaire.

Ishihara, M., Ito, S., Ishii, R., Harada, T., and Thawonmas, R. (2018). Monte-carlo tree search for implementation of dynamic difficulty adjustment fighting game ais having believable behaviors. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–8. IEEE.

Jacko, J. A. (2012). Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications.

Jackson, S. A. and Csikszentmihalyi, M. (1999). *Flow in sports.* Human Kinetics.

Jacobs, R. S., Werning, S., Jansz, J., and Kneer, J. (2020). Procedural arguments of persuasive games. *Journal of Media Psychology.*

Jalife, K., Harteveld, C., and Holmgård, C. (2021). From flow to fuse: A cognitive perspective. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–30.

Javadi, A.-H., Emo, B., Howard, L. R., Zisch, F. E., Yu, Y., Knight, R., Pinelo Silva, J., and Spiers, H. J. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. *Nature communications*, 8(1):14652.

Jemmali, C., Seif El-Nasr, M., and Cooper, S. (2022). The effects of adaptive procedural levels on engagement and performance in an educational programming game. In *Proceedings of the 17th International Conference on the Foundations of Digital Games*, pages 1–12.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661.

Jennings-Teats, M., Smith, G., and Wardrip-Fruin, N. (2010). Polymorph: dynamic difficulty adjustment through level generation. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*, pages 1–4.

Johanson, C., Gutwin, C., Bowey, J. T., and Mandryk, R. L. (2019). Press pause when you play: Comparing spaced practice intervals for skill development in games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 169–184.

Johnson, D., Gardner, M. J., and Perry, R. (2018). Validation of two game experience scales: the player experience of need satisfaction (pens) and game experience questionnaire (geq). *International Journal of Human-Computer Studies*, 118:38–46.

Juul, J. (2004). Introduction to game time/time to play: An examination of game temporality. *First person: New media as story, performance and game*, pages 131–142.

Juul, J. (2007). Swap adjacent gems to make sets of three: A history of matching tile games. *Artifact: Journal of Design Practice*, 1(4):205–217.

Juul, J. (2009). Fear of failing? the many meanings of difficulty in video games. *The video game theory reader*, 2(237-252).

Juul, J. (2011). *Half-real: Video games between real rules and fictional worlds*. MIT press.

Kaplan, O., Yamamoto, G., Taketomi, T., Plopski, A., Sandor, C., and Kato, H. (2018). Exergame experience of young and old individuals under different difficulty adjustment methods. *Computers*, 7(4):59.

Karpinskyj, S., Zambetta, F., and Cavedon, L. (2014). Video game personalisation techniques: A comprehensive survey. *Entertainment Computing*, 5(4):211–218.

Katz, D. (1960). The functional approach to the study of attitudes. *Public opinion quarterly*, 24(2):163–204.

Kayama, H., Okamoto, K., Nishiguchi, S., Yamada, M., Kuroda, T., Aoyama, T., et al. (2014). Effect of a kinect-based exercise game on improving executive cognitive performance in community-dwelling elderly: case control study. *Journal of medical Internet research*, 16(2):e3108.

Keates, S., Langdon, P., Clarkson, J., and Robinson, P. (2000). Investigating the use of force feedback for motion-impaired users. In *Proceedings of the 6th ERCIM Workshop*, pages 207–212. Florence Italy.

Khajah, M. M., Roads, B. D., Lindsey, R. V., Liu, Y.-E., and Mozer, M. C. (2016). Designing engaging games using bayesian optimization. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5571–5582.

Khalili-Mahani, N., Assadi, A., Li, K., Mirgholami, M., Rivard, M.-E., Benali, H., Sawchuk, K., De Schutter, B., et al. (2020). Reflective and reflexive stress responses of older adults to three gaming experiences in relation to their cognitive abilities: mixed methods crossover study. *JMIR mental health*, 7(3):e12388.

Khoshkangini, R., Valetto, G., Marconi, A., and Pistore, M. (2021). Automatic generation and recommendation of personalized challenges for gamification. *User Modeling and User-Adapted Interaction*, 31:1–34.

Kickmeier-Rust, M. D. and Albert, D. (2010). Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, 26(2):95–105.

Kim, Y. (2009). *The role of task complexity and pair grouping on the occurrence of learning opportunities and L2 development.* Northern Arizona University.

Kimberlin, C. L. and Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American journal of health-system pharmacy*, 65(23):2276–2284.

Kitakoshi, D., Suzuki, K., and Suzuki, M. (2020). A study on coordination of exercise difficulty in cognitive training system for older adults. In *2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*, pages 1–6. IEEE.

Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., and Ravaja, N. (2011). A review of the use of psychophysiological methods in game research. *journal of gaming & virtual worlds*, 3(3):181–199.

Klarkowski, M., Johnson, D., Wyeth, P., Smith, S., and Phillips, C. (2015). Operationalising and measuring flow in video games. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, pages 114–118.

Klimmt, C., Blake, C., Hefner, D., Vorderer, P., and Roth, C. (2009). Player performance, satisfaction, and video game enjoyment. In *International conference on entertainment computing*, pages 1–12. Springer.

Kling, R. and Star, S. L. (1998). Human centered systems in the perspective of organizational and social informatics. *Acm Sigcas Computers and Society*, 28(1):22–29.

Klisch, Y., Miller, L. M., Wang, S., and Epstein, J. (2012). The impact of a science education game on students' learning and perception of inhalants as body pollutants. *Journal of science education and technology*, 21(2):295–303.

Knorr, J. and Vaz de Carvalho, C. (2021). Using dynamic difficulty adjustment to improve the experience and train fps gamers. In *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*, pages 195–200.

Koskinen, A., McMullen, J., Hannula-Sormunen, M., Ninaus, M., and Kiili, K. (2023). The strength and direction of the difficulty adaptation affect situational interest in game-based learning. *Computers & Education*, 194:104694.

Krath, J., Schürmann, L., and Von Korflesch, H. F. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, 125:106963.

Kristensen, J. T. and Burelli, P. (2022). *Operationalising difficulty in puzzle games*. IT University of Copenhagen, Department of digital design.

Kristensen, J. T., Guckelsberger, C., Burelli, P., and Hämäläinen, P. (2022). Personalized game difficulty prediction using factorization machines. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–13.

Kueider, A. M., Parisi, J. M., Gross, A. L., and Rebok, G. W. (2012). Computerized cognitive training with older adults: a systematic review. *PloS one*, 7(7):e40588.

Laamarti, F., Eid, M., and El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology*, 2014.

Lach, E. (2017). Dynamic difficulty adjustment for serious game using modified evolutionary algorithm. In *International Conference on Artificial Intelligence and Soft Computing*, pages 370–379. Springer.

Laffan, D. A., Greaney, J., Barton, H., and Kaye, L. K. (2016). The relationships between the structural video game characteristics, video game engagement and happiness among individuals who play video games. *Computers in Human Behavior*, 65:544–549.

Larsen, J. M. (2010). Difficulty curves.

Lee, A., Song, K., Ryu, H. B., Kim, J., and Kwon, G. (2015). Fingerstroke time estimates for touchscreen-based mobile gaming interaction. *Human movement science*, 44:211–224.

Lee, Y.-H. and Heeter, C. (2017). The effects of cognitive capacity and gaming expertise on attention and comprehension. *Journal of Computer Assisted Learning*, 33(5):473–485.

Lemmens, J. S. and von Münchhausen, C. F. (2023). Let the beat flow: How game difficulty in virtual reality affects flow. *Acta Psychologica*, 232:103812.

Lewis, M., Haviland-Jones, J. M., and Barrett, L. F. (2010). *Handbook of emotions*. Guilford Press.

Li, X., Niksirat, K. S., Chen, S., Weng, D., Sarcar, S., and Ren, X. (2020). The impact of a multitasking-based virtual reality motion video game on the cognitive and physical abilities of older adults. *Sustainability*, 12(21):9106.

Li, Y. and Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information processing & management*, 44(6):1822–1837.

Li, Y.-N., Yao, C., Li, D.-J., and Zhang, K. (2014). Adaptive difficulty scales for parkour games. *Journal of Visual Languages & Computing*, 25(6):868–878.

Lin, D., Bezemer, C.-P., Zou, Y., and Hassan, A. E. (2019). An empirical study of game reviews on the steam platform. *Empirical Software Engineering*, 24:170–207.

Lin, J.-H. T., Wu, D.-Y., and Tao, C.-C. (2018). So scary, yet so fun: The role of self-efficacy in enjoyment of a virtual reality horror game. *New Media & Society*, 20(9):3223–3242.

Linehan, C., Bellord, G., Kirman, B., Morford, Z. H., and Roche, B. (2014). Learning curves: analysing pace and challenge in four successful puzzle games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, pages 181–190.

Liu, C., Agrawal, P., Sarkar, N., and Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529.

Liu, P. and Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, 42(6):553–568.

Logan, G. D., Zbrodoff, N. J., and Williamson, J. (1984). Strategies in the color-word stroop task. *Bulletin of the Psychonomic Society*, 22(2):135–138.

Lomas, J. D., Koedinger, K., Patel, N., Shodhan, S., Poonwala, N., and Forlizzi, J. L.

(2017). Is difficulty overrated? the effects of choice, novelty and suspense on intrinsic motivation in educational games. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 1028–1039.

Lopes, R. and Bidarra, R. (2011). Adaptivity challenges in games and simulations: a survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(2):85–99.

Løvoll, H. S. and Vittersø, J. (2014). Can balance be boring? a critique of the "challenges should match skills" hypotheses in flow theory. *Social indicators research*, 115(1):117–136.

Luyten, L., Lowyck, J., and Tuerlinckx, F. (2001). Task perception as a mediating variable: A contribution to the validation of instructional knowledge. *British Journal of Educational Psychology*, 71(2):203–223.

MacKinnon, D. P., Krull, J. L., and Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention science*, 1:173–181.

MacLean, K. E. (2000). Designing with haptic feedback. In *Proceedings 2000 icra. millennium conference. ieee international conference on robotics and automation. symposia proceedings (cat. no. 00ch37065)*, volume 1, pages 783–788. IEEE.

MacPherson, S. E. (2018). Definition: Dual-tasking and multitasking. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*.

Maertens, M., Vandewaetere, M., Cornillie, F., and Desmet, P. (2014). From pen-and-paper content to educational math game content for children: A transfer with added difficulty. *International Journal of Child-Computer Interaction*, 2(2):85–92.

Maillot, P., Perrot, A., and Hartley, A. (2012). Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. *Psychology and aging*, 27(3):589.

Malaby, T. M. (2007). Beyond play: A new approach to games. *Games and culture*, 2(2):95–113.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive science*, 5(4):333–369.

Martin, S. (2014). Measuring cognitive load and cognition: metrics for technology-enhanced learning. *Educational Research and Evaluation*, 20(7-8):592–621.

Martin, W. and Magerko, B. (2020). The game as a classroom: Understanding players ' goals and attributions from a learning perspective. In *Proceedings of the 15th International Conference on the Foundations of Digital Games*, pages 1–4.

Martinez, M. E. (1998). What is problem solving? *The Phi Delta Kappan*, 79(8):605–609.

Masanobu, E., Mikami, K., et al. (2017). Dynamic pressure cycle control: Dynamic diffculty adjustment beyond the flow zone. In *2017 Nicograph International (NicoInt)*, pages 9–14. IEEE.

Maynard, D. C. and Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance*, 10(4):303–330.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.

Megawati, R., Listiani, H., Pranoto, N. W., Akobiarek, M., et al. (2023). The role of gpt chat in writing scientific articles: A systematic literature review. *Jurnal Penelitian Pendidikan IPA*, 9(11):1078–1084.

Mekler, E. D., Bopp, J. A., Tuch, A. N., and Opwis, K. (2014). A systematic review of quantitative studies on the enjoyment of digital entertainment games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 927–936.

Miller, M. K. and Mandryk, R. L. (2016). Differentiating in-game frustration from at-game frustration using touch pressure. In *Proceedings of the 2016 ACM international conference on interactive surfaces and spaces*, pages 225–234.

Miyake, Y. (2015). Current status of applying artificial intelligence for digital games. *Journal of the Japanese Society for Artificial Intelligence*, 30(1):45.

Moniaga, J. V., Chowanda, A., Prima, A., Rizqi, M. D. T., et al. (2018). Facial expression recognition as dynamic game balancing system. *Procedia Computer Science*, 135:361–368.

Monterrat, B., Desmarais, M., Lavoué, E., and George, S. (2015). A player model for adaptive gamification in learning environments. In *International conference on artificial intelligence in education*, pages 297–306. Springer.

Moon, J., Choi, Y., Park, T., Choi, J., Hong, J.-H., and Kim, K.-J. (2022). Diversifying dynamic difficulty adjustment agent by integrating player state models into monte-carlo tree search. *Expert Systems with Applications*, page 117677.

Moore, G. C. and Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information systems research*, 2(3):192–222.

Mora-Cantallops, M. and Sicilia, M.-Á. (2018). Moba games: A literature review. *Entertainment computing*, 26:128–138.

Nacke, L. and Lindley, C. A. (2008). Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the 2008 conference on future play: Research, play, share*, pages 81–88.

Nagle, A., Wolf, P., and Riener, R. (2016). Towards a system of customized video game mechanics based on player personality: Relating the big five personality traits with difficulty adaptation in a first-person shooter game. *Entertainment computing*, 13:10–24.

Nakamura, J. and Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the foundations of positive psychology*, pages 239–263. Springer.

Neisser, U. (2014). *Cognitive psychology: Classic edition*. Psychology press.

Nirme, J., Duff, A., and Verschure, P. F. (2011). Adaptive rehabilitation gaming system: on-line individualization of stroke rehabilitation. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6749–6752. IEEE.

Norsworthy, C., Jackson, B., and Dimmock, J. A. (2021). Advancing our understanding of psychological flow: A scoping review of conceptualizations, measurements, and applications. *Psychological Bulletin*, 147(8):806.

Nuutila, K., Tapola, A., Tuominen, H., Molnár, G., and Niemivirta, M. (2021). Mutual relationships between the levels of and changes in interest, self-efficacy, and perceived difficulty during task engagement. *Learning and Individual Differences*, 92:102090.

O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955.

Ordnung, M., Hoff, M., Kaminski, E., Villringer, A., and Ragert, P. (2017). No overt effects of a 6-week exergame training on sensorimotor and cognitive function in older adults. a preliminary investigation. *Frontiers in human neuroscience*, 11:160.

Orji, R., Mandryk, R. L., and Vassileva, J. (2017). Improving the efficacy of games for change using personalization models. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(5):1–22.

Orvis, K. A., Horn, D. B., and Belanich, J. (2007). *Task difficulty and prior videogame experience: Their role in performance and motivation in instructional videogames.* US Army Research Institute for the Behavioral and Social Sciences.

Orvis, K. A., Horn, D. B., and Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames.

Computers in Human behavior, 24(5):2415–2433.

O'Shea, Z. and Freeman, J. (2019). Game design frameworks: Where do we start? In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–10.

Oshita, M. and Ishikawa, H. (2012). Gamepad vs. touchscreen: a comparison of action selection interfaces in computer games. In *Proceedings of the Workshop at SIGGRAPH Asia*, pages 27–31.

Ouellette, M., Breeding, L., and Clark, C. (2019). Using applied cognitive load theory and difficulty analysis for educational game design for understanding and transference of literacy skills in adults. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–11.

Ozkul, F., Palaska, Y., Masazade, E., and Erol-Barkana, D. (2019). Exploring dynamic difficulty adjustment mechanism for rehabilitation tasks using physiological measures and subjective ratings. *IET Signal Processing*, 13(3):378–386.

Paavilainen, J. (2017). Playability: A game-centric definition. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, pages 487–494.

Paavilainen, J. (2020). Defining playability of games: functionality, usability, and gameplay. In *Proceedings of the 23rd International Conference on Academic Mindtrek*, pages 55–64.

Paraschos, P. D. and Koulouriotis, D. E. (2023). Game difficulty adaptation and experience personalization: a literature review. *International Journal of Human–Computer Interaction*, 39(1):1–22.

Pato, V. M. Á. and Delgado-Mata, C. (2013). Dynamic difficulty adjusting strategy for a two-player video game. *Procedia Technology*, 7:315–321.

Pavlas, D. (2010). A model of flow and play in game-based learning the impact of game characteristics, player traits, and player states.

Pedersen, C., Togelius, J., and Yannakakis, G. N. (2010). Modeling player experience for content creation. *IEEE Transactions on Computational Intelligence and AI in Games*, 2(1):54–67.

Peng, X., Xie, X., Huang, J., Jiang, C., Wang, H., Denisova, A., Chen, H., Tian, F., and Wang, H. (2023). Challengedetect: Investigating the potential of detecting in-game challenge experience from physiological measures. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–29.

Petralito, S., Brühlmann, F., Iten, G., Mekler, E. D., and Opwis, K. (2017). A good reason to die: how avatar death and high challenges enable positive experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5087–5097.

Pezzera, M. and Borghese, N. A. (2020). Dynamic difficulty adjustment in exer-games for rehabilitation: a mixed approach. In *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–7. IEEE.

Pfau, J., Smeddinck, J. D., and Malaka, R. (2020). Enemy within: Long-term motivation effects of deep player behavior models for dynamic difficulty adjustment. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–10.

Picard, R. W. (2000). *Affective computing*. MIT press.

Power, J., Lynch, R., and McGarr, O. (2020). Difficulty and self-efficacy: An exploratory study. *British Journal of Educational Technology*, 51(1):281–296.

Pratama, H. A. and Krisnadhi, A. A. (2018). Representing dynamic difficulty in turn-based role playing games using monte carlo tree search. In *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 207–212. IEEE.

Preacher, K. J. and Hayes, A. F. (2004). Spss and sas procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, & computers*, 36:717–731.

Przybylski, A. K., Rigby, C. S., and Ryan, R. M. (2010). A motivational model of video game engagement. *Review of general psychology*, 14(2):154–166.

Purnama, A., Akbar, S., and Dharma, D. (2018). Adjustment of difficulty level on wobble board-based game using monte carlo tree search algorithm. In *2018 5th International Conference on Data and Software Engineering (ICoDSE)*, pages 1–6. IEEE.

Pusey, M., Wong, K. W., and Rappa, N. A. (2021). The puzzle challenge analysis tool. a tool for analysing the cognitive challenge level of puzzles in video games. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–27.

Qin, H., Rau, P.-L. P., and Salvendy, G. (2010). Effects of different scenarios of game difficulty on player immersion. *Interacting with computers*, 22(3):230–239.

Raven, J. C. and Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press Oxford.

Rawlings, T. (2012). My cotton picking life. *GameTheNews*.

Ren, X. (2016). Rethinking the relationship between humans and computers. *Computer*, 49(8):104–108.

Ren, X., Silpasuwanchai, C., and Cahill, J. (2019). Human-engaged computing: the future of human–computer interaction. *CCF transactions on pervasive computing and interaction*, 1(1):47–68.

Rigby, S. and Ryan, R. (2007). The player experience of need satisfaction (pens): An applied model and methodology for understanding key components of the player experience. 2007. sep. *URL: https://natronbaxter. com/wp-content/uploads/2010/05/PENS_Sept07. pdf [accessed 2020-12-20]*.

Risi, S. and Togelius, J. (2020). Increasing generality in machine learning through procedural content generation. *Nature Machine Intelligence*, 2(8):428–436.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, 22(1):27–57.

Rodgers, W. M., Conner, M., and Murray, T. C. (2008). Distinguishing among perceived control, perceived difficulty, and self-efficacy as determinants of intentions and behaviours. *British journal of social psychology*, 47(4):607–630.

Roohi, S., Relas, A., Takatalo, J., Heiskanen, H., and Hämäläinen, P. (2020). Predicting game difficulty and churn without players. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 585–593.

Rosner, B., Glynn, R. J., and Lee, M.-L. T. (2006). The wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*, 62(1):185–192.

Rossoff, S., Tzanetakis, G., and Gooch, B. (2010). Adapting personal music for synesthetic game play. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, pages 163–170.

Rozado, D., Moreno, T., San Agustin, J., Rodriguez, F., and Varona, P. (2015). Controlling a smartphone using gaze gestures as the input mechanism. *Human–Computer Interaction*, 30(1):34–63.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3):450.

Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, 30(4):344–360.

Sakaue, S., Kimura, T., and Nishino, H. (2023). Reducing objective difficulty without influencing subjective difficulty in a video game. In *Proceedings of the 5th ACM*

International Conference on Multimedia in Asia, pages 1–5.

Salehzadeh Niksirat, K., Silpasuwanchai, C., Ren, X., and Wang, Z. (2017). Towards cognitive enhancement of the elderly: A ux study of a multitasking motion video game. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems*, pages 2017–2024.

Sampayo-Vargas, S., Cope, C. J., He, Z., and Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education*, 69:452–462.

Sánchez, J. L. G., Vela, F. L. G., Simarro, F. M., and Padilla-Zea, N. (2012). Playability: analysing user experience in video games. *Behaviour & Information Technology*, 31(10):1033–1054.

Sarkar, A. and Cooper, S. (2019). Transforming game difficulty curves using function composition. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion*, 24(7):1153–1172.

Schättin, A., Arner, R., Gennaro, F., and de Bruin, E. D. (2016). Adaptations of prefrontal brain activity, executive functions, and gait in healthy elderly following exergame and balance training: a randomized-controlled study. *Frontiers in aging neuroscience*, 8:278.

Scheepers, D. and Keller, J. (2022). On the physiology of flow: Bridging flow theory with the biopsychosocial model of challenge and threat. *International Journal of Psychophysiology*, 182:119–128.

Schell, J. (2019). *Tenth anniversary: The art of game design: A book of lenses.* AK Peters/CRC Press.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., and Zelaznik, H. N. (2018). *Motor control and learning: A behavioral emphasis.* Human kinetics.

Schoene, D., Valenzuela, T., Toson, B., Delbaere, K., Severino, C., Garcia, J., Davies, T. A., Russell, F., Smith, S. T., and Lord, S. R. (2015). Interactive cognitive-motor step training improves cognitive risk factors of falling in older adults–a randomized controlled trial. *PLoS one*, 10(12):e0145161.

Sellers, M. (2017). *Advanced game design: a systems approach.* Addison-Wesley Professional.

Sepulveda, G. K., Besoain, F., and Barriga, N. A. (2019). Exploring dynamic difficulty adjustment in videogames. In *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, pages 1–6. IEEE.

Seyderhelm, A. J., Blackmore, K. L., and Nesbitt, K. (2019). Towards cognitive adaptive serious games: A conceptual framework. In *Joint International Conference on Entertainment Computing and Serious Games*, pages 331–338. Springer.

Shakhova, M. and Zagarskikh, A. (2019). Dynamic difficulty adjustment with a simplification ability using neuroevolution. *Procedia Computer Science*, 156:395–403.

Sharek, D. and Wiebe, E. (2014). Measuring video game engagement through the cognitive and affective dimensions. *Simulation & Gaming*, 45(4-5):569–592.

Shute, V. J., Ventura, M., and Ke, F. (2015). The power of play: The effects of portal 2 and lumosity on cognitive and noncognitive skills. *Computers & education*, 80:58–67.

Sideridis, G. D., Kaplan, A., Papadopoulos, C., and Anastasiadis, V. (2014). The affective experience of normative-performance and outcome goal pursuit: Physiological, observed, and self-report indicators. *Learning and Individual Differences*, 32:114–123.

Silva, M. P., do Nascimento Silva, V., and Chaimowicz, L. (2015). Dynamic difficulty adjustment through an adaptive ai. In *2015 14th Brazilian symposium on computer games and digital entertainment (SBGames)*, pages 173–182. IEEE.

Smeddinck, J. D., Mandryk, R. L., Birk, M. V., Gerling, K. M., Barsilowski, D., and Malaka, R. (2016). How to present game difficulty choices? exploring the impact on player experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5595–5607.

Smith, G. E., Housen, P., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., and Zelinski, E. M. (2009). A cognitive training program based on principles of brain plasticity: results from the improvement in memory with plasticity-based adaptive cognitive training (impact) study. *Journal of the American Geriatrics Society*, 57(4):594–603.

Smyth, M. M. and Scholey, K. A. (1992). Determining spatial span: The role of movement time and articulation rate. *The Quarterly Journal of Experimental Psychology Section A*, 45(3):479–501.

Soderman, B. (2021). *Against Flow: Video Games and the Flowing Subject*. MIT Press.

Sorenson, N., Pasquier, P., and DiPaola, S. (2011). A generic approach to challenge modeling for the procedural creation of video game levels. *IEEE Transactions on*

Computational Intelligence and AI in Games, 3(3):229–244.

Sparrow, B., Liu, J., and Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science*, 333(6043):776–778.

Speelman, C. P. and Kirsner, K. (2005). *Beyond the learning curve: The construction of mind.* Oxford University Press, USA.

Spence, I. and Feng, J. (2010). Video games and spatial cognition. *Review of general psychology*, 14(2):92–104.

Spiel, K., Bertel, S., and Kayali, F. (2019). Adapting gameplay to eye movements-an exploration with tetris. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pages 687–695.

Stajkovic, A. D. and Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological bulletin*, 124(2):240.

Stammer, D., Günther, T., and Preuss, M. (2015). Player-adaptive spelunky level generation. In *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 130–137. IEEE.

Steam (2023). Welcome to steam.

Steele, J. (2020). What is (perception of) effort? objective and subjective effort during task performance. *PsyArXiv*.

Stein, A., Yotam, Y., Puzis, R., Shani, G., and Taieb-Maimon, M. (2018). Eeg-triggered dynamic difficulty adjustment for multiplayer games. *Entertainment computing*, 25:14–25.

Sternberg, R. J. and Kaufman, S. B. (2011). *The Cambridge handbook of intelligence.* Cambridge University Press.

Streicher, A. and Smeddinck, J. D. (2016). Personalized and adaptive serious games. In *Entertainment Computing and Serious Games: International GI-Dagstuhl Seminar 15283, Dagstuhl Castle, Germany, July 5-10, 2015, Revised Selected Papers*, pages 332–377. Springer.

Suovuo, T. “ ., Skult, N., Joelsson, T. N., Skult, P., Ravyse, W., and Smed, J. (2020). The game experience model (gem). *Game User Experience and Player-Centered Design*, pages 183–205.

Sutoyo, R., Winata, D., Oliviani, K., and Supriyadi, D. M. (2015). Dynamic difficulty adjustment in tower defence. *Procedia Computer Science*, 59:435–444.

Swann, C., Piggott, D., Schweickle, M., and Vella, S. A. (2018). A review of scientific progress in flow in sport and exercise: normal science, crisis, and a progressive shift.

Journal of Applied Sport Psychology, 30(3):249–271.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312.

Sykes, J. and Federoff, M. (2006). Player-centred game design. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1731–1734.

Tan, C. H., Tan, K. C., and Tay, A. (2011). Dynamic game difficulty scaling using adaptive behavior-based ai. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(4):289–301.

Tan, D. and Nijholt, A. (2010). *Brain-computer interfaces and human-computer interaction*. Springer.

Tan, E. S.-H. (2008). Entertainment is emotion: The functional architecture of the entertainment experience. *Media psychology*, 11(1):28–51.

Tavakol, M. and Dennick, R. (2011). Making sense of cronbach's alpha. *International journal of medical education*, 2:53.

Tekinbas, K. S. and Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. MIT press.

Teoh, A. N., Kaur, D., Dillon, R., and Hristova, D. (2020). Developing gaming instinctual motivation scale (gims): item development and pre-testing. *Game User Experience And Player-Centered Design*, pages 163–182.

Thomée, S., Härenstam, A., and Hagberg, M. (2011). Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults-a prospective cohort study. *BMC public health*, 11:1–11.

Tombaugh, T. N. and McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, 40(9):922–935.

Tondello, G. F. and Nacke, L. E. (2019). Player characteristics and video game preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 365–378.

Toril, P., Reales, J. M., Mayas, J., and Ballesteros, S. (2016). Video game training enhances visuospatial working memory and episodic memory in older adults. *Frontiers in human neuroscience*, 10:206.

Trafimow, D., Sheeran, P., Conner, M., and Finlay, K. A. (2002). Evidence that perceived behavioural control is a multidimensional construct: Perceived control and perceived difficulty. *British journal of social psychology*, 41(1):101–121.

Tsai, F.-H. (2016). The effectiveness evaluation among different player-matching mechanisms in a multi-player quiz game. *Journal of Educational Technology & Society*, 19(4):213–224.

Tudge, J. R. and Winterhoff, P. A. (1993). Vygotsky, piaget, and bandura: Perspectives on the relations between the social world and cognitive development. *Human development*, 36(2):61–81.

Tyack, A. and Mekler, E. D. (2021). Off-peak: An examination of ordinary player experience. In *Proceedings of the 2021 CHI Conference on Human factors in computing systems*, pages 1–12.

Ulmer, J., Braun, S., Cheng, C.-T., Dowey, S., and Wollert, J. (2022). Gamification of virtual reality assembly training: Effects of a combined point and level system on motivation and training results. *International Journal of Human-Computer Studies*, 165:102854.

Urakami, J., Hu, Y. Z., and Chignell, M. (2021). Monitoring cognitive performance with a serious game: A longitudinal case study on online cognitive assessment using serious games. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Vahlo, J. and Karhulahti, V.-M. (2020). Challenge types in gaming validation of video game challenge inventory (cha). *International Journal of Human-Computer Studies*, 143:102473.

Valencia, Y., Majin, J., Guzmán, D., and Londoño, J. (2018). Dynamic difficulty adjustment in virtual reality applications for upper limb rehabilitation. In *2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA)*, pages 1–6. IEEE.

Valve (2008). Left 4 dead. *Valve Corporation.*

van Dinther, M., Dochy, F., Segers, M., and Braeken, J. (2014). Student perceptions of assessment and student self-efficacy in competence-based education. *Educational Studies*, 40(3):330–351.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., and Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of applied psychology*, 87(3):506.

Veltman, J. and Gaillard, A. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5):656–669.

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., and Cerella, J. (2003). Aging and dual-task performance: a meta-analysis. *Psychology and aging*, 18(3):443.

Vorderer, P., Klimmt, C., and Ritterfeld, U. (2004). Enjoyment: At the heart of media entertainment. *Communication theory*, 14(4):388–408.

Wang, X., Niksirat, K. S., Silpasuwanchai, C., Wang, Z., Ren, X., and Niu, Z. (2016). How skill balancing impact the elderly player experience? In *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pages 983–988. IEEE.

Wehbe, R. R., Mekler, E. D., Schaekermann, M., Lank, E., and Nacke, L. E. (2017). Testing incremental difficulty design in platformer games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5109–5113.

Wen, Z. and Ye, B. (2014). Analyses of mediating effects: the development of methods and models. *Advances in psychological Science*, 22(5):731.

Wenger, E., Schaefer, S., Noack, H., Kühn, S., Mårtensson, J., Heinze, H.-J., Düzel, E., Bäckman, L., Lindenberger, U., and Lövdén, M. (2012). Cortical thickness changes following spatial navigation training in adulthood and aging. *Neuroimage*, 59(4):3389–3397.

Wheat, D., Masek, M., Lam, C. P., and Hingston, P. (2016). Modeling perceived difficulty in game levels. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–8.

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). *Engineering Psychology and Human Performance*. Psychology Press.

Wiechmann, A., Hall, J. R., and O'Bryant, S. E. (2010). The utility of the spatial span in a clinical geriatric population. *Aging, Neuropsychology, and Cognition*, 18(1):56–63.

Wiemeyer, J. and Hardy, S. (2013). Serious games and motor learning: concepts, evidence, technology. In *Serious games and virtual worlds in education, professional development, and healthcare*, pages 197–220. IGI Global.

Wiemeyer, J., Nacke, L., Moser, C., and ‘Floyd’ Mueller, F. (2016). Player experience. *Serious games: Foundations, concepts and practice*, pages 243–271.

Wilkinson, R. T. and Allison, S. (1989). Age and simple reaction time: decade differences for 5,325 subjects. *Journal of gerontology*, 44(2):P29–P35.

Winne, P. (1985). Cognitive processing in the classroom. *The international encyclopedia of education*, 2:795–808.

Wobbrock, J. O., Cutrell, E., Harada, S., and MacKenzie, I. S. (2008). An error model for pointing based on fitts' law. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1613–1622.

Woolson, R. F. (2007). Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3.

Wu, W.-H., Hsiao, H.-C., Wu, P.-L., Lin, C.-H., and Huang, S.-H. (2012). Investigating the learning-theory foundations of game-based learning: a meta-analysis. *Journal of Computer Assisted Learning*, 28(3):265–279.

Xue, S., Wu, M., Kolen, J., Aghdaie, N., and Zaman, K. A. (2017). Dynamic difficulty adjustment for maximized engagement in digital games. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 465–471.

Yanase, Y. and Narumi, T. (2016). Transparently adjusting difficulty in a jump-action game. *Transactions of the Virtual Reality Society of Japan*, 21(3):415–422.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & behavior*, 9(6):772–775.

Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., et al. (2024). Gpt (generative pre-trained transformer)–a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*.

Yildirim, O., Surer, E., et al. (2021). Developing adaptive serious games for children with specific learning difficulties: A two-phase usability and technology acceptance study. *JMIR Serious Games*, 9(2):e25997.

Yin, H., Luo, L., Cai, W., Ong, Y.-S., and Zhong, J. (2015). A data-driven approach for online adaptation of game difficulty. In *2015 IEEE conference on computational intelligence and games (CIG)*, pages 146–153. IEEE.

Young, G., Zavelina, L., and Hooper, V. (2008). Assessment of workload using nasa task load index in perianesthesia nursing. *Journal of PeriAnesthesia Nursing*, 23(2):102–110.

YouTube (2023). Youtube.

Yu, H. and Trawick, T. (2011). Personalized procedural content generation to minimize frustration and boredom based on ranking algorithm. In *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*.

Yun, C., Trevino, P., Holtkamp, W., and Deng, Z. (2010). Pads: enhancing gaming experience using profile-based adaptive difficulty system. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pages 31–36.

Zajkac-Lamparska, L., WilkoscDkebczyska, M., Wojciechowski, A., Podhorecka, M., Polak-Szabela, A., Warchol, L., Kkedziora-Kornatowska, K., Araszkiewicz, A., and Izdebski, P. (2019). Effects of virtual reality-based cognitive training in older adults living without and with mild dementia: a pretest–posttest design pilot study. *BMC* *research notes*, 12:1–8.

Zhai, S., Kong, J., and Ren, X. (2004). Speed–accuracy tradeoff in fitts ’ law tasks— on the equivalency of actual and nominal pointing precision. *International journal of human-computer studies*, 61(6):823–856.

Zhang, J. (2021). Directly controlling the perceived difficulty of a shooting game by the addition of fake enemy bullets. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–5.

Zhang, P. and Dillon, A. (2003). Hci and mis: shared concerns. *International Journal of Human-Computer Studies*, 59(4):397–402.

Zhu, J. and Ontañón, S. (2020). Player-centered ai for automatic game personalization: Open problems. In *Proceedings of the 15th International Conference on the Foundations of Digital Games*, pages 1–8.

Zohaib, M. (2018). Dynamic difficulty adjustment (dda) in computer games: A review. *Advances in Human-Computer Interaction*, 2018.

Zook, A. and Riedl, M. (2012). A temporal data-driven player model for dynamic difficulty adjustment. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 8, pages 93–98.

Zung, W. W. (1965). A self-rating depression scale. *Archives of general psychiatry*, 12(1):63–70.

Zung, W. W. (1971). Self-rating anxiety scale. *BMC Psychiatry*.

Appendix A

Online Survey Materials

A.1 Online Survey Materials

A.1.1 Materials of the card sorting procedure

Subjective game difficulty refers to the player's perceptions of difficulty with the game. These perceptions arise from the player's overall evaluation of three aspects: the game's attributes that may cause difficulty, the actual difficulties the player faced while playing, and the player's motivational and emotional states when encountering difficulty in play. We provide you with six dimensions related to subjective game difficulty and introduce these dimensions (see Table A.1). We provide the 60 initial items in random order for classification into these dimensions.

Table A.1 The Definitions of the Six Dimensions.

Dimensions	Definitions
Game Complexity	The player's perception of the game's attributes that may cause difficulty. Players can describe these attributes without the game playing.
Game Completion Difficulty	The player's general perception of the game's completion difficulty. Players can describe this difficulty without the game playing.
Game-playing Difficulty	The player's perception of the actual difficulties they faced while playing. These difficulties are caused by the game's demands on players' skill level and are experienced by players during the game playing. This type of difficulty can be described after the game playing.
Player Competence	The player's perception of their performance and competence in playing this game. Playing experience in this game is necessary for players' competence self-evaluation. Therefore, competence is usually evaluated after the game playing.
Player Pressure	The player's perception of stress and other accompanying negative feelings from completing the game. Pressure feeling is experienced during game playing and is usually evaluated after the game playing.
Player Effort	The player's perception of their effort and investment in completing the game. Effort feeling is experienced during game playing and is usually evaluated after the game playing.

Please classify each of the following 60 initial items into one of the six provided dimensions. For instance, if you believe the item "This game is..." pertains to the player's perception of "Game Complexity", select this dimension from the six available options.

You can review the introduction to these dimensions when unsure how one item can be classified, or you can select the "Other" option and provide your opinion. The number of items classified into each dimension does not need to be the same.

A.1.2 Materials of the Scale Testing

To attend this survey, you must have played the game "*Plants vs. Zombies*" produced by Popcap Games because we need you to complete this survey based on your experience in this game. Therefore, before the formal survey, you must correctly answer five questions about this game to prove you have played it and are familiar with it (see Table A.2).

Table A.2 Questions about *Plants VS. Zombies* before the formal survey

NO.	Question contents	Types	Options
1	In the game, the resource needed to grow plants is:	Single choice	A Coins, B. Sunlight C. Water, D. Seeds
2	In the game, there is a type of plant that needs to to be planted in night conditions; they are:	Single choice	A. Fruits, B. Flowers C. Peas D. Mushrooms
3	In the game, when zombies encounter plants, they will ? plants.	Single choice	A. Eat, B. Attack C. Dig, D. Pull-out
4	In the game, what is the name of your neighbor?	Single choice	A. Crazy Jack, B. Crazy Dave C. Crazy Steve, D. Crazy Inventor
5	In the game, the factors you need to consider are:	Multiple choice	A. Types of plants B. Locations of plants C. Order of planting D. Types of zombies E. Routes of zombies F. Quantity of resources

Appendix B

Final Version of Subjective Game Difficulty Scale

B.1 English Version of the SGDS

The Subjective Game Difficulty scale (Below are listed all 25 items that can be used as needed. To score this instrument, you must reverse score the items for the Player Competence subscale by subtracting each item response from 8 and score other items as usual. The higher the average score of each subscale, the more difficult the player perceives that dimension to be.)

For each of the following statements, please indicate how much you agree with it, using the following scale:

1-Strongly disagree, 2-disagree, 3-Somewhat disagree, 4-Neutral, 5-Somewhat agree, 6-Agree, 7-Strongly agree

Game Complexity

I think relationships among game elements are complex.

I think the information provided by this game is too much.

I think the rules of this game are complex.

This game is very difficult to understand.

Game Completion Difficulty

This game is very difficult to complete.

This game looks impossible to win.

The goal of this game is unachievable.

Game-playing Difficulty

I had to observe very carefully when playing this game.

I had to identify different things carefully in this game playing.

Thinking fast was an important part of playing this game.

I had to act quickly when playing this game.

Playing this game demanded precision in my actions.

Player Competence

I feel competent in this game.

I feel very capable and effective in this game.

I am pretty skilled in this game.

I am satisfied with my performance at this game.

I think I did pretty well in this game, compared to other players.

I am better than average in this game.

Player Pressure

I felt very tense while playing this game.

The stress of this game was beyond my scope.

I felt very pressured while playing this game.

Playing this game made me very discouraged.

Player Effort

Playing this game required me to put great effort.

I put much effort into this game.

I invested much energy into this game.

B.2 Chinese Version of the SGDS

主 观游戏难度量表

（以下列出了所有 25 个 问题，可根据需要使用。要 对这个量表 进行 评分，您必须对“玩家 胜任感” 分量表的条目 进行反向 计分，方法是用 8减去 每个条目的回答分数，而其他条目则按常 规方式 计分。每个分量表的平均分越高，表示玩家 认为该维度越困 难。）

对于以下 每个 陈述，请使用下列尺度表明 您对它 们的同意程度：

1-非常不同意，2-不同意，3-有些不同意，4-中立，5-有些同意，6-同意，7-非常同意

游戏复杂度

我 认为游戏要素之 间的 关系是复杂的。

我 认为游戏给的信息太多了。

我 认为游戏的 规则是复杂的。

这个游戏很 难理解。

游戏完成 难度

这个游戏很 难完成。

这个游戏看起来不可能 赢。

这个游戏的目标是不可能 达成的。

游戏游玩 难度

这个游戏很 难完成。

玩游戏时我必须非常仔 细地 观察。

玩游戏时我必须仔 细识别不同的事物。

快速思考是玩游戏的重要 组成部分。

玩游戏时我必须迅速行 动。

玩游戏时需要我的操作精 确。

玩家 胜任感

我感到自己有能力玩 这个游戏。

我 觉得自己在 这个游戏中非常有能力和有效率。

我在 这个游戏上很熟 练。

我 对自己在 这个游戏中的表 现感到 满意。

我 认为与其他玩家相比，我在 这个游戏中表 现得很好。

我在 这个游戏上超 过了平均水准。

玩家 压力感

玩游戏时我感 觉非常紧张。

这个游戏的 压力超出了我的 极限。

我在玩游戏时感到了很大 压力。

玩 这个游戏让我十分灰心丧气。

玩家努力感

玩 这个游戏需要我付出很大的努力。

我为这个游戏付出了很多努力。

我在游戏中投入了很多精力。

B.3 Japanese Version of the SGDS

主観的ゲーム難易度尺度

(以下に 25 の質問がすべて列挙されており、必要に応じて使用することができます。このツールを採点に使用するには、「プレイヤー有能感」サブスケールにある質問のスコアを反転させる必要があります。つまり、質問のスコアを 8 から引きます。その他の部分は通常通り使用します。各サブスケールの平均スコアが高いほど、プレイヤーはその次元をより難しいと感じることになります。)

次のそれぞれの記述について、以下の尺度を使用してどの程度同意するかを示してくだ

さい。

1-強く不同意、2-不同意、3-やや不同意、4-どちらでもない、5-やや同意、6-同意、7-強く同意

ゲームの複雑さ

このゲーム要素間の関係性が複雑だと思う。

このゲームは提供される情報が多すぎると思う。

このゲームのルールは複雑だと思う。

このゲームを理解するのはとても難しい。

ゲーム完了の難度

このゲームを完了するのは難しい。

このゲームで勝つことは不可能に見える。

このゲームの目標を達成することは不可能だ。

ゲームプレイの難度

ゲームプレイしているとき私は細心の注意を払わなければならなかった。

ゲームプレイ中私は様々なことを注意深く識別しなければならなかった。

素早く考えることはゲームをプレイする上で重要な部分であった。

ゲームプレイしているとき私は素早く動作しなければならなかった。

ゲームプレイでは動作に正確さが求められた。

プレイヤーの有能感

このゲームで私は有能だと感じる。

このゲームで私はとても有能で効率的だと感じる。

私はこのゲームに熟練している。

私はこのゲームでの自分のパフォーマンスに満足している。

他のプレイヤーと比較して、私はこのゲームでとても上手だったと思う。

このゲームで私は平均以上に優れている。

プレイヤーの重圧感

このゲームをプレイしている間、私はとても緊張していた。

このゲームで感じるストレスは私の限界を超えた。

このゲームをプレイしている間、私はとてもプレッシャーを感じた。

このゲームをプレイして私はとても気落ちした。

プレイヤーの努力感

このゲームをプレイするには多大な努力が求められた。

私はこのゲームに多くの努力を注いだ。

私はこのゲームに多くのエネルギーを注いだ。

Appendix C

Player Experience Questionnaire

Player Experience Questionnaire

Below are 51 items about your experience in the game, please answer them according to you real feelings. For each of the following statements, please indicate how much you agree with them, using the following scale: 1-Strongly disagree, 2-disagree, 3-Somewhat disagree, 4-Neutral, 5-Somewhat agree, 6-Agree, 7-Strongly agree

1) I feel competent at the game.
2) I feel very capable and effective when playing.
3) My ability to play the game is well matched with the game's challenges.
4) The game provides me with interesting options and choices.
5) The game lets you do interesting things.
6) I experienced a lot of freedom in the game.
7) Learning the game controls was easy.
8) The game controls are intuitive.
9) When I wanted to do something in the game, it was easy to remember the corresponding control.
10) I enjoyed playing this game very much.
11) This game was fun to play.
12) I thought this was a boring game.
13) This game did not hold my attention at all.
14) I would describe this game as very interesting.
15) I thought this game was quite enjoyable.
16) While I was playing this game, I was thinking about how much I enjoyed it.
17) I put a lot of effort into this.
18) I didn ' t try very hard to do well at this game.

19) I tried very hard on this game.
20) It was important to me to do well at this game.
21) I didn ’ t put much energy into this.
22) I believe this game could be of some value to me.
23) I think that doing this game is useful.
24) I think this is important to do because it can make me happy.
25) I would be willing to do this again because it has some value to me.
26) I think doing this activity could help me to keep healthy.
27) I believe playing this game could be beneficial to me.
28) I think this is an important activity.
29) Playing this game challenges me.
30) Playing this game could provide a good test of my skills.
31) I find that playing this game stretches my capabilities to my limits.
32) I was challenged by this game, but I believed I am able to overcome these challenges.
33) I knew clearly what I wanted to do in this game.
34) I knew what I wanted to achieve in this game.
35) My goals were clearly defined.
36) While playing this game, I had a good idea about how well I was doing.
37) I was aware of how well I was performing in this game.
38) I receive immediate feedback on my actions.
39) My attention was focused entirely on the game that I was playing.
40) When playing this game, I was totally concentrated on what I was doing.
41) When playing this game, I felt in control over what I was doing in the game.
42) I feel comfortable with the controls of this game.
43) I often find myself doing things spontaneously and automatically without having to think.
44) The When I play the game, I feel I am in a world created by the game.
45) I kind of forgot about myself when playing this game.
46) I lost the consciousness of my identity and felt like “ melted ” into the game.
47) When I played this game, I sometimes felt like things were happening in slow motion.
48) When I play this game, I tend to lose track of time.
49) Playing this game is rewarding in itself.
50) I loved the feeling of that performance and want to capture it again.
51) I enjoyed the experience.