Title	データ駆動型ゲーム開発ービデオゲームを題材としたプラットフォーム公開,コンテンツ生成,経験駆動型デザインに関する分析
Author(s)	Muhammad, Nazhif Rizani
Citation	
Issue Date	2023-06
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18704
Rights	
Description	Supervisor:飯田 弘之, 先端科学技術研究科, 博士

Doctoral Dissertation

Data-Driven Game Development: Analysis of Publishing Platform, Content Generation, and Experience-Driven Design of Video Games

ASRIZAL, Muhammad Nazhif Rizani

Supervisor: Hiroyuki Iida

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
Information Science

June, 2023

Abstract

The video game industry is highly competitive and rapidly growing, making it crucial for game developers to conduct thorough market research, gather feedback from players, and engage in effective communication with them to develop successful and engaging games. As the market becomes more competitive, it becomes increasingly difficult for game developers to create successful games that stand out in such a crowded field. To achieve success, game developers need to conduct extensive market research to understand the preferences and interests of their target audience. They also need to gather feedback from players during and after the development process, using strategies such as early access releases to receive feedback and improve their games. Historically, game development has not been data-driven, but as the game industry matures, data is becoming an essential and integral part of the game development life cycle to support decision-making across all stages of the process. The uses of data in-game development can be split into two stages: creation and optimization. The creation stage includes concepting, pre-production, and production stages, while the optimization stage includes testing or beta testing, launch, and post-production or live operation stages.

The main objective of this dissertation is to gain a deeper understanding of how game refinement and motion-in-mind theories can be applied in data-driven game development, and how they can be used to measure the entertainment aspect and content quality of video games, which can be useful for game developers and researchers to create better and more engaging video games. To achieve it, we are guided by two purposes: (1) To measure the entertainment aspect of video games from their steam storefront data and to improve the game's visibility on the Steam Platform?, and (2) To define the indicator to measure content quality (difficulty) and player performance in *FlowFree*, To explore how the value of this indicator differ based on player type. This dissertation provide insights for game developers to create better and more engaging video games as well as providing a new perspective on data-driven approach for game development through the lens of motion-in-mind.

Keyword: Data-Driven, Game Development Life Cycle, Motion-in-Mind, Procedural Content Generation, Puzzle Solver

Acknowledgments

First of all, I would like to express my sincere gratitude to Professor Hiroyuki Iida for being my great advisor and supporting me in many things during my life not only during master's study but also during my doctoral study at Japan Advanced Institute of Science and Technology (JAIST). He taught me countless of great lessons which are not limited to research but also life in general. His guidance and supports helped me in all the time of research and writing of this thesis. If I did not meet him during his visit at UNSRI in 2017, I would have not be able to achieve all of this.

Secondly, I would like to acknowledge Professor Kokolo Ikeda, Associate Prof. Sakti Sakriani, Professor Jin Yoshimura and Dr. Anggina Primanita for being my committee. They gave me a lot of inputs and suggestions to improve my dissertation. I cannot thank Professor Mineo Kaneko and Dr. Saajid Abuluaih enough for being my Minor Research Advisor and supporting me through it. I am also grateful to Senior Lecturer Kotona Motoyama who gave me the opportunity to participate in her diversity studies projects and discussion and Prof. Shungo Kawanishi who gave me the opportunity to participate in ISA's Empowerment Programs which facilitated fruitful discussions with brilliant Japanese high school students on topics such as leadership, identity, and tackling global issues together.

Next, I would like to extend my gratitude to Assistant Professor Mohd Nor Akmal Khalid for his guidance and constant support while I am working on my thesis and research. Also, I would like to thank Mrs. Anggina Primanita for all the support she gave and also for introducing me to Professor Hiroyuki Iida in 2017.

Furthermore, I would like to thank my dear friends, The "Three Girls and Their Daddy" (PaPa, Bell and Saggu), Kangkang, Yamamoto-san, Kak Albert, Masaya, Gary, Kak Shofi, Candy, Amel, Mas Hasan and Mbak Dian for their unwavering support during both the ups and downs of my journey. Without them, I would not have mentally survived this challenging path. I also would like to thank Iida lab members, the Indonesian community (both in JAIST, Kanazawa and Awazu), JAIST Badminton club, my juniors and of course, my beloved Yosakoi team members, 粟津おすえべ花吹雪, with whom I have shared countless precious memories. Without all of them, my life would be much less vibrant. Additionally, I thank JAIST staffs for helping me with many things.

Finally, I must express my very profound gratitude to my grandma, parents, sisters and family for providing me with unlimited support and continuous encouragement through the years of my study and life. This accomplishment would not have been possible without them. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Chapter Introduction

In this chapter, as the main interest of this thesis lies in data-driven game development, we briefly introduced an overview of the state of the video game industry, the importance and use of data in game development life cycle and the problems that exists in both game academics and industry field. Finally, we summarize our contributions and the structure of this thesis.

1.2 Background

The game market is massive, it has generated US$196.8 billion in worldwide revenue and consists of over 3.1 million people consuming the game as reported in July 2022 [1]. One of the video game publishing platforms, *Steam*[2], with over 46 thousand titles in their library, is predicted to have over $8 billion revenue in 2022 [3] making them one of the most successful video game publishing platform that allows a multitude of game developers to publish their games as well as providing their players with an abundant choice of games from *Steam*'s library to enjoy. While the video game market grows larger, it becomes harder for a game developer to make their game stand out in such a competitive crowd as it requires the developer not only to produce such high-quality

[1] https://newzoo.com/key-numbers
[2] https://store.steampowered.com/
[3] https://www.statista.com/topics/4282/steam/

content and mechanics in their game to keep engaging for their players but also to be eye-catching on the game market to attract the players into purchasing their games. To produce such games, the game developer needs to do a lot of market research [92] before developing their games to gain insights into the kind of games that the players might be interested in as well as after they published their games to gain insights from the players to improve their games and keep the player attracted to their games because ultimately, these games are developed for the players. Ullmann et al. [98] investigated a set of factors (*team size, level of independence, game genre, game platform, graphical perspective, and development problems* that may impact a game's success from 200 video game project data they collected. Though none of the factors have a strong relationship with game success, their results show that higher-rated games present more occurrences of crunch time but have fewer problems with scope, delays, budget, and cutting features during their development phase. Another strategy that some game developers have adopted is to release an unfinished version of their game through early access to gather feedback from players. Lin et al. [62] conducted an empirical study of early access games on Steam platform. Although the result showed that using *early access* (EA) strategy does not lead to more satisfied players, they found that there exists a correlation between Early Access Games (EAGs) to higher positive review rate. They concluded that the communication between the game developer and the players of EAGs is crucial as the players enjoy and get emotionally involved in the decision-making of the game. This emphasizes the importance of communication and collaboration between game developers and players in the game development process.

Overall, the video game market is highly competitive and constantly evolving, making it essential for game developers to conduct thorough market research, gather feedback from players, and engage in effective communication with players to develop successful and engaging games. By understanding the factors that may impact a game's success, game developers can make strategic decisions about their game development and marketing strategies. Additionally, strategies such as early access can be used to gather feedback from players, which can help game developers improve their games and attract more players. However, it is important for game developers to strike a balance between gathering feedback and maintaining a stable development process to avoid delays and budget issues.

Ultimately, the key to success in the video game market is to understand the players' needs and preferences and to constantly strive to improve the game development process.

Historically, game development has not been data-driven[28]. Through out the year, as the game industry matures, the method used in game development and production processes also mature and are further optimized[82]. The uses of data-driven approach became an essential and integral part of game development life cycle to support decision-making across all the stages of the game development life cycle[41]. The benefit of data-driven decision-making can be seen in every game development life cycle stage. Although there are various methodologies to game development [4, 18, 89, 23, 88, 53], in her GDC Talk [36], Emily Greer (Kongregate[4]) discussed that the uses of data in-game life cycle could be split into two stages: creation, which covers concepting, pre-production, and production stages, and optimization, which covers testing or beta testing, launch and post-production or live operation stages. Eberhard et al. [30] investigated the helpfulness of video game reviews on the steam platform through a large dataset of video game reviews they extracted from *Steam*. The result shows that reviews voted as *helpful* tend to be longer, use more complex language, and expresses a more negative sentiment in which they are more critical towards the product and have deeper detail about the individual aspects of it. Moreover, Lin et al. [63] found that the developers need to pay attention to the design of their game's first 7 hours of gameplay, as most negative reviews are posted within that period. Data are utilized to answer the questions such as "What game should we make?", "How should we make it?" and "What can we do better?" which are crucial as they decide the direction of the game that will be developed as well as to learn from the success and failure of existing games. Meanwhile, on the optimization side, data are used to answer questions related to the game, such as "what's working and what's not working?" and "what can we make better?" as well as questions related to the players, such as "how the player engages with the game?" and "how can we keep players engaged?" which is crucial to understand how the players consume their games as well as to extend their games' lifespan. Rizani et al. [79] observed the active user and review data of Counter-Strike: Global Offensive (CS:GO) and Team Fortress 2 (TF2) before and after their business model transition from Pay-to-Play (P2P) to Free-to-Play (F2P). The authors found that the number of active users in both CS:GO and TF2 increased by 30%

[4]https://www.kongregate.com/

and 150% respectively but both games also received numerous negative reviews about the security or community aspect of their game from their players, which shows that although F2P transition strategy will result in the increase the number of active users, it can also bring unforeseen problems that the developer have to be prepared for. Martins Kummer et al. [66] presented a commitment-based approach to predict churn and remaining lifetime on CIG2017 Game Data Mining competition data. The authors proposed a new attribute on top of the original commitment-based approach that measures all activities related to players' engagement and generates new attributes based on them to identify a tendency behavior of the player. The result shows that the generation of attributes related to the tendency of each player action gave extra information to the models, allowing better performance in predicting churn value.

Due to the scale of video game industry, developing a successful game is challenging [63] and video games are consumer-oriented, where the success of a game relies on the end-user [74], thus making it crucial for the game companies to understand the consumer — the players of their games [70]. Despite the numerous benefit of data-driven approaches, many companies (not only game companies) find it challenging to derive meaningful insights from data [101]. This thesis focuses on utilizing game refinement and motion-in-mind theory as a data-driven approach in the stages of game development. The contributions of this thesis are indicators to measure the entertainment aspect and engagement in video games through game refinement and motion-in-mind theory.

1.3 Statement of Research Question

The thesis aims to investigate how game refinement and motion-in-mind theories can be applied in data-driven game development and how they can be used to measure the entertainment aspect and content quality of video games. Additionally, the thesis will explore the relationship between game refinement and motion-in-mind theory and arousal theory. The research questions that will be addressed in the thesis are:

- **Research Question 1:** How can we measure the entertainment aspect of video games from their Steam storefront data and how can we improve their visibility on the Steam platform? (Chapter 3)

- **Research Question 2:** What is the indicator to measure the difficulty of the contents as well as player performance in FlowFree?. How does the value of this indicator differ based on player type? (Chapter 4).

The main goal of this research is to gain a deeper understanding of how game refinement and motion-in-mind theories can be applied in data-driven game development, and how they can be used to measure the entertainment aspect and content quality of video games, which can be useful for game developers and researchers to create better and more engaging video games. The research will also provide insights into how to improve the visibility of video games on the Steam platform, which can be beneficial for game developers and publishers.

Overall, this thesis aims to provide a comprehensive understanding of game refinement and motion-in-mind theories in video game development and how they can be used to measure the entertainment aspect and content quality of video games. The insights and findings of this research can be applied in the video game industry to create better and more engaging video games for players.

1.4 Structure of The Thesis

This thesis comprises six main chapters, given as follows:

- **Chapter 1: Introduction**

  The objective of this chapter is to introduce the broad view of this research, such as the backgrounds and motivations, how each of the keywords relates to each other in the study. The introduction chapter also includes a statement of the research problem that the research aims to solve as well as the contribution and significance of the research. At the end of this chapter, the structure of the dissertation will be stated.

- **Chapter 2: Literature Review**

  The objective of this chapter is to serve as a review of the theoretical background related to this research and present state-of-the-art research in the field. The first section of this chapter covers existing data-driven techniques and applications, which will be split into the creation and optimization stages. The second section of this chapter covers the game refinement theory and motion-in-mind, a measure of entertainment in the game domain. Finally, a conclusion that justifies the research carried out in the dissertation will be presented at the end of this chapter.

- **Chapter 3: Application of Meta-Gaming Concept to The Publishing Platform: Analysis of The Steam Games Platform**

  The third chapter in this dissertation covers the result of the analysis of Steam data, a video game publishing platform. The contents of this chapter include findings on annual releases and multi-player support on the Steam platform, game prices' effect on games' ratings, the types of game developer and their strategies in publishing their games on the steam platform, as well as an in-depth analysis of Steam Games Achievements in term of how they are uniquely utilized depending the game's type, how they are consumed by the player and how they affect games' rating. Ultimately, this chapter provides an in-depth empirical analysis of Steam as a publishing platform and its game, where the data analyzed consists of a combination of public data provided by *Steam Store* and third-party steam statistic gathering service, *Steam Spy*, that provides a valuable metrics including the total owner's estimation on steam. Additionally, we will use *Steam Achievements* data to provide deeper insight into steam games. In this chapter, we will answer RQ1, which discusses the indicator to measure the entertainment aspect of video games from their Steam storefront data. Additionally, at the end of this chapter, we will discuss an opportunity for a new business model that utilizes games' auxiliary data (such as game type, rating, price and etc.) to increase players' engagement for the video game publishing platform and will be presented as a conclusion.

- **Chapter 4: Generating, Solving and Analysis of FlowFree Puzzles**

  The fourth chapter in this dissertation covers generating and solving puzzles in *FlowFree* game and analysis of *FlowFree* puzzles with Motion-in-Mind measures. Additionally, an experiment is conducted to collect human gameplay data and player-perceived difficulty data through a clone of *FlowFree* game we developed. The mechanism of *FlowFree* puzzle generation will be presented. At the end of this chapter, the result of analysis on *FlowFree* puzzles and its correlation to human gameplay and players' perceived difficulty and interestingness data will be discussed and presented as a conclusion.

- **Chapter 5: Conclusion**

  The last chapter is the conclusion of the dissertation. It concludes the whole dissertation relative to the main aim and objectives of the dissertation. Some potential future works are also outlined.

Chapter 2

Literature Review

2.1 Chapter Introduction

The objective of this chapter is to serve as a review of the theoretical background related to this research and present state-of-the-art research in the field. The first section of this chapter covers existing data-driven techniques and applications, which will be split into the creation and optimization stages. The second section of this chapter covers the game refinement theory and motion-in-mind, a measure of entertainment in the game domain. Finally, a conclusion that justifies the research carried out in the dissertation will be presented at the end of this chapter.

2.2 Data-Driven Game Development

Data-driven game development is defined by two concepts, "data-driven" and "game development". Data-driven is defined as making decisions based on analysis and interpretation of data rather than intuition only. Game development is defined as a process in which a game is produced, involving skills such as concept generation, design, build, testing and release. Therefore, Data-driven game development can be defined as a process of developing a video game that involves decision-making based on analysis and interpretation of data in its development cycle. In order to stand out in a highly competitive game industry, game companies leverage game-playing data to make game design-related decisions to provide players with more meaningful experiences in their games [111]. Thousands of games are developed by game companies and released across several platforms and distribution methods every year, although there are various methodologies for game development [18, 89, 23, 88, 53], the process of the game development life cycle can be categorized into three phases[4]: (1) pre-productions, which covers market research, concepting, etc, (2) production, which covers assets creation, implementation, etc, and (3) post-production, which covers launch, testing, and live operation.

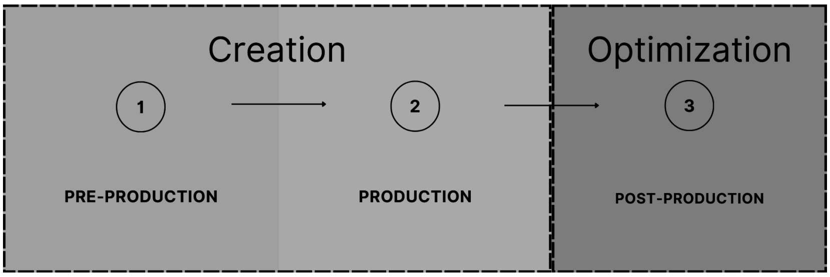


Figure 2-1: the uses of data in game development life cycle

In her GDC Talk [36], Emily Greer (Kongregate[1]) presented that the uses of data in a game development life cycle can be split into two stages: creation and optimization as shown in figure 2-1.

2.2.1 Creation Stage

The creation stage consists of pre-production and production phases. Pre-production is the process of testing the feasibility of a game, its requirement, and its design [33]. In most cases, the deliverable of the pre-production phase is a game design document (GDD) which is a blueprint for the design and development of a game. Some game companies would also develop a minimum viable product (MVP) [78] to demonstrate and test the essential features of the game early in the game development life cycle to save time and money before the company fully committed to developing the game [46]. The aim of the data-driven approach in this phase is to gain insights into the GDD or blueprint of the game that will be developed. Some key examples from existing work are presented below.

Traditionally, video games are only available in physical copy and players used to gain information about video games through magazines, gaming outlet, or their friends.

[1]https://www.kongregate.com/

However, more online marketplaces for video games have become more accessible and available on multiple platforms, video game players do not need to struggle to purchase or find reviews about the video game of their choice [6]. Reviews are essential for video game players to gain information about the game they are interested in and for game developers to gain insights for updates and future releases of their games [63]. Eberhard et al. [30] investigated the helpfulness of video game reviews on the steam platform through video game review dataset from *Steam* platform. The reviews section on *Steam* platform allows the players to write reviews on the games they have purchased, as it is also a medium for the players to communicate their thought to the game developer. The authors found that reviews with more helpfulness votes differ from the majority of the rest, tend to be longer, use more complex language, and tend to be more critical towards the product and go into greater detail about the individual aspect of the game.

On the same note, Lin et al. [63] performed an empirical study of the reviews of 6,224 games on the *Steam* platform. The author found that the developers need to pay attention to the design of their game's first 7 hours of gameplay, as most negative reviews are posted within that period. Additionally, the authors found that a large number of reviews for free-to-play games are posted approximately after one hour of playing hours, whereas negative reviews are often posted after only half the playing hours of the positive reviews, and players complain more about game design rather than bugs in their reviews. Guzsvinecz [37] investigated the correlation between positive reviews, game design elements, and mechanics of 21 soul-like games. The author found that factors such as medieval setting, 2D graphical dimensions, drawn graphical style, interconnected world, no difficulty settings, single-player mode, no weapon/armor upgrades, having equipment durability features, an in-game map, extra penalties upon death and not a classic level-up system will lead the player to leave a positive review on the game. Review is one of video game data that is commonly utilized by game developers in the pre-production phase because they are openly available, can be easily accessed, and contains rich information that can be processed using various approaches to gain insights such as important elements and factors of a game [37], and insights to improve the gameplay experience of the game [63].

Once the game's concept, design, and schedule are established, the production phase of the game development life cycle begins. The production phase executes the design and planning of the pre-production phase, which takes most of the time and budget of the game development process [33]. The production phase involves a wide range of tasks and activities, including programming, art and animation, sound design, and testing. This phase is critical to the game's success, as it determines the game's final quality and player experience. Some key examples from existing work are presented below.

Procedural Content Generation (PCG) is a technique for creating content algorithmically. This content can be anything such as terrain[72], levels[103], stories[54], quests[55], characters[32], rule-set [14] and other contents of the game that affect gameplay other than nonplayer character (NPC) and the game engine itself [95]. PCG allows content to be generated automatically and it can greatly reduce the amount of time and money that a designer/company can spare in their game development process[100] as well as increase game replayability for the players with a continual introduction to novel contents in game [38]. Spelunky [2] is an example of a published game that implemented a procedural content generation system. The game procedurally generates different levels every time you play, which makes the game have infinite replayability value, and the player will always experience different challenges. Although PCG allows automatic content generation in games, the content generation is often random—adjusting the content according to user needs and preferences are essential steps toward effective and meaningful PCG [110]. Stammer et al. [85] adapted a dynamic difficulty adjustment (DDA) system in Spelunky based on the survey they have collected. The authors did a user study where 58 participants were randomly assigned to one of three groups A (no adjustment), B (only difficulty adjustment) and C (both player profile and difficulty are estimated and adapted). Overall, the result of the experiment survey showed that there are high "good" and "very good" ratings from group B and even higher ratings from group C. Another example is the random encounter system in "No Man's Sky" [3], a game that procedurally generates a virtual universe, including planets, creatures, and plants, which allows players to explore new and unique environments every time they play.

In Drachen et al. [28], Georg Zoeller [4], discussed the *SkyNet*'s telemetry system to track

[2]https://spelunkyworld.com/
[3]https://www.nomanssky.com/
[4]https://www.bioware.com/

player behavior as well as developer's usage for the purposes of evaluating and enhancing not only game design but also the production pipeline, quality assurance methods, and workflow. One of the lessons they learned is that it is important to always test your hooks if you plan to make decisions based on the data. Important hooks need to be retested frequently. "Game hooks" refers to a feature or part of the game that is intended to draw the players' intention to immerse them further into the game [68]. Data-driven approaches can give benefits such as dynamic difficulty adjustment of the game [116, 42, 85], game experience balancing [39, 69, 95, 110], understanding players' behaviors [21], reducing the cost of the game production [56, 38] and so on.

2.2.2 Optimization Stage

The optimization stage consists of post-production phase, where it involves making final adjustments and fixing any remaining issues or bugs in the game as well as implementing data analytics and making data-driven decisions to optimize the game where improvements can be made to enhance the user experience.

Aung et al. [9] analyzed play history of 5,000 players of *Just Cause 2* to profile the spatio-temporal behavior of the players using DEDICOM. One of the main objective the authors investigated is the behavioral differences between player types defined by early abandonment and commitment, where they found out that Early Dropouts (players who stopped shortly after starting to play) selected a difficulty higher than most other players, which subsequently made them into stopped playing even though they are only in the very early stage of the game. The insights that we can gain here is that misunderstanding or mismatching between players' ability for game difficulty might cause the players to quit the game in the very early stage of the game.

The aim of the optimization stage is to create a more polished, engaging, and profitable game by improving the user experience, while addressing any issues that may be preventing players from fully enjoying the game.

2.3 Game Refinement Theory

Game refinement (GR) theory has been studied the game outcome uncertainty, where game dynamics are evaluated based on an innovative view on the outcome uncertainty of the game simulated via the analogy of Newton's law of motion [44] [43]. It has been evaluated in the domain of game such as board games and sports games, also studied in non-game domain such as education and business. Game refinement theory fundamentally involves the measures that define the game sophistication that converges towards a common range, where the most stochastic game located in $GR \in [0.07, 0.08]$, corresponds to the lower bound (fairness) and upper bound (engagement), respectively. Later, it has been involved to measure the attractiveness of a game [90], where the lower bound and upper bound are corresponds to the game that more relies on skill and chance, respectively.

From game playing point of view, the information on reaching a game outcome for a player is regarded as a function of time t, and the information on the game results is regarded as the solved uncertainty (information) $x(t)$. In other words, the process of solving the uncertainty is an increasing function of time achieving such an outcome. Then, (2.1) is obtained to illustrate the velocity in game, where the parameter $n(1 \leq n \in \mathbb{N})$ is the number of possible options (branching factors), the parameter t is the game length, which is depends on players.

$$x'(t) = \frac{n}{t}x(t) \tag{2.1}$$

However, such a formulation implies that the game outcome is known. In reality, the game outcome is unknown until the game ends. As such, a realistic formulation considering the uncertainty of the game outcome is given by (2.2). Note that $0 \leq t \leq T$ and $0 \leq x(t) \leq 1$, and $x(0) = 0$ and $x(T) = 1$. Here, from the game objective point of view, the game length is assumed as T, and the game outcome is regarded as $x(T) = 1$.

$$x(t) = \left(\frac{t}{T}\right)^n \tag{2.2}$$

The velocity in a game process can be seen as Equation (2.1), based on the accelerated velocity in physics is used to describe changes in velocity, Equation (2.3) is given to illustrate the rate of change of the solved information $x(t)$ of the game progress, where the solved information of (2.2) is assumed to be twice derivable at $t \in [0, T]$. This implies that game is fascinating if this value increases or decreases, it will make the game even more fascinating and entertaining. thus, this character is considered to be the one that deserves the most attention in a well-refined game domain. This thesis is studied in determining the deterministic and stochastic characters in puzzle field, which is a popular single agent game domain.

$$x''(t) = \frac{n(n-1)}{T^n} t^{(n-2)} \mid_{t=T} = \frac{n(n-1)}{T^2} \tag{2.3}$$

Then, the GR measure is given as (2.4) in the root square of Equation (2.3), the accelerated velocity. This measure has been verified to reflect some aspects of the entertainment of games, such as attractiveness, engagement, and playing comfort. This thesis focus on doing entertaining analysis with the game refinement theory in the puzzle domain.

$$GR = \frac{\sqrt{n(n-1)}}{T} \tag{2.4}$$

2.3.1 Gamified experience for board games and sports games

Based on the study of game refinement theory, the ratio of solving uncertainty at different depths is given as v, and the solved uncertainty of the game $y(t)$ is an increasing function of time t, which can be given by (2.5). Let p be the probability of selecting the best choice among n number of options (branching factors). Hence, $p = \frac{1}{n}$ holds the moving velocity in a game. Based on such notation, the risk frequency ratio m (risk frequency over the whole game length) is defined as $m = 1 - p = 1 - v = \frac{n-1}{n}$. Then, gamified experience is gained only when the risk of failure occurs with $m \geq \frac{1}{2}$, which implies $n \geq 2$, and has been verified kinds of fun games.

$$y(t) = vt \tag{2.5}$$

The slope (v) with the time (t) of a game progress model in (2.5) and mass in the game playing have been determined in two domains: (1) scoring sports games, and (2) board games. For scoring sports game, let G and T be the total scores of goals and shoot attempts per game, respectively. Score rate v (the total scores of goals over the shoot attempts per game) is given by (2.6), where the slope v ($v = p$) of game progress model is equivalent with score rate in (2.5). Note that the score rate v in some sports (e.g., table tennis, badminton, soccer) is given by $v = \frac{1}{2}$, this situation is because one would have a point with the possibility of $\frac{1}{2}$ at each round.

$$v = \frac{G}{T} \quad \text{and} \quad m = 1 - v \tag{2.6}$$

For board games, let B and D be the average number of possible moves and game length. Score rate p is approximated as (2.7), by which p is equivalent with the slope v ($v = p$) of game progress model in (2.5). Note that the v in board games is approximated based on the number of plausible moves b, where $n \simeq \sqrt{B}$ is used in the best-case analysis of an efficient $\alpha\beta$ algorithm that is useful for pruning.

$$v \approx \frac{1}{2}\frac{B}{D} \quad \text{and} \quad m = 1 - v \tag{2.7}$$

2.4 Motion in Mind

When players play games from the beginning to the end, the game progress can be treated as solving uncertainty. In other words, the game is full of uncertainty at the beginning, as the player play the game and the game process moves forward, the game's uncertainty information becomes less until zero at the end. Over time, the process of playing the game is one of decreasing uncertainty. In a puzzle game, the game information becomes certain when the player gets solutions to solve the puzzle.

Similar to physics in the world, vital physical quantities in mind are the velocity and mass, with the assumptions of $v + m = 1$ which are based on the zero-sum assumption, where gain or loss utility of one player is exactly balanced by the losses or gains of the utility of its opponent; thus deriving a reliable measurement of players' game experience, such as engagement and comfort [43]. Moreover, in puzzle games, different levels players may choose differently at each step based on skills, differing in velocity to move and solutions to solve the puzzle; portraying different solve experiences, such as attractiveness and engagement.

By analogically defining the game-winning (or success) rate and winning hardness (or difficulty) as the velocity (v) and mass (m), respectively, various motions in mind quantities can be determined [43]. Table 2.1 provides the analogical link of the related physics in mind notations and its in-game context (specific to the current study).

Table 2.1: Analogical link between physics and game (adopted from [43])

Notation	Physics context	Game context
y	Displacement	Solved uncertainty
t	Time	Progress or length
v	Velocity	Win rate (p)
M	Mass	Win hardness (m)
g	Acceleration (gravity)	Acceleration, a
F	Newtonian force	Force in mind
$\vec{p}$	Momentum	Momentum
U	potential energy	Potential energy, E_p

As the table shown, the displacement (y) in physic corresponds to the solved uncertainty in game context, and time (t) stands for the game progress or length in game domain. Force is determined as a product of mass and acceleration ($F = ma$), which relates to the acquiring engagement of player's movement ability in the game playing, from Newton's second law of motion. In classical physic, the gravitational potential energy U is given by (2.8) where g and h stand for gravitational acceleration and height (or displacement), respectively. Then, the potential energy (E_p) given by (2.9) can be obtained by the correspondence of $M = m$, $g = a$, and $h = y(t)$, where m and a stands for the win hardness and acceleration in game, respectively. A game's energy is defined as the amount of the required information (energy) needed by the playing in the game process, which is equivalent to the expectation of player to finish the game or the anticipation that the player expect the game give.

$$U = Mgh \tag{2.8}$$

$$E_p = ma\left(\frac{1}{2}at^2\right) = \frac{1}{2}ma^2t^2 = 2mv^2 \tag{2.9}$$

Meanwhile, the notion of momentum in game-playing process is given by (2.10), which defines the product of m and v, which is the moving difficulty (or hardness) and ability to move, respectively. This equation states that momentum (p_1) is directly proportional to the velocity of a game, and directly proportional to the mass of a game. In other words, such quantities describe the freedom magnitude of the player to use their ability to address the difficulty in games. Note that momentum in game playing is relied on two factors: the game progress ratio v and the hardness to move in a game m.

$$\vec{p_1} = mv \tag{2.10}$$

The game experience depends on the game itself (objective), but also on the player (subjective) such as skill, experience. Assumptions of both momentum and mass as the manifestation of energy lead to the discussion on the notion of potential energy (E_p) being conserved over time [50]. Then, such energy is transformed into the game's momentum ($\vec{p_1}$) and the mind's momentum ($\vec{p_2}$) of players, as given by (2.11). And, the $\vec{p_1}$ is considered the objective point of view, whereas the $\vec{p_2}$ is from the subjective point of view. The former is associated with the game's motion, while the latter is associated with the player's play experience [50], which is obtained based on equations (2.10), (2.11) and (2.12).

$$E_p = \vec{p_1} + \vec{p_2} \tag{2.11}$$

$$\vec{p_2}(m) = E_p - p_1 = 2m^3 - 3m^2 + m \tag{2.12}$$

$$\vec{p_2}'(m) = 6m^2 - 6m + 1 \tag{2.13}$$

Then, (2.13) is obtained by the first derivative of (2.12). Solving $\vec{p_2} = 0$, then $m = \frac{3\pm\sqrt{3}}{6}$ is obtained. It was conjectured that $m \simeq 0.79$ is the upper limit for competitive play mode, where $m \simeq 0.21$ is the lower limit for easy-win mode associated with the addictive zone (Figure 2-2). Each limit value corresponds to risk-taking engagement and profit-winning engagement, respectively. Interestingly, the cross point of $\vec{p_2} = E_p$ occurred when $m = 0.5$, which implies the moment where the game's motion is the greatest while the mind's motion is non-existence since E_p reflects energy conservation of objective and subjective motions. That means the game experience becomes fully stochastic, and predicting the game outcome becomes impossible.

2.5 Chapter Summary

In this chapter, related work prior to the current thesis were introduced. Works related to the important keywords, such as data-driven game development where data played

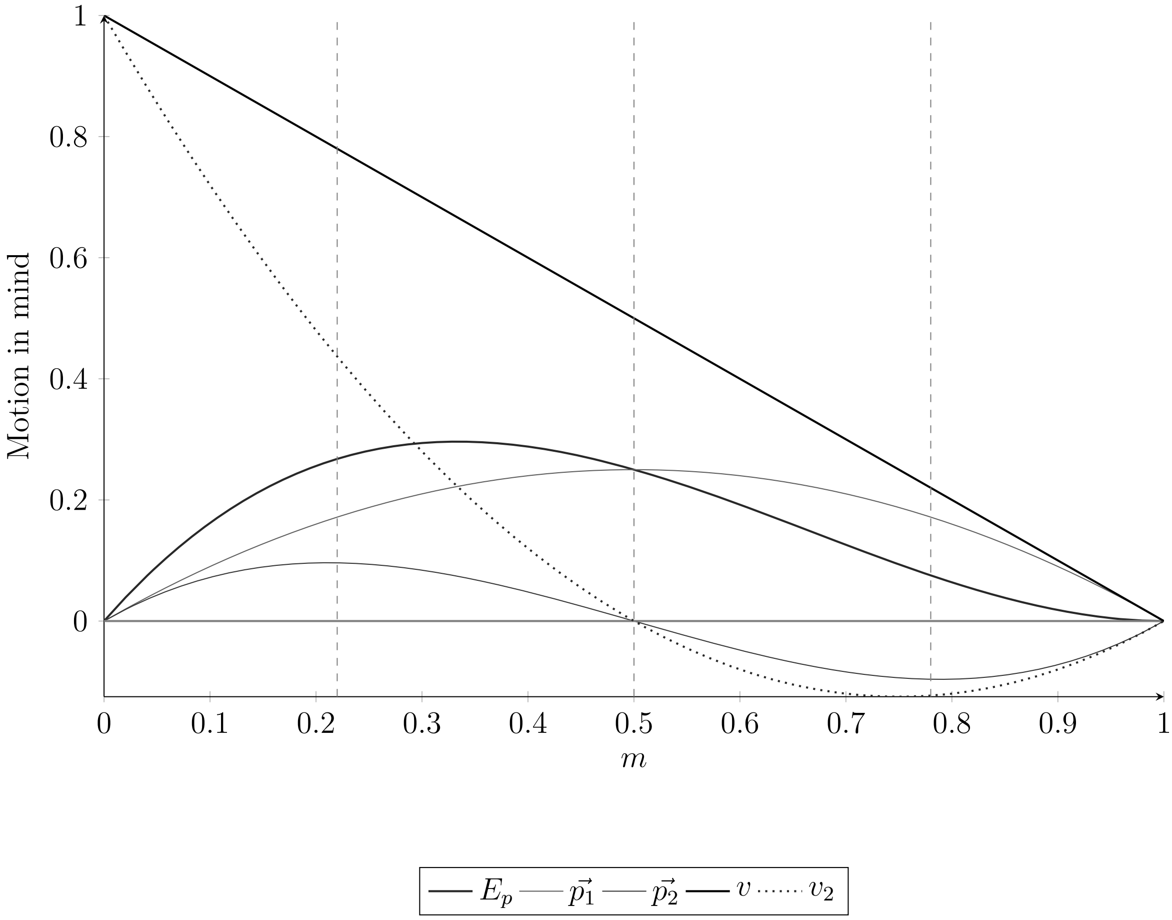


Figure 2-2: Illustration of law of motion in mind over various mass (m). The subjective motion(p_2) is derived from the objective ones(p_1), where subjective velocity (v_2) was established. $\vec{p_2}$ is derived based on the conservation of E_p.

meaningful role in different game development life cycle stages. In relation to the entertainment aspects, a measurement of entertainment of the game that highly depends on the uncertainty in the game, the Game Refinement Theory and Motion-in-Mind Theory, are introduced. These studies are significant as it serves as the base to the research carried out in this thesis.

Chapter 3

Application of Meta-Gaming Concept to The Publishing Platform: Analysis of The Steam Games Platform

This chapter is an updated and abridged version of the following publications:

- M. N. Rizani, S. Thavamuni, M. N. A. Khalid and H. Iida. (2021). Steam Game Achievement Analysis. The First Artificial Intelligence and Entertainment Science Workshop (AIES 2021), pp. 67-69, 2021.
- Muhammad Nazhif Rizani, Mohd Nor Akmal Khalid and Hiroyuki Iida. (2022). Application of Meta-Gaming Concept to The Publishing Platform: Analysis of The Steam Games Platform. Information 2023, 14, 110.

3.1 Chapter Introduction

The third chapter in this dissertation covers the result of the analysis of Steam data, a video game publishing platform. The contents of this chapter include findings on annual releases and multi-player support on the Steam platform, game prices' effect on games' ratings, the types of game developer and their strategies in publishing their games on the steam platform, as well as an in-depth analysis of Steam Games Achievements in term of how they are uniquely utilized depending the game's type, how they are consumed by the player and how they affect games' rating. Ultimately, this chapter provides an in-depth empirical analysis of Steam as a publishing platform and its game, where the data analyzed consists of a combination of public data provided by *Steam Store* and third-party steam statistic gathering service, *Steam Spy*, that provides a valuable metrics including the total owner's estimation on steam. Additionally, we will use *Steam Achievements* data to provide deeper insight into steam games. In this chapter, we will answer RQ1, which discusses the indicator to measure the entertainment aspect of video games from their Steam storefront data. Additionally, at the end of this chapter, we will discuss an opportunity for a new business model that utilizes games' auxiliary data (such as game playability type, rating, price and etc.) to increase players' engagement for the video game publishing platform and will be presented as a conclusion.

The video game market has exploded in the digital marketplace as one of the rapidly growing digital industries where it is estimated that 2.3 billion gamers across the globe will spend $137.9 billion on games in 2018 [97]. Furthermore, as one of the biggest game publishing platforms, Steam had 120 million monthly active users and over 50 thousand games on their catalog as of 2021[1]. Furthermore, with the transition of the video game market from isolated local experiences to more networked ones, millions more users can access the internet for an expanded universe of gamers' games and virtual communities. In ever-competitive and expanding business markets, satisfying such a growing consumer base generates massive data. Therefore, data-driven analysis is becoming an essential tool for analyzing consumer behavior, which is helpful for gaming developers, marketers, and streaming platforms [28, 97].

As a rapidly growing game publishing platform, Steam's popularity became well-known for the well-established game studios and independent game developers, typically known as the "indie" developers (or indie studios). The shift of game production using "free" and accessible all-in-one game engines had dominated the market for the development of game products and services [112]; allowing more flexible 'open-close' production[2] that being supported by the Steam's platform.

[1]https://backlinko.com/steam-users

[2]The 'opened' production game makers are where multiple professional and leisure-based game-making identities were shared, and 'closed' production was adopted under platform governance policies, proprietary technical requirements, and multisided market strategies [112].

Since creating video games is a lengthy and demanding process [67], which could cause management and production problems in the same proportion [73], over-bearing and over-confident of developers [31, 73], and even requires urgent updates[3] [61]. As such, a game studio's financial success often depends on providing exciting experiences and access to a diverse audience, especially in a densely populated platform such as Steam. Such risks were relevant to both small and large development studios, highlighting the importance of knowledge support and understanding the current market situation. Steam's game review is one of the sources for knowledge wealth on discerning the suitable monetary models to be adopted [79] and everyday needs of players and flaws in existing games [96]. Nevertheless, an epistemological problem has occurred where developers must distinguish between actual contributors to improvements in the game and those that merely express their subjective wishes, especially when community reviews are regarded as a form of user feedback.

Game distribution platforms, such as the Steam platform, are expected to provide continual improvement on the games owned by the players. In such a context, incorporating feedback from players is paramount [96]. However, making sense of the overwhelming volume of data available on such a platform to discern useful or beneficial content is challenging, making values from insights given in the feedback overlooked by developers and publishers alike [61, 96]. As such, helpful attributes on the Steam platform can be a valuable tool to uncover players' underlying intentions and wishes while identifying beneficial insights into the fast-paced behavior of the digital marketplace.

Considering data analytics in game-playing, the motion in mind model has been used to induce subjective association based on the objective matrices in the game-playing process [43]. By adopting the analogy of motions and physical phenomenon, information progression from uncertainty to certainty can be modeled, representing the ratio of difficulty (or challenge) to solve such information [43, 50]. Associating such information analysis with the empirical data from the Steam platform may uncover the underlying characteristics of players' interactions and trends of their game content consumption.

Therefore, this study aims to provide an in-depth empirical analysis of Steam as a publishing platform and its games. The data that will be analyzed consists of a combination

[3] An urgent update is a software update that fixes problems deemed critical enough not to be left unfixed until a regular-cycle update.

of public data provided by *Steam Store* and third-party steam statistic gathering service, *Steam Spy*, providing valuable metrics, including the total owner's estimation on Steam. Moreover, one feature of the Steam Platform is the *Steam Achievements* of the games. The developer can set in-game goals for the players to achieve, and players are rewarded with an achievement emblem to be showcased on their Steam profile. The achievement can also be extrinsic motivation to motivate players to explore game's content [58]. In addition to the RQ1 stated in Chapter 1, this chapter attempts to address the following one research question (RQ) and four sub research questions (SRQs):

- **RQ1:** How can we measure the entertainment aspect of video games from their Steam storefront data and how can we improve their visibility on the Steam platform?
- **SRQ1:** What kind of game releases & multi-player support on the Steam platform?
- **SRQ2:** How does rating affected by game prices on the Steam platform?
- **SRQ3:** What kind of developers & publishers were dominant on the Steam platform?
- **SRQ4:** Does Steam achievement affect game rating & type of games?

3.2 Steam and related works

3.2.1 Steam

Steam is a video game digital distribution service and storefront by Valve. It was initially released in 2003 as a way for Valve to provide automatic updates for their games but later it is expanded to what currently is in late 2005. In recent years, Steam has become a popular digital game distribution platform that has drawn much attention from academia. First developed by Valve corporation, Steam offered services related to digital distribution, digital rights management (DRM), multiplayer gaming, and social networking [48, 63]. It became the world's largest gaming platform and started with an official release on September 12, 2003. Various game genres were listed in Steam, including first-person shooters (FPS), role-playing, racing, and even independent games for their digital management and distribution (i.e., Indie games). Steam is a cross-platform that supporting multiple gaming environments [97]. Games distributed on steam platform have their own store page where it can be utilize to attract player into buying the game by adding videos and screenshots of the game, including the reviews from the users that have already bought the game as shown in Figure 3-1.

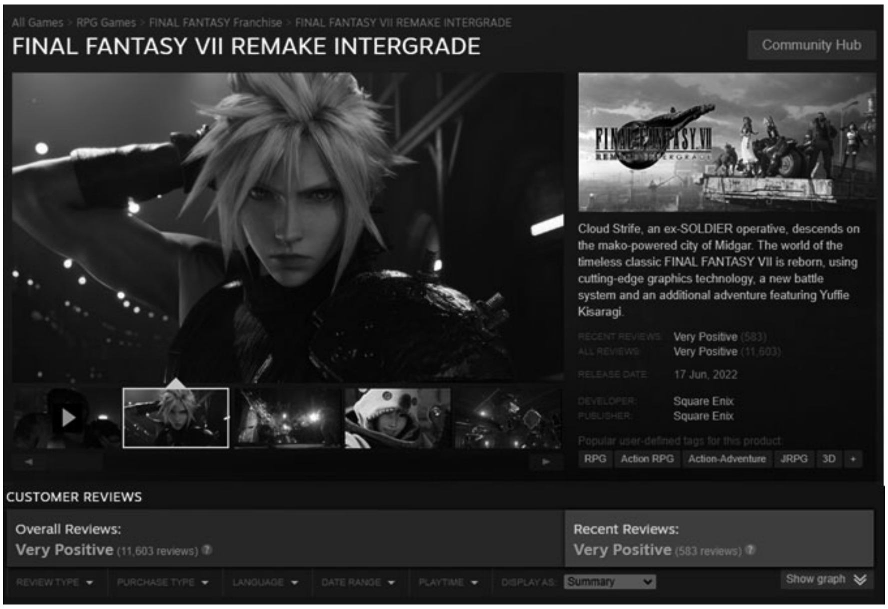


Figure 3-1: Steam game store page [4]

The Steam platform users interact with it via a local Steam client, available for an operating system such as Windows, Mac, and Linux. Games can be purchased from the Steam Store or third-party vendors, which are then activated through the Steam platform and playable after logging in on Steam using the Steam client [63]. Payments are available in various currencies, and licenses are registered to the user library [48]. Games ownership (or license) and updates will be automatically verified and installed since it is mandatory to play a game through Steam. Users can easily download and enjoy games from their library with their account information anywhere at any time. Additionally, the steam user able to showcase the game they owned, achievements they have received and many other things on their steam account page 3-2 as well as to write comments on other steam user's account similar to some features in other social media platform (such as Twitter[5], etc).

[5]https://twitter.com/home

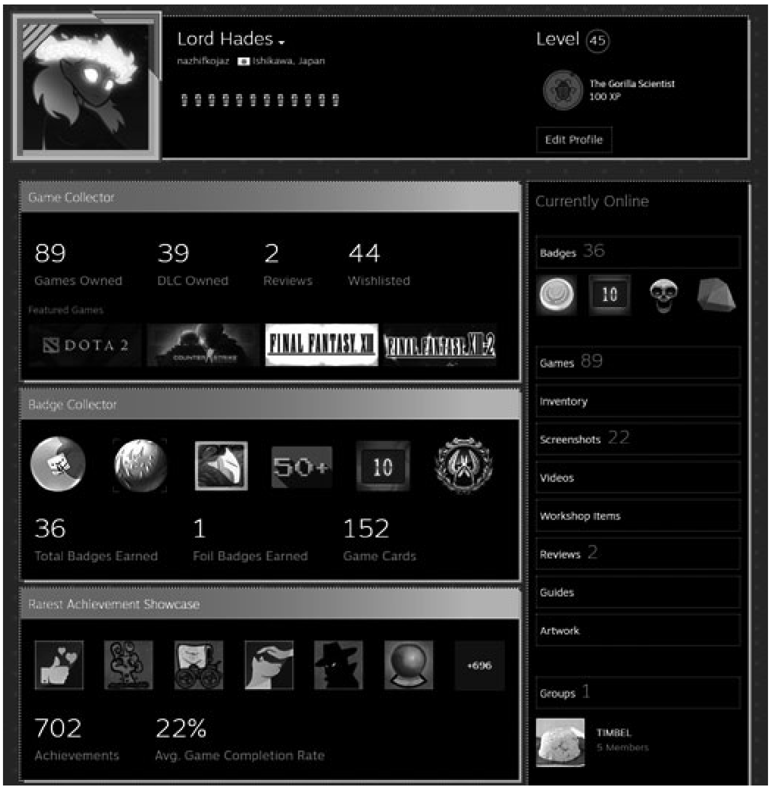


Figure 3-2: Steam user profile page [6]

3.2.2 Related works

Several veins of research had been conducted on the Steam platform and database. For example, Windleharth et al. [108] describes a conceptual analysis of all user-generated tags applied on video games in the Steam video game distribution system, where the categories were sorted and compared to the video game metadata schema, where emergent terms beneficial to players were presented and discussed to uncover issues in organization and its implications for the future work. Then, Li and Zhang [57] propose an alternative approach to understand video game genre classification via preliminary network analysis of the user-generated game tags on the Steam platform using centrality analysis and community detection. Such an approach is intended to lay the ground for and encourage the further investigation of the intertwined connections between genres, inconsistently defined abstraction levels, and different user focuses. Meanwhile, some researchers focus on game reviews' reliability and their implications for game development. For instance, Kang et al. [48] identify factors affecting the helpfulness of reviews uploaded by users on the communities by analyzing unrefined game data via data mining techniques, such as classification and regression trees (CART) and a multi-layered perceptron network, to predict the most significant variable in Steam community's game review. Moreover, Eberhard et al. [30] conducted on the steam games reviews dataset to discover aspects of reviews usefulness from the community point-of-view. They extracted the text bodies from the review, obtained 58 different features from the reviews, and separated them into three categories (unhelpful, helpful, and top review) based on the number of votes they received from the community. They found that reviews with more helpfulness votes tend to be longer, use more complex language and express more negative sentiment, as well as more critical towards the product and go into greater detail about the individual aspect of the game. While these reviews tend to be longer or critical about the product, Eberhard et al. [30] also found that there exists reviews with a large number of helpfulness vote with a short or meaningless text where the number of votes derived from humor or the author being a popular personality. Meanwhile, Lin et al. [63] performed an empirical study on games reviews on the Steam platform to have a better understanding of the user-perceived quality in games where the number and the complexity of reviews, the type of information that is provided in the reviews, and the number of playing hours before posting a review is analyzed. Based on the results, it was found that positive and negative reviews provide helpful insights, and their association with playing hours is unique between different dimensions of game types while being distinctive compared to the mobile app reviews. Busurkina et al. [17] utilizes the netnography research that adopted a Structural Topic Model (STM) to evaluate game-playing experience based on player reviews on the Steam platform. Seven dimensions were identified, which can be disentangled to generate more knowledge on the evaluation processes and the game itself. The findings extend the comprehension of consumer retention mechanisms and better understand users' motives and criteria in comparing games.

From another perspective, some work also explored user profiling to determine the players' specific behavioral characteristics or personalized content delivery. The first comprehensive analysis of hardcore gamer profiling was conducted on a dataset of over 100 million Steam platform users, with over 700,000 hardcore players (users playing more than 20 hours per week). It covers over 3,300 games using a k-mean clustering algorithm to determine the specific behavioral categories of hardcore players [13]. The results identify six hardcore gamers' behavioral clusters, where some were related to the sense of motives, consciousness, and openness to experience depending on the game genres played. Meanwhile, Li et al. [58] uncover the underlying structures of the Steam user profiles using exploratory factor analysis to define the player's preference and personalized behavior characteristic of the Steam community. Finally, Vihanga et al. [104] conducted a study to explore player population fluctuations within online games to identify weekly seasonality, archetypal weekly population patterns, and relative frequency of these patterns from an extensive steam player population data of 1,963 games. The study identified that 77% of games displayed a recurring weekly pattern clustered into nine diverse weekly player population fluctuation patterns. Out of nine clusters, the two highly similar dominant clusters indicated that most games display a weekly pattern where the player population increase towards the weekend.

Other aspects of the players were also explored in conjunction with the market influences, business model, and decision-support system. Toy et al. [97] discovered patterns among game ownership, genre, and geographical region from a vast Steam database via basic Heat map and clustered Heat map analysis. The result analysis revealed several interesting patterns, trends, and correlations of popular genres in the gaming industry (i.e., action games), shifting of current market practice and strategies (i.e., early access), and potentially lead to improved markets, business models, and a more responsive market in general. Ranti et al. [76] proposed a k-prototypes algorithm that integrates both k-means and k-modes algorithms to cluster mixed numeric and categorical attributes of Steam's user behavior telemetry data (40% or more of their total accumulative playtime) from the World of Warcraft game, resulting in three groups of a total of 15 clusters. It was found that there is a good correlation between sales data from the sample and actual sales data reported by game development companies. Also, better insights into the play patterns of the games bought and played by steam users, patterns about the user themselves, and the importance of differentiating users (i.e., doubling the player base does not double the revenue). Also, Wang et al. [106] proposed a solution for a new video game recommendation system for the Steam platform called STEAMer, which utilizes the Steam user data and applies additional user data in conjunction with a deep autoencoder learning model to generate potential recommendations. Performance evaluation included comparing STEAMer with a baseline deep neural network-based system. The results showed that adding additional public Steam user data has a noticeable and positive effect on the game recommendation with a noticeable increase in the test metrics over the traditional deep neural network using the same features. furthermore, Ahmad Kamal et al. [3] conducted implementation of genre-based and topic modelling model in a recommender system to predict rating of games using public steam dataset. Though the result shows that genre-based model outperforms topic modelling model, it doesn't outperform the model performance from the previous research. Therefore, they concluded that genre is not a suitable parameter for recommending games.

Other related research on Steam data and platform includes determining network feasibility of the Steam In-Home streaming services in comparison to the regular network infrastructure [15], discovering security vulnerability to serve as a guideline for computer forensic for Steam game platform [91], and determining the impact of shifts of business model changes [79], based on the analysis of Steam review data. However, limited studies were conducted on the Steam platform, focusing on game-level analysis, which provides valuable knowledge and intuitive insights for the game developer.

Studies on game-level analysis were also conducted on the Steam platform focusing on different perspectives. Some studies explored the Steam platform in conjunction with other distinct platforms (i.e., Twitch.tv that focuses on streaming) and their influences on a specific game experience. For example, Gandolfi [34] visualized the dynamics and trends of game platform analysis mediated by network-oriented software Gephi on a role-playing game Dark Souls 3, along with an exploratory counter-example using the action game The Division. It was found that such a media trend generates two different reactions: a positive one when the game is no longer just a game but a performance to watch, and a negative one when the interactive affordances were questioned. Meanwhile, Lin et al. [62] conducted an empirical study on the characteristics of 1,182 Early Access Games (EAGs) where the interaction between players and developers of EAGs, and the Steam platform during and after leaving the early access stage, and the tolerance of players of the quality of EAGs are analyzed. The study found that EAGs tend to be "indie" games (adopted by smaller development studios), and lower reviews were given during the early access stage compared to the review after leaving the early access stage, whereas the rating is vice versa. Bailey and Miyata [12] conducted data mining on the Steam "achievement" data of the video games in the Steam platform to discover trends in the game completion rates and correlates to the factors outside of the game's length. The study found that the completion rate can indicate the rate of players completing the game content and provide a benchmark for future scoping decisions in individual projects, which influences the game development decisions and success rate.

Li et al. [59] analyze and evaluate the playability of video games by mining players' opinions from their reviews guided by the game-as-system definition, where sentiment analysis, binary classification, multi-label text classification, and topic modeling are sequentially performed. A total of 99,993 player reviews on the Steam platform were evaluated, which focused on the collective opinions relative to the maintenance and evolution of video games and helped game developers to understand it. Ullmann et al. [98] investigated the aspects that describe a high-rated game through 200 video game projects on the Steam Platform. Though genre, graphical perspective, game modes, and platforms do not correlate to ratings of the games, the study found that games from smaller teams are often linked to higher ratings. Additionally, they analyzed post-mortems discussed by the developers of high-rated games. Furthermore, Du [29] conducted a study to predict whether a game on steam is on discount or not using machine learning methods through data collected from the steam database. The study compared Logistic Regression and Random Forest Classification and concluded that Random Forest reaches the top performance of 79.5% accuracy. This model will benefit players by allowing them to purchase a game at the right time while saving their money and for game publishers to optimize their discount strategies. Meanwhile, Badoni et al. [10] conducted observation based on a survey from 315 participants regarding which aspect of the game (graphics, gameplay, mechanics, and audio) is the most attractive among various desktop and mobile games. Based on the survey result, gameplay and graphics are mutually beneficial. In summary, the related works and their contribution summary relative to the current study were provided in Table 3.1.

Relative to the physical motion, the motion in mind concepts takes the concept further by adopting motion formulation to describe the entertainment aspects of games from the objective and subjective standpoints [50], based on the basic assumption of move selection,

Table 3.1: Literature summary of the previous works conducted on Steam platform

Citation	Contribution summary	Relation to our works
[108]	Conceptual analysis	Data features
[48]	Predict significant reviews	Data features
[57]	Genre classification	Data features
[15]	Networking feasibility	Data features
[30]	Useful review features	Data features
[91]	Discover security problem	Data features
[13]	Behavioral categorization	User profiling
[34]	Media influence on platform	Decision support
[97]	Market pattern	Business modeling
[62]	Early access impact	Business modeling
[63]	Perceived quality from review	Data features
[58]	Personality and preferences	User profiling
[104]	Play pattern	User profiling
[12]	Development decision	Decision support
[17]	User motivation from review	Data features
[76]	Play and purchasing pattern	Business modeling
[79]	Impact of business shift	Business modeling
[106]	Game recommendation	Decision support
[3]	Predict game rating	Decision support
[29]	Predict game discounting	Decision support
[59]	Playability based on review	Data features
[98]	Aspects of highly-rated games	Data features
[10]	Attractive aspect	Decision support

game progression, and the ratio of winning (m) and challenge (m) is equalized ($v+m=1$) [43]. The motion-in-mind model had been previously adopted to identify meta-gaming elements from the perspective of game evolution and its influence on culture [109], linking entertainment with the game-tree search processes [75], educational structure [7], process fairness [8], defining game features that make it addictive [49], and bridging comfort in physical to the comfort in mind [114]. More recent work takes the motion in mind concept to identify the mechanisms to retain entertainment in long-term arcade games [35], and key entertainment aspects (challenge, anticipation, unpredictability) between different God of War series [115]. This previous study showed that the motion-in-mind concept provides a versatile metric that is suited for analyzing varying aspects of meta-gaming elements of the Steam publishing platform, in addition to other conventional analyses, which serves as the primary motivation of the study.

3.3 Methodology

3.3.1 Motion-in-Mind

Considering the zero-sum assumption in game playing[7], the essence of uncertainty can be determined [50]. In the schedule of reinforcement of the operant condition originally designed by Skinner [84], a variable-ratio schedule is a reinforcement schedule where the response is reinforced after an unpredictable number of responses, creating a steady, high rate of responding [45]. From a reward-driven standpoint, this condition is a typical example of a reward system based on a variable ratio schedule found in stochastic games (such as gambling and lottery games).

Meanwhile, mind sports games (such as chess and Go) are essentially stochastic games when applying the move selection model [43]. This condition implies that a game is characterized by a reward of a variable-ratio reinforcement schedule. In essence, the game is characterized by the reward function, a variable rate (denoted as $VR(N)$) of the reinforcement schedule. Then, velocity v (win rate) and mass m (win hardness) of the motion in the mind model are given by (4.1).

[7] zero-sum assumption can be defined as the gain or loss utility of one player that precisely equalized by the losses (or gains) utility of its opponent [65]

$$v = \frac{1}{N} \quad \text{and} \quad m = 1 - v, \quad \text{where} \quad 1 \leq N \in \mathbb{R} \tag{3.1}$$

Table 4.1 describes the analogy of the motion in mind model from the physics and games context. Note that there is a distinctive computation of the v for the board and scoring games as previously defined by Iida and Khalid [43]. In scoring games, the success rate is defined as $v = \frac{G}{T}$, where G and T are the average successful and total scores, respectively. Meanwhile, the success rate in board games is defined as $v = \frac{B}{2D}$ where B is the average branching factor, and D is the average game length.

Table 3.2: Analogical link between physics and game (adopted from Iida and Khalid [43])

Notation	Motion context	Game context
y	displacement	solved uncertainty
t	time	progress or length
v	velocity	solving rate
M	mass	solving hardness, m
g	acceleration (gravity)	acceleration, a (Thrills)
F	Newtonian force	force in mind
$\vec{p}$	Momentum	momentum (Freedom)
U	potential energy	potential energy, E_p

The notion of energy conservation had been proposed by [50], which provided a deeper knowledge of games' engagement and addictive mechanisms is made possible by the objectivity and subjectivity perspectives [50]. The formulation of momentum in the game ($\vec{p_1}$) and potential energy in the mind (E_p) are given by (3.2) and (3.3), respectively. Then, based on the conservation of energy in mind, given by (3.4), the momentum in mind ($\vec{p_2}$) can be derived, associated with the player's engagement, given by (3.5). Applying (3.5) by assuming the formulation of $\vec{p_2} = mv_2$ where the subjective reward v_2 is given by (3.6).

$$\vec{p_1} = mv \tag{3.2}$$

$$E_p = 2mv^2 \tag{3.3}$$

$$E_p = \vec{p_1} + \vec{p_2} \tag{3.4}$$

$$\vec{p_2} = E_p - \vec{p_1} = 2m^3 - 3m^2 + m \tag{3.5}$$

$$\vec{v_2} = 2m^2 - 3m + 1 = (1 - 2m)(1 - m) \tag{3.6}$$

The relationship between objective velocity v and subjective velocity v_2 can be established. Let v_0 be the reward function over various masses for the perfect player, which corresponds to the objectivity given by (3.7). Then, $v_k(m)$ be a reward function over various m for a player with ability parameter k, which is given by (3.8). Ability parameter k stands for the strength of players in the competitive game context or error-tolerance in the social or non-competitive context. For example, there is no error tolerance for the perfect player v_0.

$$v_0 = 1 - m, \quad \text{where} \quad 0 \leq m \leq 1 \quad \text{and} \quad 0 \leq v_0 \leq 1 \tag{3.7}$$

$$v_k = (1 - km)v_0, \quad \text{where} \quad 0 \leq k \in \mathbb{R} \tag{3.8}$$

The notion of potential energy in mind was initially discussed by [43], given by 3.3. Considering the velocity derived from the reinforcement schedule $VR(N)$ with frequency N and its generalization, the objective reinforcement (E_0) refers to the potential energy in mind of the perfect player (v_0). Otherwise, the subjective reinforcement (E_k) refers to the potential energy in the minds of other players (v_k). A game would produce its potential energy in the field of play (hence, called potential energy of play) by which players would feel engagement or reinforcement.

Figure 3-3 illustrates the objective and subjective reinforcement when $k = 3$. In behavioral psychology, the term "reinforcement" refers to an enhancement of behavior. In this study, such a term was used as positive meaning, where greater reinforcement gives people a stronger interest to stay in the event under consideration. In addition, reinforcement depends on the player's ability in the game context. Reward function v_k represents a player's model or their sense of value. When assuming $k > 3$, $v_k < 0$ holds at $m = \frac{1}{3}$ where the objective reinforcement is maximized. This condition implies that most comfort point (peak of E_0) is not included in the learning context. Therefore, it is highly expected to have $k \leq 3$. Furthermore, Go ($m = 0.42$) is still not yet solved. Hence, it is expected that $2.38 < k$ holds.

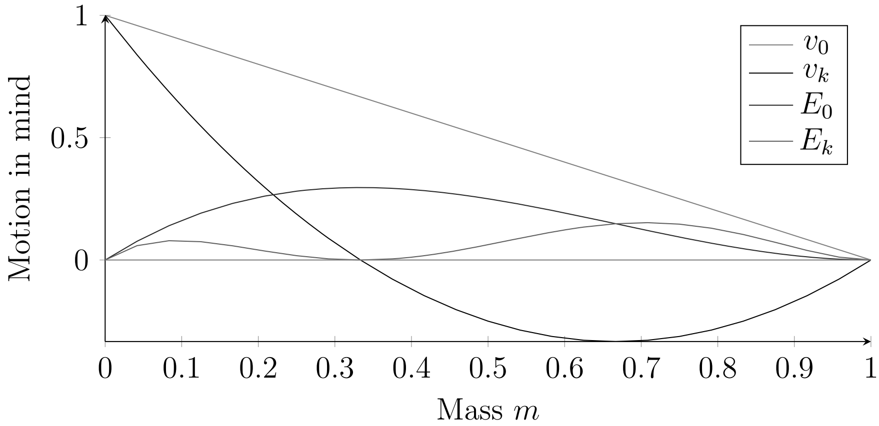


Figure 3-3: Objective and subjective reinforcement when $k = 3$

A significant difference between objective reinforcement and subjective reinforcement, for instance, $\Delta_k = E_0 - E_k$, may characterize the learning process since it represents the process of determining the truth; in some way, it is like solving a given problem (i.e., game-theoretical value in a game solving context). As such, the comfort of learning will be optimized when Δ_k is maximized at its peak $m = \frac{1}{3}$ when $k = 3$. Such a situation indicates that the theoretical success rate $\left(v_0 = \frac{2}{3}\right)$ would be a peak point to feel comfortable in the non-competitive game or learning context. On the other hand, people would feel uncomfortable (e.g., dull/anxiety in the sense of flow theory [22]) when the rate is much lower than this point.

Figure 3-4 indicates that Δ_k increases as the mass becomes larger at $0 \leq m \leq \frac{1}{3}$ and decreases as the mass becomes larger at $\frac{1}{3} \leq m \leq \frac{2}{3}$. Subjective reinforcement (E_k) was maximized at its peak $\left(m = \frac{1}{4k}\right)$ in the non-competitive game context like puzzle solving (or solving comfort), implying that puzzle solving will be highly engaged at success rate of $v_k\left(\frac{1}{4k}\right) = \frac{3}{4} - \frac{3}{16k}\,|_{k=3}= \frac{11}{16} = 0.6875$. Δ_k is maximized at its peak at $m = \frac{4}{5}$ when $k = 3$ where the game under consideration is extremely engaged due to its high competition level (called competitive comfort).

Considering these quantities to provide a metric for measuring the impacts of digital badging (i.e., game achievement) of the games on the Steam platform, it is likely that it was optimized as a meta-gaming system. In such a case, the Δ_k is maximized where the equivalent measures of success rate would be $\frac{1}{3} \leq v \leq \frac{4}{5}$. This situation implies that the

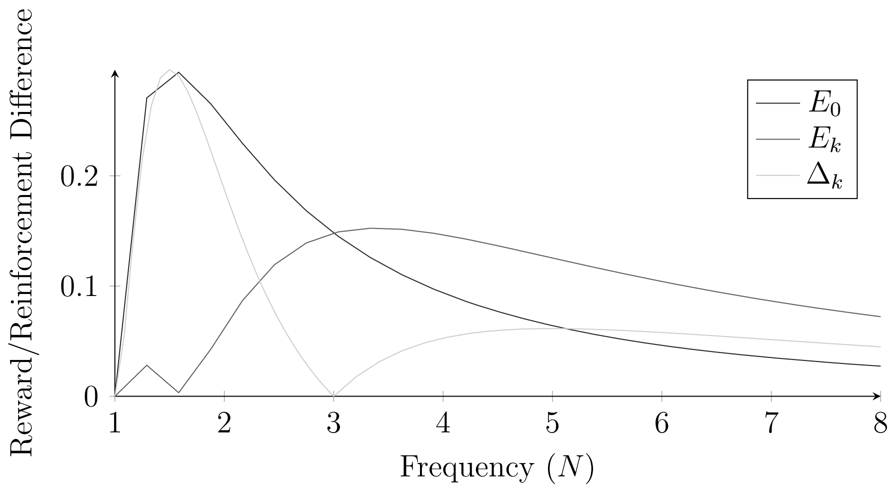


Figure 3-4: Objective and subjective reinforcement difference Δ_k when $k = 3$

potential and existing games or creators are highly engaged/reinforced since it provides a suitable condition associated with the comfort of competition and learning. As such, Δ_k is maximized at its peak $m = \frac{1}{3}$ when $k = 3$ where learning comfort is optimized. Meanwhile, Δ_k is maximized at its peak $m = \frac{4}{5}$ when $k = 3$ where the game under consideration is extremely engaged due to its high competition level.

Consequently, Johnson et al. [47] argued that understanding rewards, when and how much to be deployed would drive in-game behavior and act as an indicator for the player's progress. In the meta-gaming context, such a condition would likely be similar to where a player's cognition was focused on elements about a game's external influences instead of the game playing experience [5, 64], which was observed in the Steam platform. Previous meta-gaming solutions focus on the adopted meta-gaming approach to automate a reactive game balancing into sophisticated balance targets that were defined beyond a simple equal win requirement [40]. Meanwhile, Reis et al. [77] showed that strategy predictions and exploiting knowledge outside of a game could allow players to gain an advantage and improve the game outcome. This study is interested in discovering the underlying mechanism behind the Steam achievement indicator based on the above mentioned research questions from the perspective of game analytics and the motion in mind model. Such insights would benefit academicians and game developers in their potential future projects and creative ideation in navigating the current highly competitive game market.

3.3.2 Data Collection and Pre-processing

The data were collected from three sources: Steam Store, SteamSpy, and Steam IUserStats. A customized data collection script was developed using Python to automatically extract all of the games available on Steam Store and SteamSpy on August 10$^{\text{th}}$, 2022. Due to the large quantity of data, the process took three days to be completed. The following information is the data sources collected from the three sources:

- **Steam Store Data**: This data is consist of information about games listed on steam platform.
- **Steam Achievement Data**: This data consists of information of global achievement percentages for app on steam.
- **SteamSpy Data**: This data consists of useful metrics such as estimation for total owners of each game on steam.

Steam Store and Steam Achievement data were collected using the Steamworks application programming interface (API) provided by the Steam platform, while SteamSpy data were collected using SteamSpy API. Both APIs are publicly available to access the databases and information available on both servers. Table 3.3 shows the summary of collected data.

Table 3.3: Data summary of the three steam data sources

Data Name	Rows	Columns
Steam Store	57,155	39
Steam Achievement	32,378	4
SteamSpy	55,915	20

There are 57,155 steam games store information, and 23,774 steam global achievement data were extracted, but since the data still contains incomplete or missing information, data cleaning is necessary. A Python script was utilized in the data processing and merging, depicted in Figure 3-5. The figure showed the overview of the steps involved in pre-processing the data before in-depth analysis, which is explained as follows:

- **Cleaning / Processing**: this process includes removing duplicate rows, columns with more than 50% missing values, removing unnecessary columns (such as screenshots, movies, support_info, etc), transforming columns information (such as price, categories, achievements, etc).

- **Merge Data**: this process merged the data based on *appid* column. *appid* columns is the game's unique id on steam store listing. Such a column is the bridge between the different data sources.

We processed the data into 3 datasets which will be analyzed in order to answer our research questions. The following are information of each datasets after merging:

- **Basic information of Steam Games**: This data consisted of information combined from *Steam Store* and *SteamSpy* data which contains 49,227 rows of Steam Games. This data consisted of the Steam game's appid, name, developer, publisher, release date, price, owners min-average-max, categories, tags, positive, negative, and rating. This data includes games without steam achievements.

- **Steam Achievement Data**: This data consisted of 27,586 rows of Steam Games with its achievement data. This data consist of the same information as above data with the addition of game's achievements name and its global percentages. This data excludes games without steam achievements.

3.4 Results and Analysis

In order to determine the current state of Steam platform, we analyze the trends of Steam games and its developer on Steam platform by studying the number of games released each year and its developer. Then, in-depth analysis relative to the research question were conducted to uncover hidden trend of the Steam achievement data.

3.4.1 What kind of game releases and multi-player support on the Steam platform?

To answer SRQ1, we explored the trend of game release on Steam. Figure 3-6 shows the number of game releases, indie games ratio, and new game developers on the Steam platform from 2006 to 2022. The number of games released on Steam steadily increased

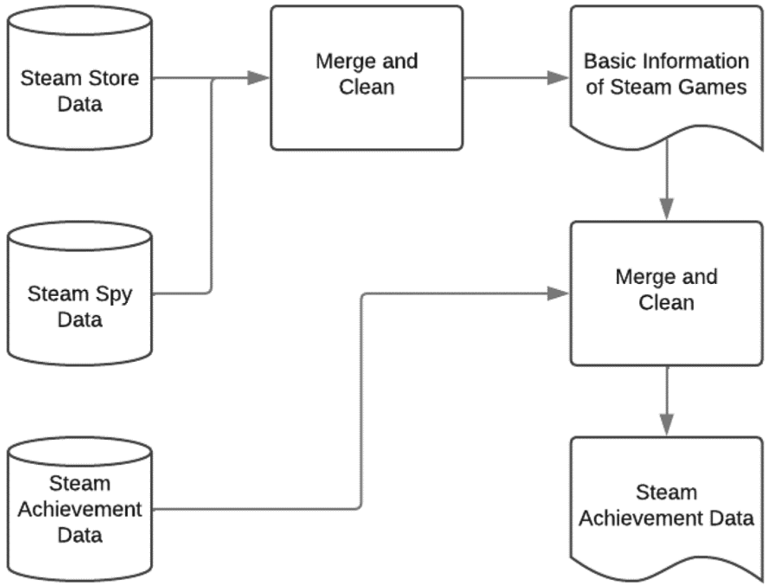


Figure 3-5: An overview of data processing and merging

every year except in 2010 and 2019 while having the highest increase in 2014 (1,498 games released, 1,079 or 350% more games than in 2013) after Steam introduced the Steam Early Access program in March 2013 [105]. The decrease that happened in 2019 possibly happened because of games' discoverability issue [93, 11], where there were changes in steam policies [11] and also the arrival of Epic Games Store [93]. Though Steam has released *Steam Labs*[8] that tackle the discoverability issues, no results were reported in achieving such outcome [26]. The number might be unreliable because changes in early access games on Steam can affect the actual release numbers on the year [2]. In 2017, Steam launched Steam Direct [27] that replaces Steam Greenlight as a new submission path designed to provide a more streamlined, transparent, and accessible way for game developers to publish their games to Steam, which results in 340% more indie games released on 2018 than the previous year. Steam Direct allows game developers to publish their games without persuading fans or steam users to vote on their game, provided they can afford the 100USD recoupable fee and meet Steam's essential criteria of legality and appropriateness. Game development/production costs can be expensive, depending on the tools, licenses, and number of people required to develop the game. Figure 3-7 and

[8]https://store.steampowered.com/labs

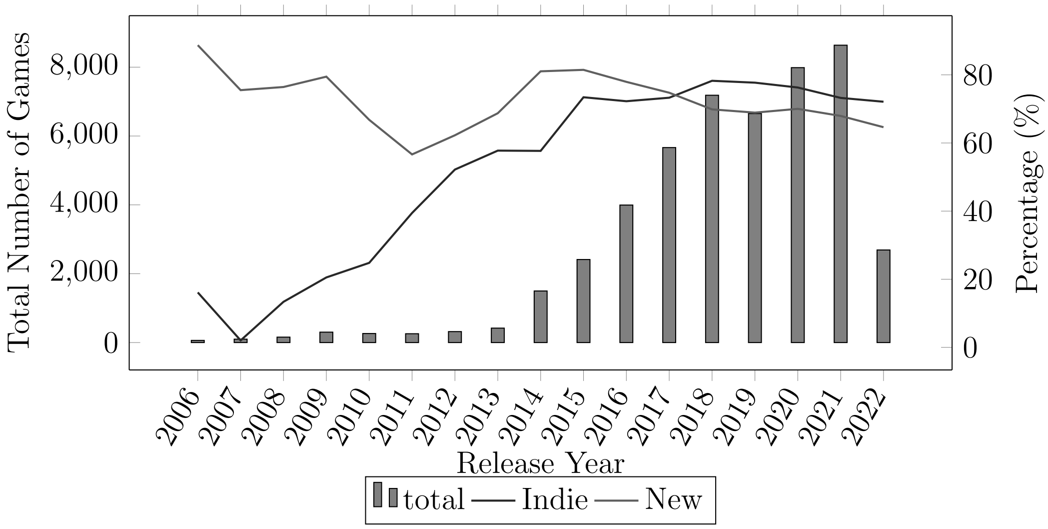


Figure 3-6: Statistics of game releases based on new developers and indie games

Figure 3-8 show the number of single-player games (single), multiplayer games (multi), and games that support both single and multiplayer gameplay (both) released on Steam from 2006 to 2022. Generally, non-indie game developers are expected to have higher financial capability compared to indie game developers, which is shown by the data where non-indie game developers published more multiplayer-supported games (30% of their overall released games) than indie game developers (20% of their overall released games) where the non-indie game developer can better afford the server costs to support their multiplayer games. We also found that 60% of all games released on Steam are single-player games developed by indie game developers.

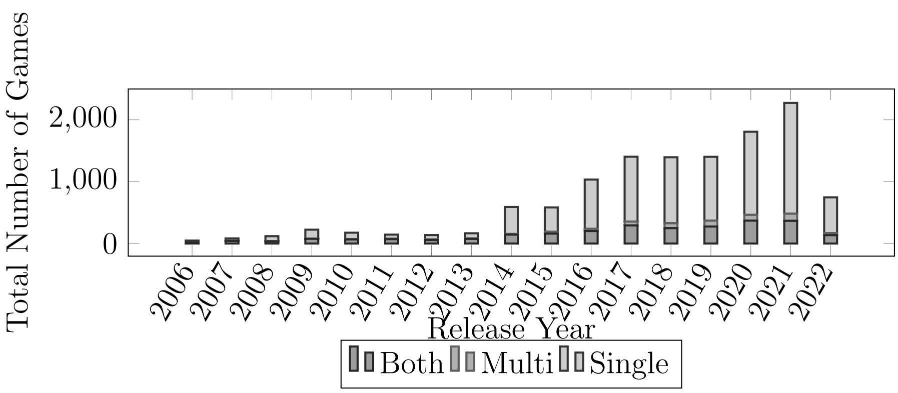


Figure 3-7: Number of game released over the year for non-indie games based on its playability (multiplayer, single player, and both)

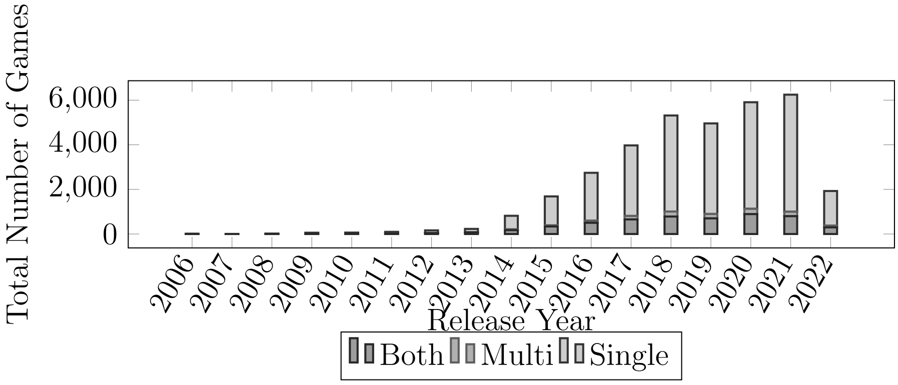
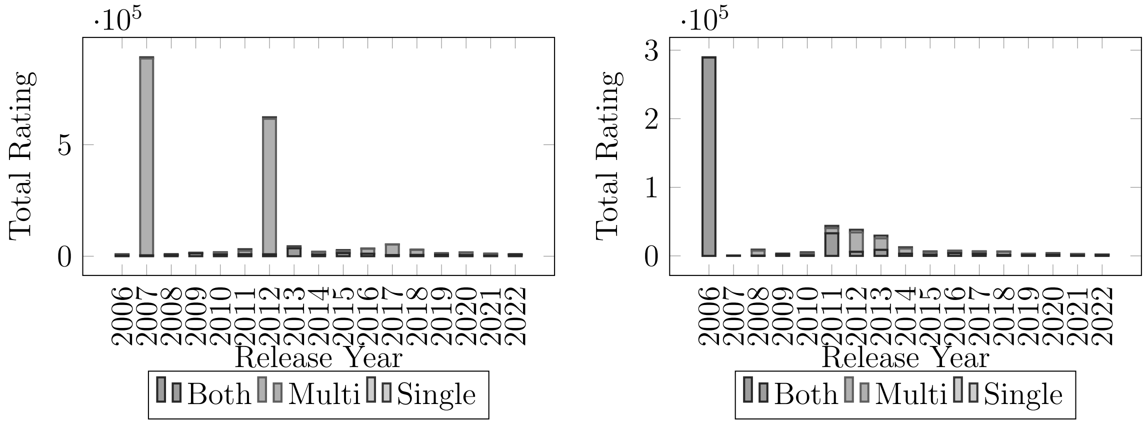


Figure 3-8: Number of game released over the year for indie games based on its playability (multiplayer, single player, and both)

3.4.2 How does rating affected by game prices on the Steam platform?

Figure 3-9, Figure 3-10, and 3-11 have shown the average player rating, Steam rating, and game price for non-indie and indie game releases from 2006 to 2022. Due to the confidentiality of the data, the exact number of owners or sales for each game could not be analyzed, so we used the *rating_total* value instead. *rating_total* shows the total number of ratings (both positive and negative ratings) a game received from its players, which we used as a metric to measure player population. In terms of *total rating* average, non-indie games received more ratings than indie games from their users, with PlayerUnknown's Battleground (rated by 2,056,746 users and 53% of positive ratings) and Terraria (rated by 991,103 users and 98% positive rating), meaning that though the steam library mostly populated by indie games, the players are still gravitated towards purchasing non-indie games than indie games. On average, the price of non-indie games is higher than that of indie games, as shown in Figure 3-11(a) and Figure 3-11(b). Interestingly, the most expensive non-indie and indie games are VR Games such as Ascent Free-Roaming VR Experience, priced at $999, and Aartform Curvy 3D 3.0 priced at $299. The production cost, the shallow player base for VR Games, and the high-spec equipment needed to develop and play a VR game might be why VR games are more expensive than non-VR games.

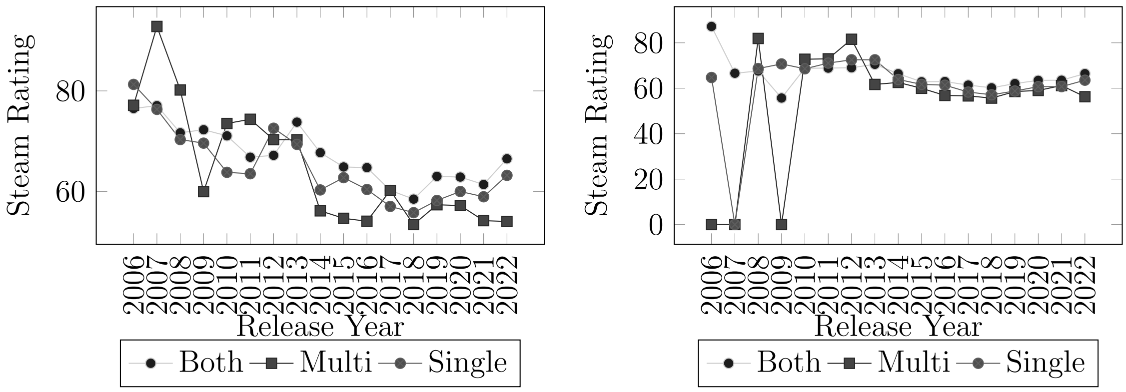
There are multiple factors when it comes to pricing a game, such as development cost,



(a) Total rating of non-indie game releases

(b) Total rating of indie game releases

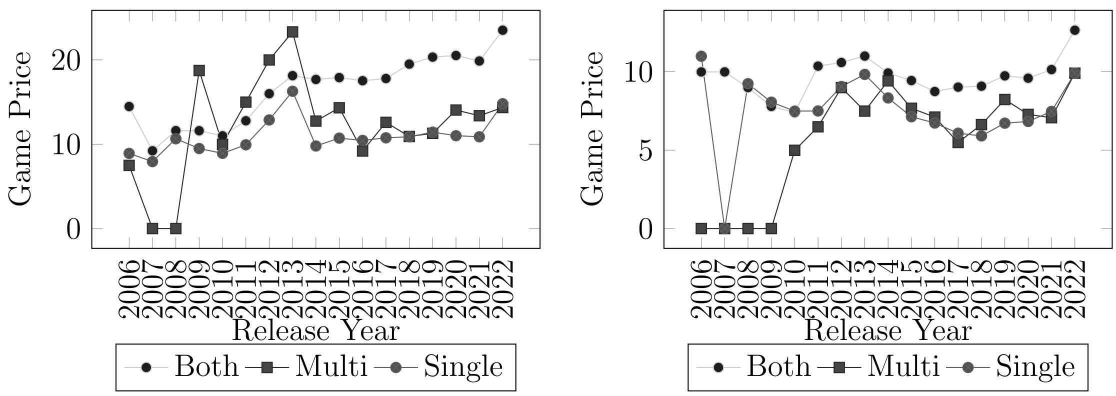
Figure 3-9: Players' rating of games released over the year for (a) non-indie games and (b) indie games based on its playability



(a) Steam rating of non-indie game releases

(b) Steam rating of indie game releases

Figure 3-10: Steam rating of games released over the year for (a) non-indie games and (b) indie games based on its playability



(a) Price of non-indie game releases

(b) Price of indie game releases

Figure 3-11: Price of games released over the year for (a) non-indie games and (b) indie games based on its playability the size of the game's contents, and other factors. For example, Final Fantasy XV [9] was priced at $60 with the content worth 60 hours of playthrough when they first launched. Players may complain if the contents available in the game are too short for the price [1]. To show the relationship between game prices and *rating total*, we did a correlation analysis as shown in Table 3.4. The value shows a positive correlation between game price and total rating for games that provide single-player and both (single-player and multiplayer) gameplay for indie and non-indie games. This condition implies that the number of ratings a game receives from its players might increase positively as the price of the game increases.

Table 3.5 provides detailed information about the total ratings received by Steam games, which are divided into two categories: indie games and non-indie games. The table shows the statistics of the total ratings based on the price range of the games. The majority of both indie and non-indie games are priced within the $1-10$ USD range, which indicates that the majority of the games available on Steam are relatively affordable. The table also shows that the highest average rating total for indie games falls within the $11-50$ USD price range, while for non-indie games it falls within the $51-100$ USD price range. This suggests that, on average, more expensive games tend to receive higher ratings. The table also shows that, on average, non-indie games receive higher rating totals compared to indie games. This could be due to the larger budget and resources available to non-indie game companies, allowing them to create higher quality games. Table 3.5 also shows the skewness and kurtosis value of the steam games data. Skewness and kurtosis are commonly used in statistics to describe the characteristics of a data where skewness is used to measure the asymmetry of a distribution, while kurtosis is used to measure its peakedness or flatness. The analysis revealed that the distribution of rating total for every price range has a positive skewness, meaning that the majority of the data points are concentrated on the right side of the distribution, and the tail of the distribution extends further to the right. This implies that there are more games with low rating totals compared to games with high rating totals. Furthermore, the kurtosis value was found to be very high, which means that the distribution of rating total is not only skewed to the right but also highly peaked. The high positive kurtosis value indicates that there are more extreme values (outliers) on the positive side of the distribution, meaning that there

[9] https://finalfantasyxv.square-enix-games.com/

are some games with very high rating totals. Moreover, the value of gini coefficient [107] on each price group is relatively high ($0.7 \sim 0.8$) except for group E, this suggests that there is a huge distribution gap between games that receive high number of rating and games that receive low number of rating regardless whether they are non-indie / AAA games or indie games. This information suggests that while the majority of games have low rating totals, there are a few games that stand out with significantly higher rating totals. The presence of these extreme values can significantly impact the central tendency and variability of the data, and may indicate that certain games are more popular and well-received by players. This also mean that the sales or performance of non-indie games will be either highly successful or a total failure, making it a high-stakes gamble for non-indie game companies in terms of the success of the games they release. Overall, the information in Table 3.5 provides valuable insights into the Steam game market and the factors that contribute to the success of games on the platform.

Table 3.4: Correlation between game prices and game rating

Playability	Indie Games	Non-Indie Games
Singleplayer	(0.33, $p < 0.05$)	(0.34, $p < 0.05$)
Multiplayer	(0.35, $p < 0.05$)	(0.36, $p < 0.05$)
Both	(0.37, $p < 0.05$)	(0.44, $p < 0.05$)

Table 3.5: statistics of price and rating total based on their price range group

Group (price range)	Quantity	Indie?	mean	skewness	kurtosis	gini	ccu	playtime forever	playtime 2weeks
A (0 or F2P)	3,377	Yes	1348	27.44	966.38	0.856	61.45	94.82	14.52
A (0 or F2P)	1,517	No	10,450	32.92	1,169.71	0.890	1423.91	456.83	115.15
B ($0 > x \geq 10$)	23,936	Yes	358.13	95.91	9,856.92	0.858	8.10	67.60	1.42
B ($0 > x \geq 10$)	7,092	No	714.26	40.81	1,786.13	0.875	24.06	96.90	3.21
C ($10 > x \geq 50$)	6,900	Yes	2,640.27	22.56	688.32	0.858	137.50	242.72	35.53
C ($10 > x \geq 50$)	3,743	No	5,868.1	18.26	496.44	0.852	319.97	514.78	49.50
D ($50 > x \geq 100$)	19	Yes	616.79	4.34	18.91	0.888	25.63	43.05	0
D ($50 > x \geq 100$)	195	No	26,501.87	4.85	27.84	0.796	2544.96	1681.28	207.41
E (> 100)	11	Yes	8.82	0.95	0.53	0.297	0	0	0
E (> 100)	11	No	28.64	2.33	5.81	0.638	8.91	0	0

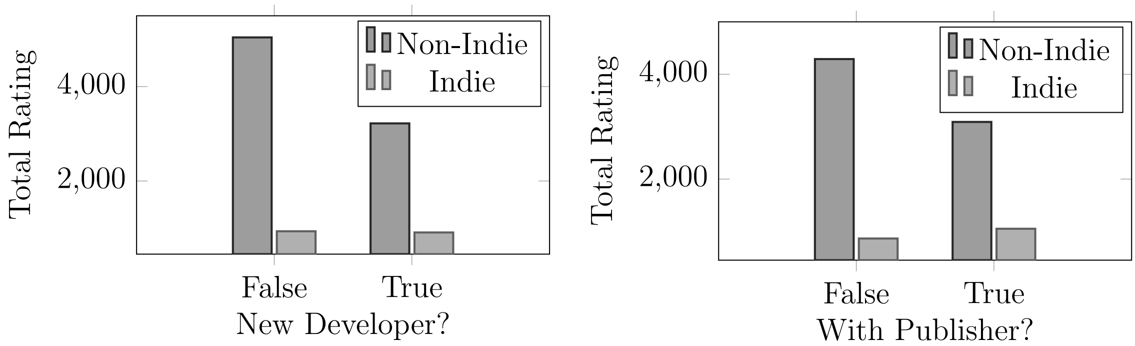
3.4.3 What kind of developers and publishers were dominant on the Steam platform?

Steam is a huge video game market, which makes it natural for it to be very competitive, both for indie and non-indie developers alike. Figure 3-12(a) shows the average *rating total* of both non-indie and indie games for new developers who just entered Steam platform market. For indie games, there is not so much difference in player ratings whether the game developer is a new or a *veteran* developer [10]. This condition indicates that Steam users do not mind the indie game developer's popularity and are more open to trying games from both new and veteran indie game developers. From the point-of-view of game developers, this indicates that both new and veteran game developers have an equal chance to be able to stand out in the steam market, as is also shown by the correlation result between the new developer indicator and total rating (Spearman's correlation coefficient, $\rho = 0.005$ and $p < 0.01$). In contrast to what have seen in indie games, for non-indie games, steam users paid more attention to the game developer's popularity and were pretty hesitant to try out the games from new developers, although statistically, it has a weak negative correlation (Spearman's correlation coefficient, $\rho = -0.18$ and $p < 0.01$).

In a competitive environment such as steam market, working with a game publisher could improve the reach and get your games noticed by steam users. Figure 3-12(b) shows the average *total rating* of both non-indie and indie games from game developers who are working and are not working together with a publisher. Figure 3-12(b) shows that there is no significant difference in terms of the average rating total received by self-published games and publisher-published games and statistically has a weak correlation (Spearman's correlation coefficient, $\rho = 0.15, p < 0.01$). Although working with a publisher might not boost the *rating total* they received from the players, it can give many benefits for indie game developers such as instant validation and hype, effective app store and media distribution, getting professional advice from the experts onboard, efficient planning and no upfront financial costs (except royalties) [51]. While it can be beneficial for indie game developers to work with publishers, especially for its financial and expert input, for non-indie game developers, publishing their games on their own will benefit them more

[10] veteran game developer is the game developer who already published more than one game in steam platform

since it gives them more flexibility and freedom on how they will sell their games.



(a) Average total rating of games from new or veteran developer based on non-indie and indie game developer)

(b) Average total rating total of games from games that's published via publisher based on non-indie and indie game developer)

Figure 3-12: Average total rating total of games based on (a) development experience and (b) publishing options for indie and non-indie game developers

3.4.4 Does Steam achievement affect game rating and type of game playbility?

To determine the influence of game achievements, we did a spearman's correlation analysis between the number of achievements and game rating for each game playability type for indie and non-indie games as shown in Table 3.6. Table 3.6 shows a positive correlation between the number of achievements and the number of ratings a game receives. Based on the data, both indie and non-indie games have a positive correlation between the number of achievements and game rating, and it varies depending on the game playbility type, number of achievements has more impact on non-indie games than indie games.

Table 3.6: Correlation between number of achievements and game rating

Playability type	**Indie Games**	**Non-Indie Games**
Singleplayer	(0.241, $p < 0.05$)	(0.410, $p < 0.05$)
Multiplayer	(0.394, $p < 0.05$)	(0.473, $p < 0.05$)
Both	(0.335, $p < 0.05$)	(0.560, $p < 0.05$)

Moreover, k-mean clustering was adopted from Scikit-learn [71] to provide an in-depth analysis of the steam games achievements data, where the parameter of the k-mean cluster was $k \in [2, 7]$ with the number of achievements, achievement percentage, game playability type, and user rating as the input. Using silhouette coefficient assessment and elbow method, it was found that a five-cluster is the most efficient for the analyzed data, as described in Table 3.7.

Table 3.7: Clustering results of the game data analysis

k	**Size**	**Labels**	**Characteristics**
1	15,200	Single Player Games	Single player games only
2	5,873	Multiplayer Games	Multiplayer supported games only (including single player games with multiplayer support)
3	6,407	Games with Great Achievements	High average global achievement percentage, mix of single and multi player games
4	105	Achievement Spam Games	Highest average achievement quantity per game, lowest average price, lowest average rating total
5	1	Highly Rated Games	Counter-Strike: Global Offensive, most rated game on steam

Thus far, five clusters has been identified from the k-mean clustering result, which clusters the games into five distinct groups. Cluster 1 consists of games that only support single-player gameplay. In contrast, Cluster 2 consists of multiplayer-supported games (multiplayer-only or single-player games with multiplayer gameplay options). In addition, we found that some multiplayer games were initially released as pay-to-play (P2P) games and transitioned to free-to-play (F2P) games, such as Counter-Strike: Global Offensive, Team Fortress 2, and PUBG: Battle Grounds. This business model transition allows the games to attract new players and return players that previously had stopped playing the game [79]. Moreover, Free-to-play (F2P) games usually adopt microtransactions within their business model to allow revenue from Loot Boxes, Character Skin, or Premium Subscription that they provide in their games.

Cluster 3 consists of games with a high global achievement percentage value. Achievements are typically adopted in games to encourage the player to explore the game world or to play the game in different play styles and extend a game's lifetime of sales [25]. However, based on Table 3.8, Cluster 3 suggests that the number of achievements available in

Table 3.8: Correlation between number and percentage of achievements, and rating of Cluster 3

Playability type	***n_acv* to percentage**	***n_acv* to rating**
Singleplayer	(-0.098, $p < 0.05$)	(0.235, $p < 0.05$)
Multiplayer	(-0.470, $p < 0.05$)	(0.199, $p < 0.05$)
Both	(-0.639, $p < 0.05$)	(0.120, $p > 0.05$)

multiplayer-supported games has a high negative correlation to their global achievement percentage value, while it is not for single-player games. This situation implies that multiplayer game players are prone to not completing the in-game achievements when the quantity is too many. This condition happens because generally, the goal in multiplayer games requires the player to compete against other players to achieve the goal of the game, which leaves them limited to no chance or time to explore additional contents in the game as well as the poor design of the progression type [25] achievements that take too much effort to achieve (i.e., requires the players to play long hours of *grinding* to achieve)

Interestingly, we found a group of *unique* games in Cluster 4 where they provide a high number of achievements in their games. These *unique* games are known as *achievement spam games*, where players can easily get plenty of in-game achievements within a relatively short time. Moreover, those achievements can be displayed on the user's Steam profile on the Steam platform. Through *achievement spam games* can be seen as a unique business model that serves a *niche* market because they only focus on the attractiveness of the sheer quantity of achievements, while the game content itself is generally not the main attractive part [87]. However, these *achievement spam games* are considered fake games that exploit the Steam system. In response to this, *Valve* introduces a *Confidence Metrics* which puts a limit of 100 achievements on games until Steam recognizes them as *real games*.

In contrast to what have observed in other clusters, Cluster 5 consists of only one game, Counter-Strike: Global Offensive (CS:GO) [99], which has received more than 6,000,000 ratings from its players. CS:GO was initially released as a pay-to-play (P2P) game in 2012, and it transitioned into a Free-to-Play (F2P) game in 2018 [79], providing content for single and multiplayer gameplay gives flexibility for the player on how they play the game as they can play the game by themselves, with their friends, or with other players online.

3.5 Discussion

Games' ratings were analyzed through motion-in-mind v value, where it was measured using $v = \frac{G}{T}$ model, where G is the number of positive ratings and T is the total rating a game received. Table 3.9 shows the correlation values between games' rating (v) to their steam market stats based on the developer type (indie and non-indie developers). Table 3.9 shows that there is a positive correlation between a game's v value and its global achievement percentage, which means that a game with a higher v value is more likely to have more achievements unlocked by its players. This suggests that players who rate a game positively are more engaged with the game, and are likely to explore its content and try to unlock achievements. While it has a positive correlation to games' global achievement percentage, v value has a negative correlation to the number of days since the game is released, which implies that the longer a game is released, the more likely it will have a lower positive response from the player, leading to losing its popularity. In a sense, the v captures a positive response (rewards) from their overall player base (total attempts). Therefore, when a game receives a high v value ($v > 0.5$), it puts the game in an advantageous position where it can lead to higher popularity and the more player attracted to play the game, including exploring its additional contents (games' achievement). Meanwhile, when the game is of low v value ($v < 0.5$), it puts the game in a disadvantageous position, making it lose its current players and future players that might be attracted to buying/playing the game. Therefore, the v value of a game might fluctuates as the time goes by and developers need to keep their games updated and engaging to maintain a high v value and keep their player base engaged. This can involve fixing bugs and adding new features, content, and incentives for players to keep playing and rating the game positively. In conclusion, the v value is a valuable metric for game developers to measure the success of their game and understand how to improve it to attract and retain more players.

Table 3.4 showed a positive correlation between the game price and the total rating received by a game. This condition implies that the higher the game price (higher price tag), the more rating the games tend to receive from the players (more people will buy

Table 3.9: v correlation to total rating, achievement percentage, number of achievements and number of days since the game is released

Developer	***v***			
Type	**total rating**	**percentage**	**n_acv**	**n_days**
Indie	(0.01, $p > 0.05$)	(0.10, $p < 0.05$)	(0.06, $p < 0.05$)	(-0.22, $p < 0.05$)
Non-Indie	(0.02, $p > 0.05$)	(0.23, $p < 0.05$)	(0.01, $p > 0.05$)	(-0.23, $p < 0.05$)

the game). However, Figure 3-13 shows that there are peak points of game prices for each game playbility type from different developer types. For instance, once a game is listed with a price tag over its peak point, it might not sell as much as other games with a lower price tag because the players feel it is too expensive or overpriced. Another insight that can gain from Figure 3-13 is that the game developers can strategize how they will price their games to sell them or attract players to buy them efficiently. For example, they can sell their single-player game on a \$20 as their regular price tag and \$15 as their *discounted* price tag when they first released their game or during *steam sales* period to attract players. Another strategy that multiplayer games can use is that by transitioning to a Free-to-Play (F2P) business model to attract players to play their game, as we saw in Counter-Strike: Global Offensive (CS:GO) in Cluster 5 that significantly increased their player population [79] which can lead to more revenue when the player spends on their microtransactions scheme.

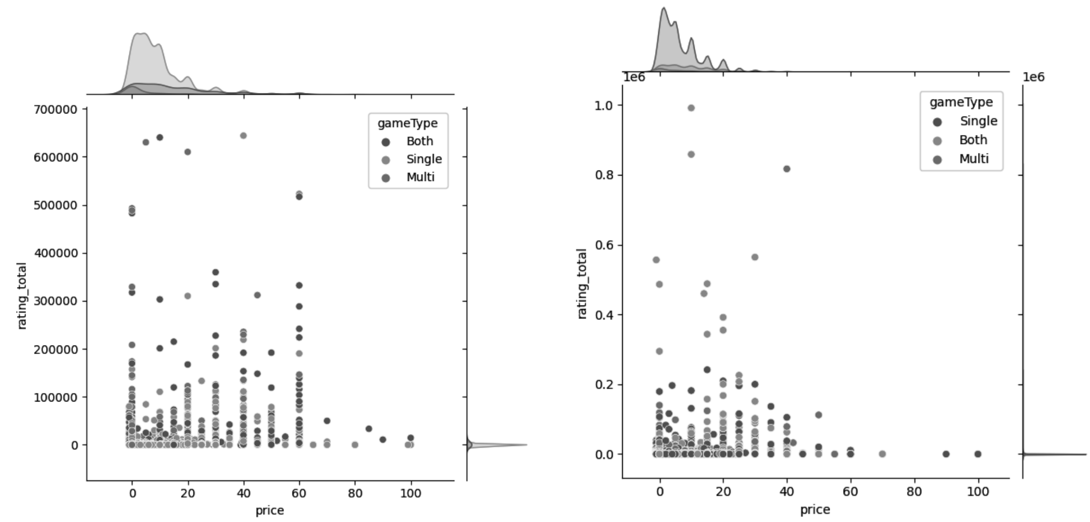
Moreover, a regression analysis was conducted through *Orange Data Mining*[11] software using their built-in *Random Forest* algorithm. First, the analysis was conducted between the independent variable (*rating total*) and the dependent variables (*game price, game playability type, number of achievements, steam rating, rating ratio, and an indicator whether the game has achievements, is released through the publisher, is an indie game, is a free-to-play game and is a game from new developer*) of the collected Steam data. Next, outliers games were removed from the data using *Orange*'s built-in outliers detection feature with *Covariance Estimator* method and removed 4,655 outliers games from the data. Then, the train-to-test ratio data of 80:20 of the sample data was considered. Based on the result in Table 3.10, R^2 value indicated that the independent models explain 75% of the variance of the dependent variable. In addition, both mean absolute error (MAE) and

[11]https://orangedatamining.com/

root mean squared error (RMSE) values were relatively low, within 261.983 and 3582.095, respectively, which is lower than the rating total's average value of 814.45 and a standard deviation of 10,106.77. This suggests that the model could be used to accurately predict a game's rating total based on various independent variables, with relatively small errors.

Table 3.10: regression analysis result for game's rating total

Model	**MSE**	**RMSE**	**MAE**	**R^2**
Random Forest	12831403.471	3582.095	261.983	0.753



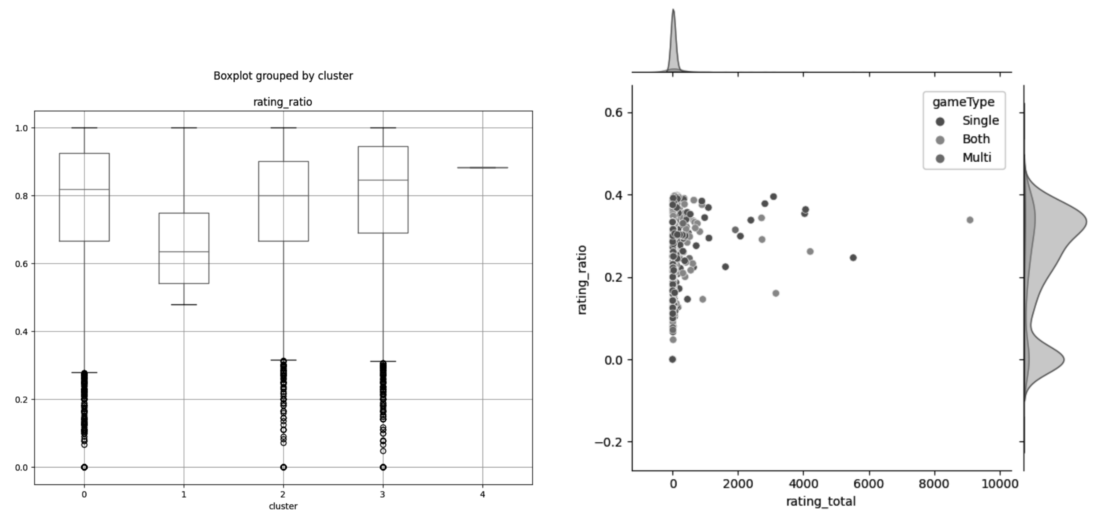
(a) Scatter plot of price against rating of all games (non-indie games)

(b) Scatter plot of price against rating of all games (indie games)

Figure 3-13: Price of game released over the year for (a) non-indie games and (b) indie games based on its playability

In terms of game playability types, the difference between single-player and multiplayer games can be observed where the multiplayer games tend to have lower achievement percentages. One of the possible explanations here is that in single-player games, players have more space or opportunities to explore the game's additional objectives to obtain in-game achievements beyond the main objectives. On the other hand, the player of multiplayer games was occupied with competing or cooperating with other players to reach the game's primary objective. Moreover, multiplayer games were preferred by non-indie game developers since they attracted a more extensive player base and, consequently, less number of releases throughout the years.

Moreover, it can be implied that the Steam platform indirectly played a role in gamifying the game releases and game ratings by providing a meta-game solution to well-known game studios and indie games. Figure 3-14 showed the boxplot and scatter plot of the positive rating ratio against the total rating of the collected Steam games. By considering $v = \frac{G}{T}$ model, where G is the number of positive ratings and T is the total rating a game received, Figure 3-14(a) showed that the majority of games regardless on which cluster they belong to, were $v \in [\frac{2}{3}, \frac{4}{5}]$ (except for some outliers). These findings substantiated that the Steam platform is still the biggest and most popular digital game distribution platform compared to other similar platforms for players and developers (such as Epic Games and itch.io). In addition, the outlier games ($v < 0.4$) that were shown in the Figure 3-14(b) implied that these outlier games are less popular or low rated (low v which also implies low *rating_ratio* values).



(a) Boxplot of positive rating ratio grouped by game clusters

(b) Scatter plot of positive rating ratio against total rating based on game playability type

Figure 3-14: The positive rating ratio given by (a) clusters and (b) against total rating

In diverse situations, the clustering results revealed that a unique solution existed that allowed for different manipulation of the Steam platform to maintain the sustainability of the game studio. For instance, the *achievement spam game* takes advantage of the achievement system to attract players to the game regardless of the content [87]. It also acts as an additional objective a player can achieve outside the intended content of the games, mainly to retain interests and continuity of the platform. Nevertheless, proper moderation is crucial to balancing and maintaining players' interests on the platform.

From the developer's perspective, rating and achievement of the Steam platform played other roles as indicators for developers to perform decision-making and risk assessment. The Steam platform helps indie developers to build reputations and user acceptance of their game release brand. On several occasions, there had been indications that the players care less about the type of developers (indie or non-indie) but care more if the developers work with a publisher. Moreover, it can be implied that fast-paced developers under a small studio or publisher would make more revenue and better received by Steam platform users by releasing single-player games. In contrast, a large development company backed by well-known publishers would be better off focusing on multiplayer game releases to take advantage of the platform achievement features while maintaining continuity by incorporating downloadable contents (or DLCs) [102].

From a business point-of-view, the Steam platform, besides being the instrument of monopoly, capitalism, commodification, and ecosystem of digital products (for instance, games) [113], providing not only a channel that connects developers and players via a gamified platform but also maintains dynamic interactions between developers and players via constant engagement and structural processes. As such, the purpose of a platform becomes a meta-game for developers and publishers to bridge the needed experience of the players. In addition, the Steam platform also provides a gaming experience beyond the game itself, a notion known as 'extraludic' [5, 64] while laying out the opportunity for learning and value-added social interactions by acting as both mechanical and social metagaming solution [60].

Finally, related to the analysis of Steam achievement, it can be implied that its proper implementation relative to the in-game and out-game contents regulates and drives a new form of "game economics" (cf. [94]); thus, the developers and publishers may take advantage over it as a form of gameplay activities or game experience to generate revenue. Although there had been some concerns regarding addiction and betting on the Steam platform [113, 94], it does provide a unique opportunity to introduce a novel business model that emphasizes player experiences rather than exploiting the Games as a Service (GaaS) model just for revenue generation [102]. Also, price and ratings ultimately make or break certain games, as found in the Steam platform analysis, highlighting the importance of cost transparency and cross-linkage to maintain existing players and attract new ones (cf. [16, 102]).

In light of this research findings, the need for a harmonic balance between the play experience (of players) and the revenue generation (of developers and publishers) could be achieved via an experience-driven business model implemented in the context of the digital games distribution platform (Figure 3-15). By having the appropriation of digital badging (such as Steam game achievement and rating) with cross-linkage services (such as chat rooms, social media interfaces, and reviews) to dynamically compute aggregated scores (such as the motion in mind model) to rank 'expected experience' players can perceive from the game listing. However, such a business model requires some investment (cost and time) from the developers and publishers. Instead of directly releasing the game, they have to undergo a moderation process (by the platform provider) to incorporate appropriate digital badges and cross-linkage services. Finally, the aggregated metrics will be updated periodically when new data from the digital badges and cross-linkage services are acquired. The envisioned business model will enable game developers to improve the visibility of their games and direct them to their targeted users, enable *steam* to recommend more personalized game recommendations to *steam* users depending on the type of games that they are interested in (i.e., based on the game's ratings, reviews, achievements, etc.) and finally, this may improve user satisfaction itself where they can discover games that they will like.

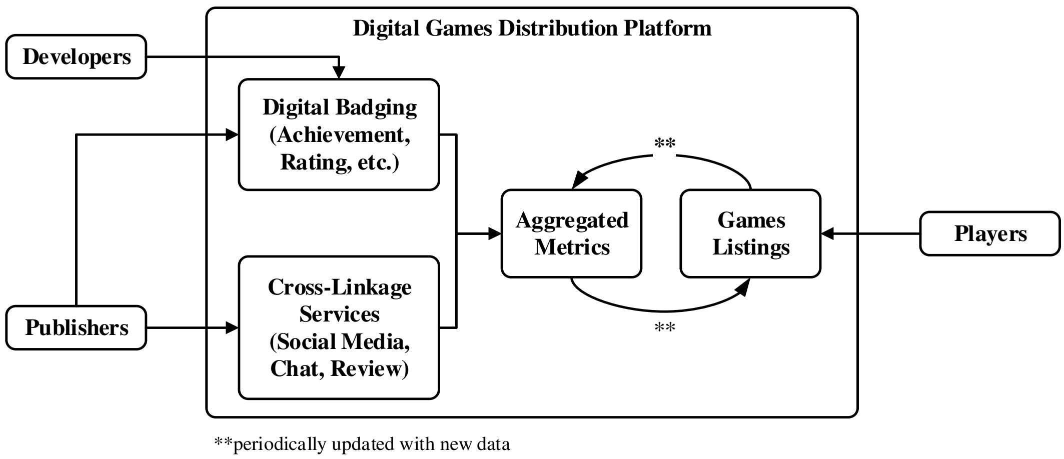


Figure 3-15: An envisioned "experience-driven" business model

3.6 Limitation

The limitations to this research include focusing on games that have in-game achievements released between 2006 and 2022 on Valve Corporation's Steam platform. This situation excludes games without in-game achievements. The primary focus was on the correlation of the data collected over causality. Also, the data collected were mainly focused on game price, categories, achievements, rating, releases, developer, publishers, and other essential auxiliary labels (such as game tags, app id, name, date, etc.). Due to the nature of game data confidentiality, the exact number of the owner of each game was not accessible. In this research, *total rating* was used to assume the number of players who own the game. Since only users who own the games in their library can review the game in the [86] platform, the actual game owners number can be higher because not every player writes reviews of the game they purchased (or played).

In addition, the measure of v, based on the number of positive rating and total rating of the Steam game received, indicated the competitive comfort that the Steam platform presented to the game players and game developers. However, such a method was based on the assumption that the level of user's ability $k = 3$ was based on the objective and subjective analysis of popular board games. Therefore, game designers and developers should have taken these findings with a grain of salt.

This research analyzed the global achievement percentage of the achievements that exist in the games. Therefore we do not differentiate the difficulty of each achievement in the games. However, there is always the possibility that completion rates could differ for games that force players to choose higher difficulty modes for obtaining achievements. Without an achievement to signify completion at the easier difficulty, there is no way to count the ratio of players who finish the game on the more accessible mode.

In particular, all games on Steam are digital downloads, so that completion rates could differ from games purchased on a physical medium. Also, the research focused on both games that require money to purchase and "free-to-play" games. Therefore, content usage will still be an issue for those games. However, the business model involved is different enough to warrant a separate investigation, particularly concerning mobile games, such as ones released on Apple's iStore or Google's Play platforms.

Since the Steam platform provides digitally downloaded games, the nature of the games and its player may be different compared to other existing or competing platforms, such as Sony's PlayStation, Microsoft's Xbox, Nintendo's Wii, and Switch hardware, and so on. Finally, the present research is empirical by nature, and causation between factors cannot precisely be determined where some other independent factors may be the primary focus. Therefore, this research focuses on the insights obtained from examining the three data repositories (Steam store, SteamSpy, and Steam IUserStats) via the publicly available Steamworks and SteamSpy application programming interfaces (APIs).

3.7 Chapter Summary

This chapter demonstrated the importance of meta-gaming of a platform based on a Steam platform, where 18,658 Steam-listed games were acquired and analyzed from Steam Store, Steam Spy, and Steam achievement databases. The study has provided a detailed analysis surrounding four key research questions relative to the game playability types, game achievement, game releases, game rating, game pricing, and their developers and publishers. First, the study found that achievement or any form of digital badging [64] can increase player's engagement in the form of 'extraludic' or play beyond the game itself while potentially inducing its own "game economics" [94]. In this direction, the managerial implications include the need for an experience-driven business model that could provide the harmonic balance between the play experience demanded by the players on one side and the revenue generation of the developers and publishers on the other side without the need for an additional intermediary.

Chapter 4

Puzzle Generation and Analysis in FlowFree

This chapter is an updated and abridged version of the following publications:

- Rizani, M. N.,Liu, C., Abuluaih, S., Khalid, M. N. A., and Iida, H., Motion-in-Mind Apporach Level Generation in FlowFree, the Artificial Intelligence and Entertainment Science Workshop (AIES 2021) Online, Japan.

- Muhammad Nazhif Rizani, Xiaohan Kang, Mohd Nor Akmal Khalid, Hiroyuki Iida and Saajid Abuluaih. Puzzle Generation and Analysis in FlowFree. The 10th ASEAN Workshop on Information Science and Technology (AWIST 2022), pp. 173-182, 2022.

4.1 Chapter Introduction

The fourth chapter in this dissertation covers about generating and solving puzzles in *FlowFree* game as well as analysis of *FlowFree* puzzles with Motion-in-Mind measures. Additionally, an experiment is conducted to collect human gameplay data and player perceived difficulty data through a clone of *FlowFree* game we developed. The mechanism of *FlowFree* puzzle generation will be presented. At the end of this chapter, the result of analysis on *FlowFree* puzzles and its correlation to human gameplay and players' perceived difficulty data will be discussed and presented as a conclusion.

The game market is massive, it has generated US$196.8 billion in world wide revenue and consisting of over 3.1 million people consuming the game as reported in July 2022 [1]. One of video game publishing platforms, *Steam*[2], with over 46 thousand titles in their library, is predicted to have over $8 billion revenue in 2022 [3] making them one of the most successful video game publishing platform that allows multitude of game developer to publish their games as well as providing their players with abundant choice of games from *Steam*'s library to enjoy. While the video game market grows larger, it becomes harder for a game developer to make their game stand out in such a competitive crowd as it requires the developer to not only produce such high-quality gameplay and mechanics in their game to keep engaging for their players but also to provide contents to keep their players attracted to play their games.

Implementation of procedural content generation (PCG) is quite common in video games to reduce the cost of hand-crafting contents in game such as levels and dungeons, its uses in puzzle generation has been limited due its strict solvability constraint [24]. Although PCG allows automatic content generation in games, the content generation is often random—adjusting the content according to user needs and preferences are essential steps toward effective and meaningful PCG [110]. Dealing with play experiences and player understanding is the foundation of motion in mind model [43], where the analogy of motion from classical mechanics is mapped to the underlying information of gaming in our mind. This condition allowed for measuring subjective events from objective ones [50] and determine the level of comfort based on reward intensity [109]. This study explores the following research question: (1) How does motion-in-mind interpret puzzles' difficulty in FlowFree? (2) How does the player interact or solve the *FlowFree* puzzles? (3) How do player perceived difficulty correlate with *FlowFree* puzzles structure and its motion-in-mind values?

4.2 Related Works

Several research had been conducted in the field of solving and generating puzzle. For example, [24] conducted a detailed survey on existing work in procedural content gener-

[1]https://newzoo.com/key-numbers
[2]https://store.steampowered.com/
[3]https://www.statista.com/topics/4282/steam/

ation in puzzles, the author discovered seven salient characteristics linked to the method and show commonalities and differences between techniques and listed promising areas for future research. In [83], the author applied a modified Evolutionary Strategy (ES) algorithm to discrete optimization problem of solving Kakuro puzzles. The proposed method modifies the mutation rate on each generation based on the best fitness of the population which allow the algorithm to outperform all implementations of the ES algorithm without dynamic mutation and a generic genetic algorithm. [19] presented Baba is Y'all, a prototype system for collaborative mixed-initiative level design where user interact with a procedural content generation system to create levels, automatically play test levels, rating level, helping level design through an evolutionary algorithm as well as suggesting levels to design. In their survey about procedural puzzle generation, [24] identified four open challenges associated with puzzle generation, which are: (1) difficulty progression for educational puzzle, (2) developing a general technique to come up with novel puzzle regardless its types, (3) assessment quality of the puzzle to measure difficulty, variety, freshness and aesthetic, and (4) ability for the procedural generation technique to create designs that are as aesthetically pleasing as human designs.

4.3 Methodology

In this section, we will discuss about developing generator for *FlowFree* puzzles as well as the *FlowFree* clone we developed to collect player perceived difficulty and gameplay data.

4.3.1 Motion-in-Mind Model

Considering the zero-sum assumption in game playing[4], the essence of uncertainty can be determined [50]. In the schedule of reinforcement of the operant condition originally designed by [84], a variable-ratio schedule is a reinforcement schedule where the response is reinforced after an unpredictable number of responses, creating a steady, high rate of responding [45]. Meanwhile, mind sports games (such as chess and Go) are essentially stochastic games when applying the move selection model [43]. From a reward-driven standpoint, this condition is a typical example of a reward system based on a variable ratio

[4]Zero-sum assumption can be defined as the gain or loss utility of one player that precisely equalized by the losses (or gains) utility of its opponent [65]

schedule found in stochastic games (such as gambling and lottery games). Meanwhile, mind sports games (such as chess and Go) are essentially stochastic games when applying the move selection model [43]. This condition implies that a game is characterized by a reward of a variable-ratio reinforcement schedule. In essence, the game is characterized by the reward function, a variable rate (denoted as $VR(N)$) of the reinforcement schedule. Then, velocity v (win rate) and mass m (win hardness) of the motion in the mind model are given by (4.1).

$$v = \frac{1}{N} \quad \text{and} \quad m = 1 - v, \quad \text{where} \quad 1 \leq N \in \mathbb{R} \tag{4.1}$$

Table 4.1 describes the analogy of the motion in mind model from the physics and games context. Note that there is a distinctive computation of the v for the board and scoring games as previously defined by [43]. In scoring games, the success rate is defined as $v = \frac{G}{T}$, where G and T are the average successful and total scores, respectively. Meanwhile, the success rate in board games is defined as $v = \frac{B}{2D}$ where B is the average branching factor, and D is the average game length.

Table 4.1: Analogical link between physics and game (adopted from [43])

Notation	Motion context	Game context
y	displacement	solved uncertainty
t	time	progress or length
v	velocity	solving rate / success rate
M	mass	solving hardness, m
g	acceleration (gravity)	acceleration, a (Thrills)

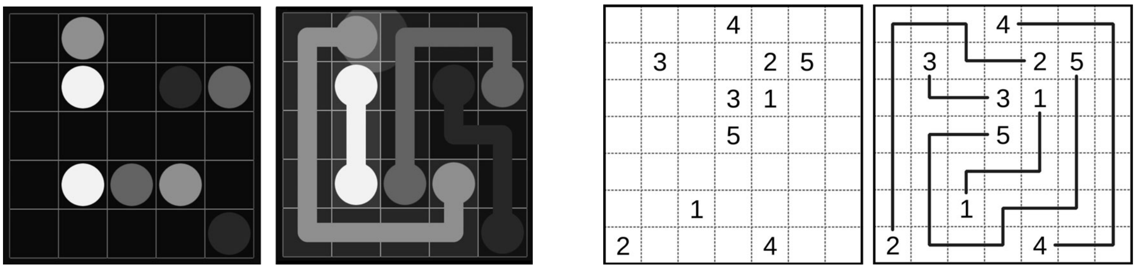
The notion of energy conservation had been proposed by [50], which provided a deeper knowledge of games' engagement and addictive mechanisms is made possible by the objectivity and subjectivity perspectives [50].

4.3.2 FlowFree

Flow Free [5] is a puzzle game released by Big Duck Games in June 2012 (Figure 4-1(a)). The game presents *Numberlink* puzzles (Figure 4-1(b)), each level has a grid of squares with colored dots occupying some of the squares, with the objective to connect dots of

[5] https://en.wikipedia.org/wiki/Flow_Free

the same color by drawing lines between them such that the entire grid is occupied by the lines in which the lines may not intersect, Numberlink is known to be NP-complete [52].



(a) FlowFree puzzle, initial state (left) and solved state (right)

(b) Numberlink puzzle, initial state (left) and solved state (right)

Figure 4-1: Puzzle initial and solved states of (a) FlowFree and (b) Numberlink

4.3.3 Generating and Solving FlowFree Puzzles

Puzzle is defined as problems to which player can find a solution based on previous knowledge and/or by exploring the solution space [20], therefore, each puzzle that will be played by the players needs to be solvable. In order to generate a solvable puzzles, solving the generated puzzle to check its solvability is required.

Due to having a strict rule in solving *FlowFree*'s puzzle, this problem falls under the category of Constraint Satisfaction Problem (CSP). Constraint Satisfaction Problem (CSP) is a set of finite constraint and a finite set of possible outcomes based on the constraint [80]. To simply put, a CSP is a path for finding solutions that are satisfying the given constraint to produce a desired output [81]. Below are the constraints we defined for *FlowFree* puzzles:

- Every cell in FlowFree puzzle is assigned to a single color
- The color of every endpoint cell is known and specified
- Every endpoint cell has exactly one neighbor which matches its color
- Every endpoint cell cannot be directly connected with another endpoint cell
- The flow through every non-endpoint cell matches exactly one of the six direction types:

`─`, `│`, `┘`, `└`, `┐` or `┌`

- The neighbors of a cell specified by its direction type must match its color
- The neighbors of a cell not specified by its direction type must not match its color

Our algorithm for *FlowFree* puzzle generation is shown as Algorithm 1. The generator will first randomly place initial positions of the puzzle boards. These initial position consists of 2 endpoint cells of each color exists in the puzzle according to its setting (for example, a puzzle is set to be generated as 8x8 size and 4 colors, the initial position will consist of 4 pairs of endpoint cells of each colors). Once the puzzle initials are declared, the solving process begins where each active cells transfer its active status to assign its color to one of their empty neighbor cells (top, bottom, left and right cells) and checks if the current puzzle state satisfy the constraints we defined in previous section. If the constraints is satisfied, the solving process continues and if the constraints is not satisfied, the current cell will assign to another empty neighbor cells or backtrack to search through new nodes. The puzzle is *solvable* when the *Solve* function return a *true* value and *unsolvable* when it returns a *false* value. To put it simply, the puzzle generator creates the *flowfree* puzzle initial position randomly and it tries to solve it using a simple depth first search (DFS) algorithm within a specified constraints.

4.3.4 Collecting Human Data

In addition to analyzing *FlowFree* puzzle using motion-in-mind, we also experimented with collecting human gameplay data and player-perceived difficulty data to provide additional analysis for this study. We developed a clone of *FlowFree* using Unity version 2020.3.26f1, and *FlowFree* client is developed for Windows and Android platform. Figure 4-2 shows the interface of *FlowFree* clone we developed, it consists of 4 clickable buttons, *free mode / tutorial* button for the player to practice of learn the basics of "how to play" flowfree, *Session 1* and *Session 2* are the button to start the experiment session for human gameplay data collection and *Exit Application* can be used to exit the application. In *FlowFree* clone, we conducted 2 experiment sessions consisting of 24 levels each, starting from puzzles with 5x5 to 10x10 board size, the board size will increase every four

Algorithm 1 *FlowFree* puzzle generator and solver

```
Put Initial position on board
Solve(board)
function SOLVE(board)
    if Board.isSolved() then
        return true                                          ▷ Board is Solved
    else
        for each cell in Board.activeCells do
            cell.setInactive()
            for each neighbor in cell.EmptyNeighbors do
                neighbor.setActive()
                neighbor.color = cell.color
                if Board.checkConstraints() then
                    if Solve(Board) then
                        return true
                Reset(neighbor)
            Backtrack(cell)
    Return False                                             ▷ Board is Unsolvable
```

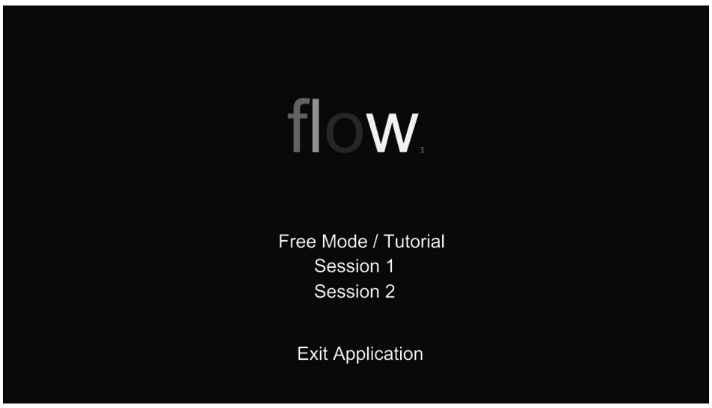


Figure 4-2: mainmenu interface of the flowfree clone

levels, and the color count will vary on each level. The difference between session one and session 2 is that puzzles in session one are taken randomly from the puzzles we have generated. In contrast, the puzzles in session two are sampled from the original sequence of *FlowFree* game. We aim to determine the difference between players' gameplay data on random difficulty level pacing (session 1, random pacing) and presumably increasing difficulty level pacing (session 2, original pacing). Whenever the players complete one

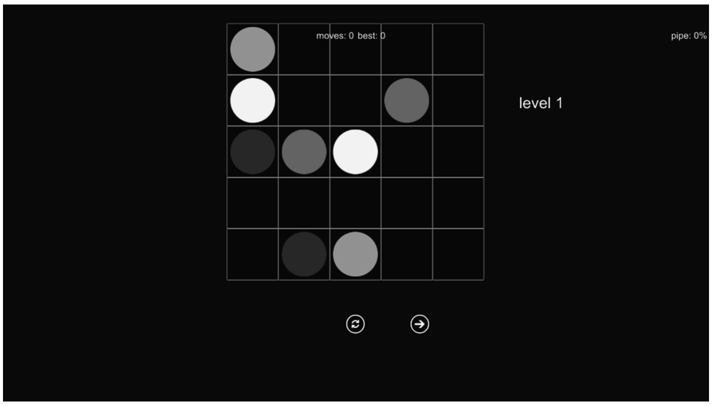


Figure 4-3: quantity of puzzle based on its board size and color count

level, the players will be asked to rate the puzzle's difficulty. We also provide a "skip" button if the player feels unable to complete the puzzle and a "retry" button to reset the state of the puzzle to the initial state (Figure 4-3. Once the player solved a puzzle, they will be prompted to rate the difficulty and interestingness of the puzzle. In this experiment, *difficulty* is defined as how difficult the puzzle was perceived by the player and *interestingness* is defined as how interesting the puzzle seems from the initial state and the solved state perceived by the player (Figure 4-4). Table 4.2 shows the data that we collected during the experiment session.

Table 4.2: Collected data information

Data Name	Description
SessionType	describe the session type
PlayerId	unique id of each player
StageId	stage / level number of the session
Moves	number of moves taken by the player to solve the puzzle
Time Taken	time taken for the player to solve the puzzle
Try	try / reload counter, it increases if player reset the puzzle state
Difficulty	player perceived difficulty value (1-very easy 5 very difficult)
Difficulty	player perceived interestingness value (1-very not interesting 5 very interesting)
IsGaveUp	a boolean value to indicate if player skipped the level

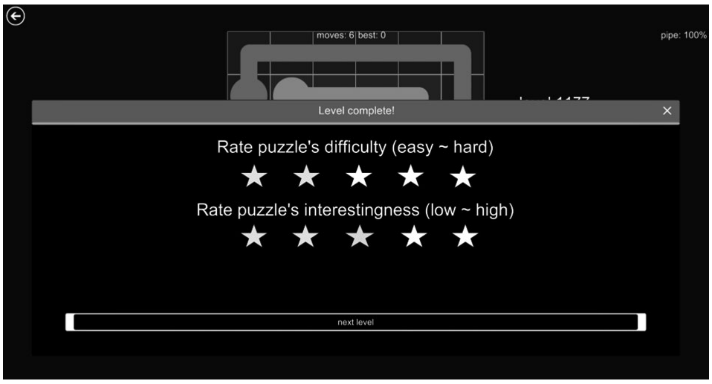


Figure 4-4: quantity of puzzle based on its board size and color count

4.4 Result Analysis

4.4.1 FlowFree Puzzles Analysis

In total, we have generated 3,373 puzzles ranging from 5x5 to 10x10 board size and various color count (number of color exist in puzzle) setting to allow for a range of puzzle difficulties. Figure 4-5 shows the quantity of puzzle based on its board size and color count. To analyze the puzzle, we consider the v value of motion-in-mind that we have discussed in section 2.4 to assume the puzzle's difficulty through its properties. Using the approach of scoring games, the success rate is defined as $v = \frac{G}{T}$, where G and T are the number of colors and the number of lines constructing the puzzle, respectively. As we can see in Figure 4-15, the v value decreases as the board size increases meaning that the solving rate decreases (m = 1 - v, solving hardness increases), making the puzzle become more difficult. Our findings showed that as the board size increased, the v value decreased, meaning that the puzzle became harder to solve. Conversely, as the number of colors in the puzzle increased, the v value also increased, making the puzzle easier to solve. This can be attributed to the fact that with more colors in the puzzle, there are fewer empty tiles and therefore less uncertainty for the player in terms of moves.

Overall, our research provides valuable insights into the relationship between board size, color count, and puzzle difficulty. These findings can be used to create more engaging and challenging puzzles in the future.

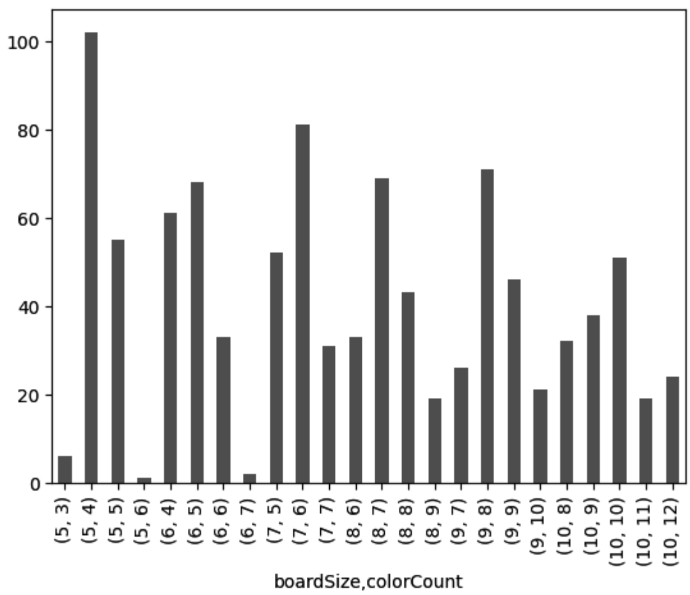


Figure 4-5: quantity of puzzle based on its board size and color count

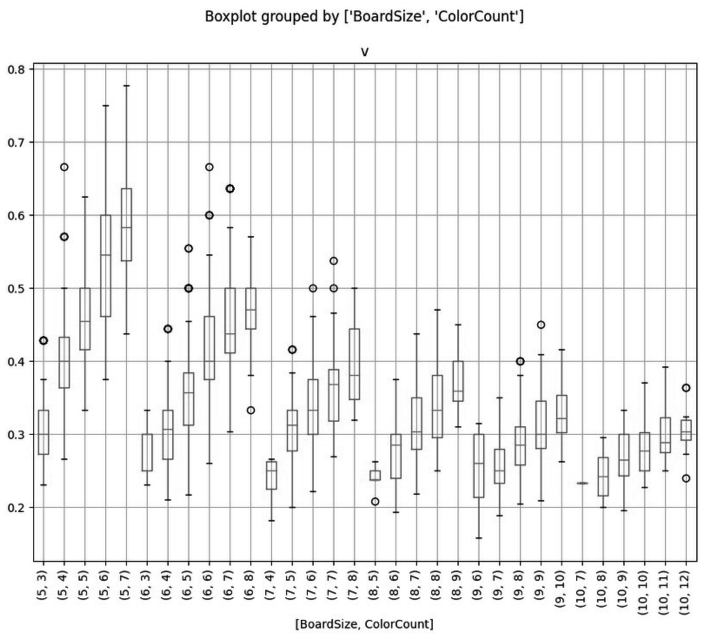


Figure 4-6: boxplot of puzzle's v based on its board size and color count

4.4.2 Human Data Analysis

Our analysis of human gameplay data involved collecting 2544 samples from 106 sessions, with 53 sessions of each session type. However, we encountered broken data, such as invalid formatting or incomplete data, in sessions where players completed less than 30% of the levels, etc. To ensure the reliability of our analysis, we conducted a data cleaning process that involved removing any invalid or incomplete data. This resulted in 2112 valid samples from 88 gameplay sessions, with 44 sessions from the original pacing and 44 from the random pacing.

In addition to the collected data, we analyzed players' performance using motion-in-mind. This involved analyzing players' movements and time taken to complete each level, as well as their overall experience of playing the game. Our analysis revealed some interesting findings. The average v of each puzzle level on the original pacing was displayed in Figure 4-7. We found that players generally performed better in the original pacing session, as the puzzle levels were carefully designed by a human designer to ensure a challenging yet enjoyable gameplay experience. On the other hand, Figure 4-8 showed the spread of v of each puzzle level that players played on the random pacing. The randomness of the pacing in the latter session made the v seem unstable. This suggests that the random pacing session did not provide a consistent level of difficulty or engagement for the players

We also compared the *moves* and *time taken* by players on both level pacing, as shown in Figures 4-9 and 4-14. Our analysis revealed that players took more moves and time to solve puzzles and fluctuated more in the random pacing sessions. Players also perceived puzzles in the random pacing session to be more difficult than in the original pacing session and were more likely to give up on them. This suggests that the random pacing session was not as enjoyable or engaging for the players as the original pacing session.

These results suggest that the puzzle generation process should provide puzzles based on the player's performance on the levels they have solved. For example, when a player takes a lot of time and moves to solve a puzzle, the following puzzle should be within or lower v value of previous levels while maintaining the board size or color count, and vice versa. This approach would ensure that the game remains challenging yet enjoyable for the players, providing them with a sense of progress and achievement.

In summary, our findings highlight the importance of designing puzzles with the players' experience and performance in mind. By doing so, we can create a more engaging and enjoyable gameplay experience, which in turn can lead to increased player retention and satisfaction.

Moreover, k-mean clustering was adopted from Scikit-learn [71] to analyze the players data, where the parameter of the k-mean cluster was $k \in [2, 7]$ with the parameter of total time player took to complete the session, average time per level, total moves player took to finish the session, average moves per level, average rating per level, total tries per session and the number of give up player did per session. The data is reduced By using silhouette coefficient assessment and its elbow point, it was found that a four-cluster is the most efficient for the analyzed data, as described in Table 4.3.

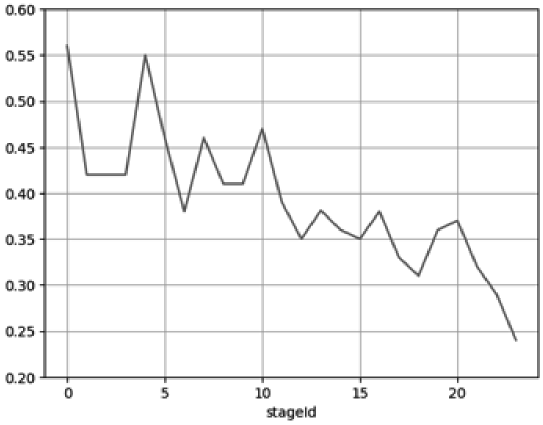


Figure 4-7: original pacing v average per level

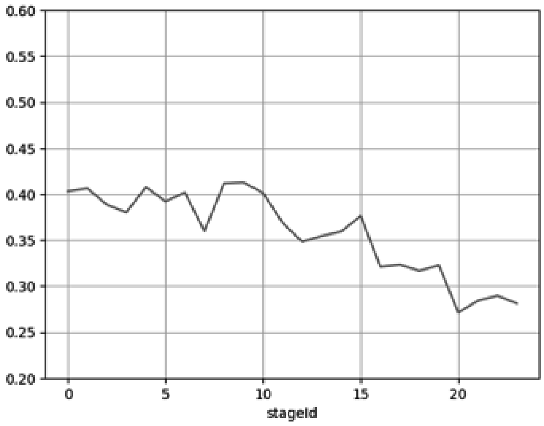


Figure 4-8: random pacing v average per level

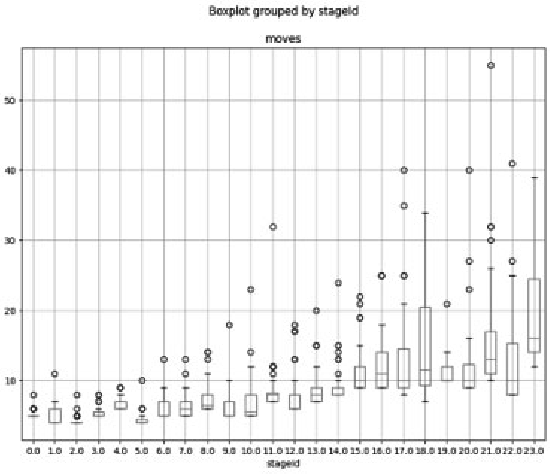


Figure 4-9: original pacing moves

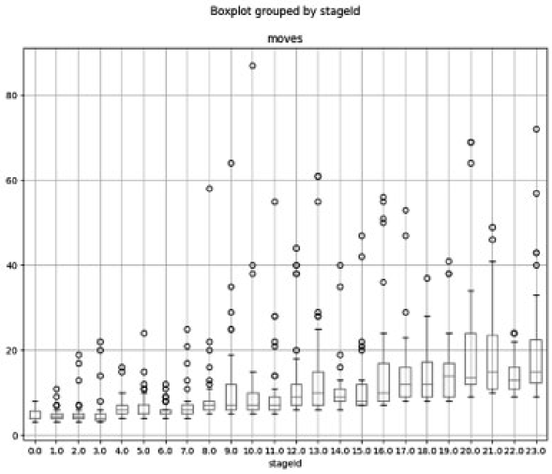


Figure 4-10: random pacing moves

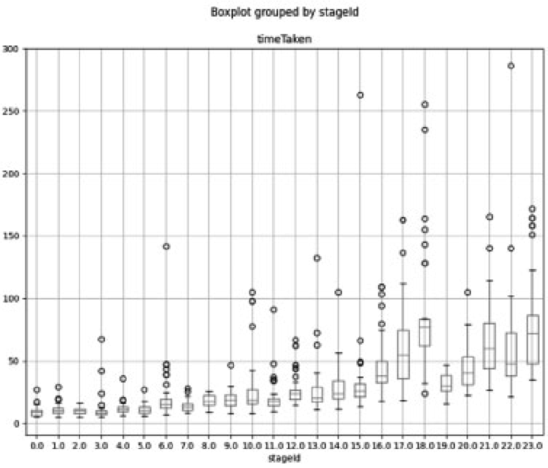


Figure 4-11: original pacing time taken

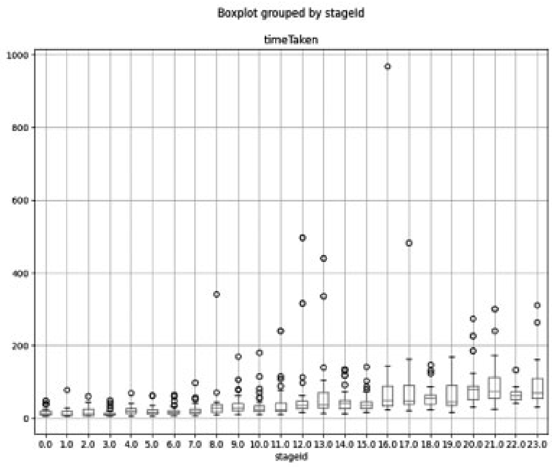


Figure 4-12: random pacing taken

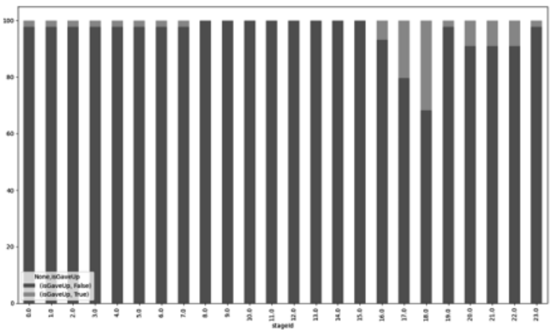


Figure 4-13: original pacing give up percentage

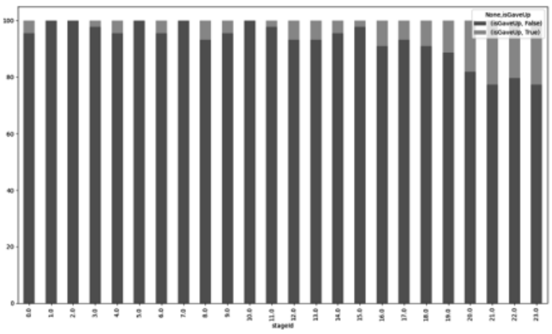


Figure 4-14: random pacing give up percentage

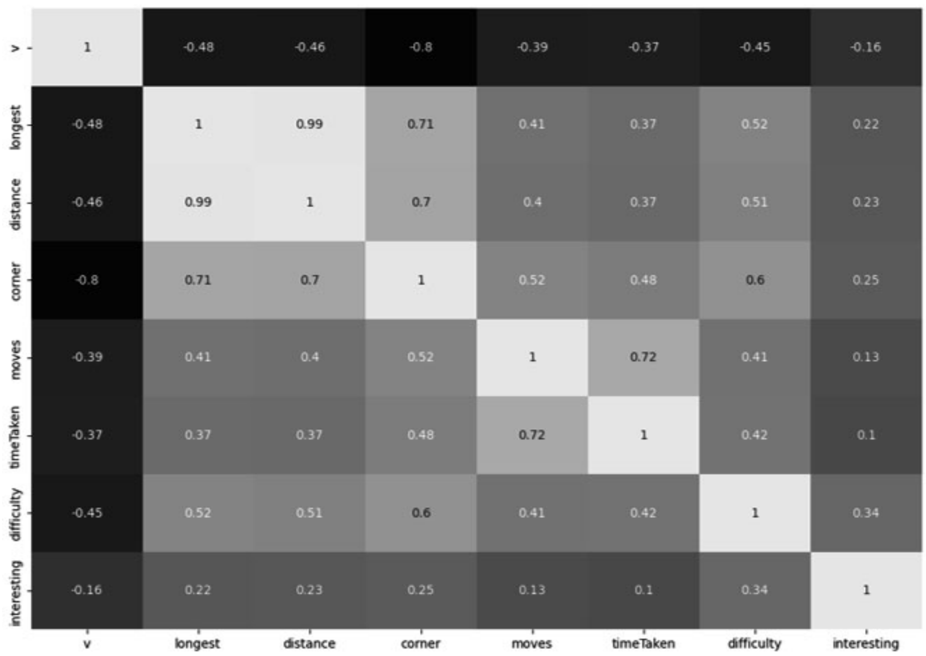


Figure 4-15: correlation map between puzzle metrics and gameplay data

Table 4.3: Clustering results of the player data

cluster	population size	total time	avg time	total moves	avg moves	avg rating	avg interestingness	total try	total giveup	remarks
1	37	740.29s	32.22s	217.7	9.52	2.27	4.7	35	0.95	Strong Players
2	12	1569s	71.97s	300	13.95	2.22	3.45	32	2.08	Time-Taking Players
3	20	1057s	45.20s	239.6	10.25	2.46	3.99	44.95	0.60	Try-and-Error Players
4	19	530s	26.21s	198	9.60	1.63	3.54	33.21	2.6	Weak Players

Table 4.4: Motion-in-Mind value of player data, $a = \frac{G}{T^2}$, $j = \frac{3G}{T^3}$

cluster	average colors	avg moves	avg time	average a	average j	remarks
1	6.704	8.455	24.986	**0.02397**	**0.00430**	Strong Players
2	6.578	9.293	62.958	0.00504	0.000418	Time-Taking Players
3	6.474	10.409	42.948	0.00909	0.001022	Try-and-Error Players
4	5.760	8.518	17.645	*0.03365*	*0.00771*	*Weak Players*

4.5 Discussion

The table 4.5 provides an interesting insight into the correlation between players' perceived difficulty (Rating) and the puzzle's properties, as well as the puzzle's v value and players' gameplay data. We can see that there is a moderate to strong positive correlation between players' perceived difficulty and the puzzle's properties such as board size, color count, and lines. This finding suggests that players do indeed find puzzles with larger board sizes, more colors, and more lines to be more difficult. Furthermore, we can observe that players' gameplay data, specifically the number of moves and time taken, have a negative correlation to the puzzle's v value. This means that as the puzzle's v value increases, indicating an easier puzzle, players tend to take fewer moves and less time to solve it. This is an interesting finding that may reflect players' willingness to expend more effort on puzzles that they perceive to be easier. Moreover, the correlation values between the puzzle's properties and players' perceived difficulty, as well as the negative correlation between players' gameplay data and the puzzle's v value, confirm the effectiveness of the v value in capturing the puzzle's difficulty through Motion-in-Mind's G/T model using the puzzle's color count and number of lines property. This suggests that the puzzle's v value can be a useful metric for designers to consider when creating puzzles, as it can provide a qualitative measure of the puzzle's difficulty level.

Table 4.5: Correlation between players perceived difficulty (rating) to puzzle properties, v and players' gameplay data

	v	corner count	moves taken	time taken
difficulty	**-0.45**	0.6	0.41	0.42
interestingness	**-0.16**	0.25	0.13	0.1

Based on the study's findings, we can further expand on the implications of the data. Firstly, we discovered that the value of v decreases as the size of the puzzle increases, which means that larger puzzles are generally more difficult for players to complete. However, we also found that the value of v decreases as the number of colors on the board increases. This suggests that increasing the number of colors can actually decrease the difficulty of a puzzle, which can be attributed to the fact that with more colors in the puzzle, there are fewer empty tiles and therefore less uncertainty for the player in terms of moves, which is a useful insight for puzzle game designers to keep in mind.

In addition, the comparison of player gameplay data between original and random level pacing allowed us to uncover important differences in player performance. Specifically, we observed that players' performance fluctuates a lot more in the random level pacing condition, which can be attributed to the inconsistent nature of each puzzle's difficulty. This inconsistency can overwhelm players, leading to more players giving up on solving the puzzles altogether.

These findings underscore the importance of providing players with puzzles that are appropriately challenging based on their performance. By using the v value to approximate puzzle difficulty through their properties, puzzle generation systems can provide players with puzzles that are at the right level of difficulty. This approach not only ensures that players are appropriately challenged, but also reduces the burden on human designers to hand-craft and design the contents (puzzles) needed for the game (*FlowFree*).

The clustering result in Table 4.3 provides valuable insights into the gameplay behavior of *FlowFree* players. It reveals the existence of four distinct groups of players based on their performance data. Cluster 2 and 4 demonstrate two different methods used by players to solve the puzzles in *FlowFree*. Cluster 4 players appear to rely on a try-and-error approach, as evidenced by their high values of total moves per session and average moves per level. In contrast, players in Cluster 2 take their time to solve the puzzle, as indicated by their high values of total time taken per session and average time taken per level. Cluster 1 is composed of strong players who are adept at solving puzzles, as evidenced by their lower total time, average time, total moves, and average moves compared to Clusters 2 and 4. Cluster 3 players also show similar characteristics to those of Cluster 1, but they are considered weak players due to their high number of total give-ups throughout the experimental session.

Furthermore, Table 4.4 shows the a, j values of player performance by cluster. a and j values represent the acceleration and jerk, respectively, in the players' performance. Acceleration refers to the rate of change in players' effort while solving the puzzle, while jerk represents the sudden or abrupt change in acceleration. Using the formula $a = \frac{G}{T^2}$ and $j = \frac{3G}{T^3}$, where G is the average number of colors in the puzzle and T is the sum of time and moves taken by the player to complete the puzzle, we found that each player type shows a different level of a and j values. Cluster 3, the weak player group, has the highest average total give-up per session, as well as the highest j and a values. This implies that when the player feels the acceleration and jerk in their mind is too high, it might overwhelm them, leading them to give up solving the puzzle or even quitting the game. This insight is crucial for game designers to pay attention to the players' experience when they progress through the puzzle or levels in the game, ensuring that the gameplay is not overwhelming for the players.

4.6 Chapter Summary

The use of PCG is quite popular for generating contents in games such as level, dungeons, board game rulesets and many other types of contents in games. Though the usage of PCG is popular in games, it is limited in puzzle games due to puzzles' strict constraints. In this study, we attempt to utilize motion-in-mind theory in puzzle generation. *FlowFree* is chosen as a test-case in this study where we developed a puzzle generation as well as its solver, generated *FlowFree* puzzles, analyzed its motion-in-mind value and collected human gameplay and human perceived difficulty data.

By analyzing the puzzle and its v value, we found that the puzzle's difficulty increases as the puzzle size grows but its difficulty also decreases when there are more colors involved in the puzzle. Moreover, by analyzing the correlation between players' perceived difficulty to puzzle's property, v value and players gameplay data we found that the more difficult the puzzle perceived by the players, the more effort required for the players to solve the puzzle. The correlation analysis also supports the motion-in-mind v value to approximate puzzle's difficulty as it shown by the negative correlation between players perceived difficulty and puzzle's v value. The study suggests that the "v" value can be a useful metric for designers to consider when creating puzzles, and that providing players with puzzles that are appropriately challenging based on their performance is important. The study also identified four distinct groups of players based on their performance data and found that each player type shows a different level of acceleration and jerk values, which designers should pay attention to when creating or generating puzzles for the players.

Chapter 5

Conclusion

5.1 General discussion on using motion-in-mind for data-driven game development

Data-driven game development has become increasingly popular in recent years, as game developers and publisher have recognized the value of using data to inform their decisions on every game development stages. Creation stage of game development involves pre-production and production phases, which focuses planning and implementation of the game. The data-driven approach in creation stage can provide developers with insights that can ensure the game can meet the preferences and expectation of the target audience as well as to tune and balance the game's contents quality to improve player engagement and experience. Optimization stage of game development involves post-production phase, which focuses on ensuring that the game is meeting the needs and preference of the target audience, understanding how the players are interacting with the game, and to make adjustment that improve the player engagement and experience. Developers should strive to create a game that is both engaging and provide a positive user experience. Although engagement and experience are related, they are not interchangeable. A game can be highly engaging but have a poor user experience if the gameplay mechanics are frustrating or the user interface is difficult to nagivate. Similarly, a game can have a great user experience but low engagement if it fails to capture the player's interest or provide enough motivation to keep playing. Furthermore, incorporating game refinement theory for data-driven approach in game development allows us to gain various insights related to the entertainment aspect and engagement in video game.

5.2 Addressing the Research Questions

The research conducted in this dissertation was driven by the following objectives: (1) How can we measure the entertainment aspect of video games from their Steam storefront data and how can we improve their visibility on the Steam Platform?; (2) What is the indicator to measure content quality (difficulty) and player performance in FlowFree? How does the value of this indicator differ based on player type?. The answer for each research questions are as follows:

5.2.1 Answer to RQ1:

The answer to RQ1 was obtained by measuring the game refinement theory and motion-in-mind values of steam games data that we have collected and processed in chapter 3. Table 5.1 shows GR, a, v, m and F values of the steam games grouped based on the total rating they received. By considering $v = \frac{G}{T}$ model, where G is the number of positive ratings and T is the total rating a game received, we found that there are an increasing pattern of steam games' v value as they get more rating from their players. In this context, v can be seen as a measure of the quantitative impression of the game by players. It represents the ratio of positive ratings to the total number of ratings received by the game. Thus, the v value is an indicator of the overall impression of the game by players and how well it is perceived by potential new players. It is a useful measure of a game's ability to attract new players and generate interest, and can ultimately impact the game's success. Interestingly, the a and F value decreases as the game received more rating from the players. a can be seen as a measure of the rate of changes in game's ability to attract new players, in other words, it is a measure of a game's growth to increase its player base, revenue or other metrics over a period of time. One explanation to this is that when a game becomes popular (for example, having a huge population of players), its ability to attract new players into the game is high while at the same time, the impact of getting new players into their game to the game's overall growth becomes lower since the focus of the game is no longer solely getting new players but also keeping the current players to play their game as well as converting free players into paying players to gain more revenue in the game. As a result, the impact of new players on the game's overall growth becomes less significant over time. Table 5.2 summarize the insights and direction of analysis in game based on its v and a. Additionally, we envisioned a new "experience-driven" business model (see Figure 3-15 in Chapter 3) which will enable game developers to improve the visibility of their games and direct them to their targeted users, as well as enable *steam* to recommend a more personalized game recommendation to *steam* users depending on the type of the game they interested in. Although it would take some investment (cost and time) from both developers and publisher to implement such system to dynamically compute aggregated scores of games' storefront data to rank 'expected experience' player can perceive through motion-in-mind model, this may improve user satisfaction where the user can discover games that they will like.

Table 5.1: G, T, a, GR, v, m and F of steam games based on their rating group

$Group(rating\,range)$	$G(positive)$	$T(total\,rating)$	a	GR	v	m	F	isIndie?
A (0 < x <= 20)	6.605	8.731	0.174	0.357	0.754	0.246	0.042804	TRUE
A (0 < x <= 20)	6.708	8.87	0.173	0.352	0.749	0.251	0.043423	FALSE
B (20 < x <= 100)	36.831	48.183	0.019	0.135	0.764	0.236	0.004484	TRUE
B (20 < x <= 100)	36.07	49.267	0.018	0.131	0.736	0.264	0.004752	FALSE
C (100 < x <= 1000)	267.313	335.419	0.003	0.056	0.786	0.214	6.4^{-4}	TRUE
C (100 < x <= 1000)	284.769	371.394	0.003	0.053	0.763	0.237	7.1^{-4}	FALSE
D (1000 < x <= 10000)	2636.479	3058.156	4^{-4}	0.019	0.848	0.152	6.1^{-5}	TRUE
D (1000 < x <= 10000)	2708.53	3377.045	3.4^{-4}	0.018	0.799	0.201	7^{-5}	FALSE
E (10000 < x <= 50000)	18586.69	20660.16	5.2^{-5}	0.007	0.895	0.105	5.4^{-6}	TRUE
E (10000 < x <= 50000)	19270.22	22930.05	4.6^{-5}	0.007	0.837	0.163	7.4^{-6}	FALSE
F (50000 < x <= 100000)	64859.489	70879.702	1.3^{-5}	0.004	0.913	0.087	1.1^{-6}	TRUE
F (50000 < x <= 100000)	62586.219	71643.734	1.3^{-5}	0.004	0.875	0.125	1.6^{-6}	FALSE
G (100000 < x <= 250000)	138026.458	151330.167	7^{-6}	0.002	0.916	0.084	5.8^{-7}	TRUE
G (100000 < x <= 250000)	132549.565	152131.935	6^{-6}	0.002	0.869	0.131	7.8^{-7}	FALSE
H (x >250000)	564420.125	611227.125	1.73^{-6}	0.001	0.916	0.084	1.45^{-7}	TRUE
H (x >250000)	694338.909	817800.273	1.9^{-6}	0.001	0.869	0.131	2.5^{-7}	FALSE

Table 5.2: the implication of v and f based on table 5.1

v	a	**implications**
low	low	unattractive games, low impact for new players, insights we can gain are "what to avoid" in your game
low	high	unattractive but growing games, insights should be analying their game updates
high	low	attractive games, insights should be the strategy to keep players' engagement stable.
high	high	attractive and growing games, insights should be the games' uniqueness and what makes it different with other similar games

5.2.2 Answer to RQ2:

The answer to RQ2 was obtained by generating *FlowFree* puzzles, collecting human gameplay data, measuring the motion-in-mind values of puzzle through its properties and analyzing human gameplay data as well as measuring its motion-in-mind value. We found that the v value of motion-in-mind can be used as an indicator to measure the puzzle's quality (difficulty) as it has negative correlation to players' perceived difficulty, meaning that the more difficult the puzzle perceived by the player, the lower v value will be. The players tend to take fewer moves and less time to solve the puzzle as the puzzle's v value increases, indicating the puzzle becomes easier.

Moreover, we performed k-mean clustering on the human gameplay data and found that there are 4 distinct groups of players in *FlowFree* (See Table 4.3). To evaluate players' performance, we measure the a and j which represent the rate of change in players' effort in solving the puzzle and the sudden or abrupt change in acceleration in mind of the player, respectively. We found that each player type shows a different level of a and j values where the weak player group (cluster 3), has the highest average total give-up per session, as well as the highest j and a values. This implies that when the player feels the acceleration and jerk in their mind is too high, it might overwhelm them, leading them to give up solving the puzzle or even quitting the game.

This insight is crucial for game designers to pay attention to the players' experience when they progress through the puzzle or levels in the game, ensuring that the gameplay is not overwhelming for the players.

5.2.3 Final remarks

In this study, we have adopted a data-driven approach that combines game refinement theory and motion-in-mind theory throughout various stages of game development. Table 5.3 presents the specific values derived from these theories and their application in each stage of game development. During the pre-production stage, the focus is on acquiring insights from existing games that can contribute to the blueprint of the new game. These insights serve as a foundation for analyzing referenced games and provide guidance for the kind of valuable information we can expect to gain from the analysis. Moving into the production stage, the focus shifts towards achieving a balance within the game's content and assessing its qualities, including factors such as difficulty. To accomplish this, we examine the individual elements of the game, and in this study, we explore the *flowfree* puzzle's difficulty shows a pattern in relation to the number of colors and board size. Finally, as we enter the post-production stage, the objective is to gain a deep understanding of how players engage with the game. This involves identifying various player types and characterizing their behaviors and preferences. By comprehending how players consume the game, we can further refine and enhance the overall gaming experience. By adopting the game refinement theory and motion-in-mind theory in a data-driven approach, we aim to optimize the game development process and create games that resonate with players on a profound level. The incorporation of these theories across different stages of development allows for a comprehensive and systematic exploration of game design, leading to more engaging and fulfilling gaming experiences.

Table 5.3: game development stages and game refinement/motion-in-mind values used

stage	**vars**	**usage**
pre-production	v, a	to determine game's ability to attract new players as well as the impact of gaining new players for the game's overall growth
production	v or m	to determine the game contents' difficulty
post-production	a, j	to measure the upper limit of discomfort can be handled by each user types

5.3 Future works

Further investigation could include acquiring information such as purchasing patterns, content usage, historical price changes and patterns of individual games; at the moment, unfortunately, the information could not be publicly obtained. In addition, a comparison between multiple game publishing platforms can be beneficial for game developers to understand the different strategies they can use as the users' consumption behavior might differ on each game publishing platforms as well as the storefront data that visible to the players. In this dissertation, we focus on *FlowFree* game which has a strict solvability constraints due to its nature being a puzzle game, further exploration of other games that have less strict solvability constraints and implementation of motion-in-mind adjusted content generation as player plays the game (real-time) can be essential for future work.

Bibliography

[1] How to set the right price for your game. `https://xsolla.com/blog/how-to-set-the-right-price-for-your-game`. Accessed: 2022-10-25.

[2] Daniel Ahmad. Steam spy estimation - daniel ahmad. `https://twitter.com/ZhugeEX/status/1215016446025228288`, 2020. Accessed: 2020-12-28.

[3] Ahmad Shahiran b. Ahmad Kamal, Shuria Saaidin, and Murizah Kassim. Recommender System: Rating predictions of Steam Games Based on Genre and Topic Modelling. In *2020 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pages 212–218, 2020. doi: 10.1109/I2CACIS49202.2020.9140194.

[4] Saiqa Aleem, Luiz Capretz, and Faheem Ahmed. Game development software engineering process life cycle: A systematic review. *Journal of Software Engineering Research and Development*, 4, 12 2016. doi: 10.1186/s40411-016-0032-7.

[5] Sky LaRell Anderson. Extraludic narratives: Online communities and video games. *Transformative Works and Cultures*, 28, 2018.

[6] Gleb Andreev, Deepak Saxena, and Jitendra Kumar Verma. Impact of Review Sentiment and Magnitude on Customers' Recommendations for Video Games. In *2021 International Conference on Computational Performance Evaluation (ComPE)*, pages 992–995, 2021. doi: 10.1109/ComPE53109.2021.9752380.

[7] Punyawee Anunpattana, Mohd Nor Akmal Khalid, Hiroyuki Iida, and Wilawan Inchamnan. Capturing potential impact of challenge-based gamification on gamified quizzing in the classroom. *Heliyon*, 7(12):e08637, 2021.

[8] Htun Pa Pa Aung, Mohd Nor Akmal Khalid, and Hiroyuki Iida. What constitutes fairness in games? a case study with scrabble. *Information*, 12(9):352, 2021.

[9] Myat Aung, Simon Demediuk, Yuan Sun, Ye Tu, Yu Ang, Siva Nekkanti, Shantanu Raghav, Diego Klabjan, Rafet Sifa, and Anders Drachen. The Trails of Just Cause 2: Spatio-Temporal Player Profiling in Open-World Games. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, FDG '19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450372176. doi: 10.1145/3337722.3337765. URL `https://doi.org/10.1145/3337722.3337765`.

[10] Pankaj Badoni, Avita Katal, M. Sweety Reddy, and Mudit Bhargava. Graphics vs Gameplay: A Comparative Analysis in Gaming. In *2022 2nd International Conference on Intelligent Technologies (CONIT)*, pages 1–8, 2022. doi: 10.1109/CONIT55038.2022.9847843.

[11] Dustin Bailey. Games released in 2019. `https://www.pcgamesn.com/steam/games-released-in-2019`, 2020. Accessed: 2020-12-28.

[12] Eric Bailey and Kazunori Miyata. Improving video game project scope decisions with data: An analysis of achievements and game completion rates. *Entertainment Computing*, 31:100299, 2019.

[13] Florian Baumann, Dominik Emmert, Hermann Baumgartl, and Ricardo Buettner. Hardcore gamer profiling: Results from an unsupervised learning approach to playing behavior on the steam platform. *Procedia Computer Science*, 126:1289–1297, 2018. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2018.08.078. URL `https://www.sciencedirect.com/science/article/pii/S1877050918313565`. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia.

[14] Cameron Browne. Automatic generation and evaluation of recombination games. 01 2008.

[15] Armir Bujari, Giacomo Quadrio, Claudio E Palazzi, Daniele Ronzani, Dario Maggiorini, and Laura Anna Ripamonti. Network traffic analysis of the steam game

system. In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 716–719. IEEE, 2017.

[16] John Businge, Moses Openja, David Kavaler, Engineer Bainomugisha, Foutse Khomh, and Vladimir Filkov. Studying android app popularity by cross-linking github and google play store. In *2019 IEEE 26th international conference on software analysis, evolution and reengineering (SANER)*, pages 287–297. IEEE, 2019.

[17] Irina Busurkina, Valeria Karpenko, Ekaterina Tulubenskaya, and Denis Bulygin. Game Experience Evaluation. A Study of Game Reviews on the Steam Platform. In *International Conference on Digital Transformation and Global Society*, pages 117–127. Springer, 2020.

[18] Heather Maxwell Chandler. *The Game Production Handbook*. Jones & Bartlett Learning, 2013.

[19] Megan Charity, Ahmed Khalifa, and Julian Togelius. Baba is Y'all: Collaborative Mixed-Initiative Level Design. In *2020 IEEE Conference on Games (CoG)*, pages 542–549, 2020. doi: 10.1109/CoG47356.2020.9231807.

[20] Simon Colton. Automated puzzle generation. In *In Proceedings of the AISB'02 Symposium on AI and Creativity in the Arts and Science*, 2002.

[21] Troy Costa Kohwalter, Leonardo Gresta Paulino Murta, and Esteban Walter Gonzalez Clua. Capturing Game Telemetry with Provenance. In *2017 16th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 66–75, 2017. doi: 10.1109/SBGames.2017.00016.

[22] Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper & Row, 1990. ISBN 0060162538.

[23] Aran Davies. What is a video game development life cycle? https://www.vice.com/en/article/nz7d8q/no-mans-sky-review. Accessed: 2022-11-22.

[24] Barbara De Kegel and Mads Haahr. Procedural puzzle generation: A survey. *IEEE Transactions on Games*, 12(1):21–40, 2020. doi: 10.1109/TG.2019.2917792.

[25] Kristy de Salas and Ian Lewis. Identifying types of Achievements. In *Proceedings of CGAMES'2013 USA*, pages 23–30, 2013. doi: 10.1109/CGames.2013.6632599.

[26] Sam Desatoff. Year over year game release numbers slowed for steam in 2019. `https://gamedaily.biz/article/1489/year-over-year-game-release-numbers-slowed-for-steam-in-2019`, 2020. Accessed: 2020-12-28.

[27] Steam Direct. Steam direct. `https://partner.steamgames.com/steamdirect`, 2017. Accessed: 2020-12-28.

[28] Anders Drachen, Magy Seif El-Nasr, and Alessandro Canossa. *Game Analytics: Maximizing the Value of Player Data*. Springer, 2013.

[29] Lingyu Du. Steam Game Discount Prediction Using Machine Learning Methods. In *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pages 149–152, 2021. doi: 10.1109/MLBDBI54094.2021.00037.

[30] Lukas Eberhard, Patrick Kasper, Philipp Koncar, and Christian Gütl. Investigating Helpfulness of Video Game Reviews on the Steam Platform. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 43–50, 2018. doi: 10.1109/SNAMS.2018.8554542.

[31] Henrik Edholm, Mikaela Lidström, Jan-Philipp Steghöfer, and Håkan Burden. Crunch time: The reasons and effects of unpaid overtime in the games industry. In *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*, pages 43–52. IEEE, 2017.

[32] Fabrício Guedes Faria, Leonardo Tórtoro Pereira, and Cláudio Fabiano Motta Toledo. Adaptive Generation of Characters for Tabletop Role Playing Games. In *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 39–46, 2019. doi: 10.1109/SBGames.2019.00016.

[33] Alia Fatima, Tayyaba Rasool, and Usman Qamar. GDGSE: Game Development with Global Software Engineering. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pages 1–9, 2018. doi: 10.1109/GEM.2018.8516498.

[34] Enrico Gandolfi. Enjoying death among gamers, viewers, and users: A network visualization of dark souls 3's trends on twitch. tv and steam platforms. *Information Visualization*, 17(3):218–238, 2018.

[35] Yuexian Gao, Chang Liu, Naying Gao, Mohd Nor Akmal Khalid, and Hiroyuki Iida. Nature of arcade games. *Entertainment Computing*, 41:100469, 2022.

[36] Emily Greer. Data-driven or data-blinded? uses and abuses of analytics in games. https://www.youtube.com/watch?v=MR2rorssk9c, 2018. Accessed: 2022-11-23.

[37] Tibor Guzsvinecz. The correlation between positive reviews, playtime, design and game mechanics in souls-like role-playing video games. *Multimedia Tools and Applications*, 82(3):4641–4670, Jan 2023. ISSN 1573-7721. doi: 10.1007/s11042-022-12308-1. URL `https://doi.org/10.1007/s11042-022-12308-1`.

[38] Erin Jonathan Hastings, Ratan K. Guha, and Kenneth O. Stanley. Automatic content generation in the galactic arms race video game. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4):245–263, 2009. doi: 10.1109/TCIAIG.2009.2038365.

[39] Maurice Hendrix, Tyrone Bellamy-Wood, Sam McKay, Victoria Bloom, and Ian Dunwell. Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, 11(4):320–327, 2019. doi: 10.1109/TG.2018.2791019.

[40] Daniel Hernandez, Charles Takashi Toyin Gbadamosi, James Goodman, and James Alfred Walker. Metagame Autobalancing for Competitive Multiplayer Games. In *2020 IEEE Conference on Games (CoG)*, pages 275–282. IEEE, 2020.

[41] Kenneth Hullett, Nachiappan Nagappan, Eric Schuh, and John Hopson. Data analytics for game development: NIER track. In *2011 33rd International Conference on Software Engineering (ICSE)*, pages 940–943, 2011. doi: 10.1145/1985793.1985952.

[42] Robin Hunicke. The Case for Dynamic Difficulty Adjustment in Games. ACE '05, page 429–433, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 1595931104. doi: 10.1145/1178477.1178573. URL `https://doi.org/10.1145/1178477.1178573`.

[43] H. Iida and M. N. A. Khalid. Using games to study law of motions in mind. *IEEE Access*, 8:138701–138709, 2020.

[44] Hiroyuki Iida, Kazutoshi Takahara, Jun Nagashima, Yoichiro Kajihara, and Tsuyoshi Hashimoto. An application of game-refinement theory to Mah Jong. In *International Conference on Entertainment Computing*, pages 333–338. Springer, 2004.

[45] Richard JE James, Claire O'Malley, and Richard J Tunney. Understanding the psychology of mobile gambling: A behavioural synthesis. *British Journal of Psychology*, 108(3):608–625, 2017.

[46] Jennifer Javornik. Baby steps in game development: Discovery-only, alpha release, prototype, or mvp? https://www.filamentgames.com/blog/baby-steps-in-game-development-discovery-only-alpha-release-prototype-or-mvp/, 2018. Accessed: 2023-01-03.

[47] Daniel Johnson, Madison Klarkowski, Kellie Vella, Cody Phillips, Mitchell McEwan, and Christopher N Watling. Greater rewards in videogames lead to more presence, enjoyment and effort. *Computers in Human Behavior*, 87:66–74, 2018.

[48] Ha-Na Kang, Hye-Ryeon Yong, and Hyun-Seok Hwang. A study of analyzing on online game reviews using a data mining approach: Steam community data. *International Journal of Innovation, Management and Technology*, 8(2):90, 2017.

[49] Xiaohan Kang, Hong Ri, Mohd Nor Akmal Khalid, and Hiroyuki Iida. Addictive games: Case study on multi-armed bandit game. *Information*, 12(12):521, 2021.

[50] Mohd Nor Akmal Khalid and Hiroyuki Iida. Objectivity and subjectivity in games: Understanding engagement and addiction mechanism. *IEEE Access*, 9:65187–65205, 2021. doi: 10.1109/ACCESS.2021.3075954.

[51] Anastasia Khomych. Indie vs game publisher: What's better for your game? `https://blog.getsocial.im/indie-vs-game-publisher-whats-better-for-your-game/`, 2022. Accessed: 2022-08-22.

[52] Kotsuma Kouichi and Takenaga Yasuhiko. Np-completeness and enumeration of number link puzzle. *IEICE technical report. Theoretical foundations of Computing*, 109:1–7, 2010.

[53] Dedy Prasetya Kristiadi, Ferry Sudarto, Dian Sugiarto, Richard Sambera, Harco Leslie Hendric Spits Warnars, and Kiyota Hashimoto. Game Development with Scrum methodology. In *2019 International Congress on Applied Information Technology (AIT)*, pages 1–6, 2019. doi: 10.1109/AIT49014.2019.9144963.

[54] Ben Kybartas and Rafael Bidarra. A survey on story generation techniques for authoring computational narratives. *IEEE Transactions on Computational Intelligence and AI in Games*, 9(3):239–253, 2017. doi: 10.1109/TCIAIG.2016.2546063.

[55] Ben Kybartas and Clark Verbrugge. Analysis of regen as a graph-rewriting system for quest generation. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(2):228–242, 2014. doi: 10.1109/TCIAIG.2013.2290088.

[56] Gorm Lai, Frederic Fol Leymarie, and William Latham. On mixed-initiative content creation for video games. *IEEE Transactions on Games*, 14(4):543–557, 2022. doi: 10.1109/TG.2022.3176215.

[57] Xiaozhou Li and Boyang Zhang. A preliminary network analysis on steam game tags: another way of understanding game genres. In *Proceedings of the 23rd International Conference on Academic Mindtrek*, pages 65–73, 2020.

[58] Xiaozhou Li, Chien Lu, Jaakko Peltonen, and Zheying Zhang. A statistical analysis of Steam user profiles towards personalized gamification. In *3rd International GamiFIN Conference, GamiFIN 2019*. CEUR-WS, 2019.

[59] Xiaozhou Li, Zheying Zhang, and Kostas Stefanidis. A data-driven approach for video game playability analysis based on players' reviews. *Information*, 12(3):129, 2021.

[60] Seth Jordan Lickteig. *Metagaming: Cognition in gaming environments and systems.* Kansas State University, 2020.

[61] Dayi Lin, Cor-Paul Bezemer, and Ahmed E Hassan. Studying the urgent updates of popular games on the steam platform. *Empirical Software Engineering*, 22(4): 2095–2126, 2017.

[62] Dayi Lin, Cor-Paul Bezemer, and Ahmed E Hassan. An empirical study of early access games on the steam platform. *Empirical Software Engineering*, 23(2):771–799, 2018.

[63] Dayi Lin, Cor-Paul Bezemer, Ying Zou, and Ahmed E Hassan. An empirical study of game reviews on the steam platform. *Empirical Software Engineering*, 24(1): 170–207, 2019.

[64] Kirk Lundblade. Crowning Achievements: A historioludic analysis of the achievement metagame in historical simulation games. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*, pages 1–11, 2021.

[65] Norman Macrae. *John von Neumann: The scientific genius who pioneered the modern computer, game theory, nuclear deterrence, and much more*. Plunkett Lake Press, 2019.

[66] Luiz Bernardo Martins Kummer, Julio Cesar Nievola, and Emerson Cabrera Paraiso. Applying Commitment to Churn and Remaining Players Lifetime Prediction. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–8, 2018. doi: 10.1109/CIG.2018.8490443.

[67] Pejman Mirza-Babaei, Naeem Moosajee, and Brandon Drenikow. Playtesting for indie studios. In *Proceedings of the 20th International Academic Mindtrek Conference*, pages 366–374, 2016.

[68] Brad Morris. Game hooks. https://www.gamedeveloper.com/design/game-hooks, 2010. Accessed: 2023-01-05.

[69] Sang-Gyu Nam, Chu-Hsuan Hsueh, and Kokolo Ikeda. Generation of game stages with quality and diversity by reinforcement learning in turn-based rpg. *IEEE Transactions on Games*, 14(3):488–501, 2022. doi: 10.1109/TG.2021.3113313.

[70] George Panagiotopoulos and George Giannakopoulos. A Study on Video Game Review Summarization. pages 36–43, 12 2019. doi: 10.26615/978-954-452-058-8_006.

[71] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL `http://jmlr.org/papers/v12/pedregosa11a.html`.

[72] Aurimas Petrovas and Romualdas Baušys. Automated Digital Terrain Elevation Modification By Procedural Generation Approach. In *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, pages 1–5, 2019. doi: 10.1109/eStream.2019.8732171.

[73] Cristiano Politowski, Fabio Petrillo, Gabriel C Ullmann, and Yann-Gaël Guéhéneuc. Game industry problems: An extensive analysis of the gray literature. *Information and Software Technology*, 134:106538, 2021.

[74] Alexandru Predescu and Mariana Mocanu. A data driven survey of video games. In *2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–6, 2020. doi: 10.1109/ECAI50035.2020.9223203.

[75] Anggina Primanita, Mohd Nor Akmal Khalid, and Hiroyuki Iida. Computing games: bridging the gap between search and entertainment. *IEEE Access*, 9:72087–72102, 2021.

[76] Kiefer Stefano Ranti, Kelvin Salim, and Abba Suganda Girsang. Clustering steam user behavior data using K-Prototypes algorithm. In *IOP Conference Series: Materials Science and Engineering*, volume 725, page 012105. IOP Publishing, 2020.

[77] Simão Reis, Luís Paulo Reis, and Nuno Lau. VGC AI Competition-A New Model of Meta-Game Balance AI Competition. In *2021 IEEE Conference on Games (CoG)*, pages 01–08. IEEE, 2021.

[78] Eric Ries. *Lean Startup: How Today's Entrepreneurs Use Continous Innovation to Create Radically Successful Businesses.* Crown Business, New York, 2011. ISBN 978-0-307-88789-4.

[79] Muhammad Nazhif Rizani, Mohd Nor Akmal Khalid, Hiroyuki Iida, and Reza Firsandaya Malik. Analysis of Business Model Transition Based on Active User and Review Value. In *International Conference on the Foundations of Digital Games*, pages 1–4, 2020.

[80] Moni Sahu, Ajay Vikram Singh, and Sunil Kumar Khatri. A Classical Constraint Satisfaction Problem and its Solution using Artificial Intelligence. In *2019 Amity International Conference on Artificial Intelligence (AICAI)*, pages 429–433, 2019. doi: 10.1109/AICAI.2019.8701325.

[81] T. Schiex and G. Verfaillie. Nogood Recording for static and dynamic constraint satisfaction problems. In *Proceedings of 1993 IEEE Conference on Tools with Al (TAI-93)*, pages 48–55, 1993. doi: 10.1109/TAI.1993.633935.

[82] Magy Seif El-Nasr. Developing Games That Capture and Engage Users. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, pages 9–10, 2019. doi: 10.1109/ICSE-Companion.2019.00025.

[83] Aaron Sexton, Erik Anderson, and James Hereford. An improved Evolutionary Strategy with dynamic mutation rate applied to Kakuro puzzles. In *2019 SoutheastCon*, pages 1–6, 2019. doi: 10.1109/SoutheastCon42311.2019.9020625.

[84] Burrhus Frederic Skinner. 'superstition'in the pigeon. *Journal of experimental psychology*, 38(2):168, 1948.

[85] David Stammer, Tobias Günther, and Mike Preuss. Player-adaptive Spelunky level generation. In *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 130–137, 2015. doi: 10.1109/CIG.2015.7317948.

[86] Steam. Introducing steam reviews. `https://store.steampowered.com/reviews/`. Accessed: 2022-07-11.

[87] Steam. More changes addressing fake games. https://steamcommunity.com/groups/steamworks/announcements/detail/3077529424431732424, 2018. Accessed: 2022-01-28.

[88] Nadia Stefyn. How video games are made: the game development process. https://www.cgspectrum.com/blog/game-development-process, 2022. Accessed: 2022-11-22.

[89] Juego Studios. What is the game development life cycle? https://www.gamedeveloper.com/business/what-is-the-game-development-life-cycle-, 2017. Accessed: 2022-11-22.

[90] Arie Pratama Sutiono, Ayu Purwarianti, and Hiroyuki Iida. A mathematical model of game refinement. In *International Conference on Intelligent Technologies for Interactive Entertainment*, pages 148–151. Springer, 2014.

[91] Raquel Tabuyo-Benito, Hayretdin Bahsi, and Pedro Peris-Lopez. Forensics Analysis of an On-line Game over Steam Platform. In *International Conference on Digital Forensics and Cyber Crime*, pages 106–127. Springer, 2018.

[92] Tavrox. Three ways to do market research for your video game. https://medium.com/game-marketing/three-ways-to-do-market-research-for-your-video-game-a31597e79d3, 2017. Accessed: 2022-10-28.

[93] Haydn Taylor. Major slowdown for number of new releases on steam in 2019. https://www.gamesindustry.biz/articles/2020-01-07-major-slowdown-for-number-of-new-releases-on-steam-in-2019, 2020. Accessed: 2020-12-28.

[94] Anne Mette Thorhauge and Rune KL Nielsen. Epic, steam, and the role of skin-betting in game (platform) economies. *Journal of Consumer Culture*, 21(1):52–67, 2021.

[95] Julian Togelius, Georgios N. Yannakakis, Kenneth O. Stanley, and Cameron Browne. Search-based procedural content generation: A taxonomy and survey.

IEEE Transactions on Computational Intelligence and AI in Games, 3(3):172–186, 2011. doi: 10.1109/TCIAIG.2011.2148116.

[96] Xinge Tong. Positioning game review as a crucial element of game user feedback in the ongoing development of independent video games. *Computers in Human Behavior Reports*, 3:100077, 2021.

[97] Eric J Toy, Jaya VHH Kummaragunta, and Jin Soung Yoo. Large-Scale Cross-Country Analysis of Steam Popularity. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1054–1058. IEEE, 2018.

[98] Gabriel C. Ullmann, Cristiano Politowski, Yann-Gaël Guéhéneuc, and Fabio Petrillo. What Makes a Game High-rated? Towards Factors of Video Game Success. In *2022 IEEE/ACM 6th International Workshop on Games and Software Engineering (GAS)*, pages 16–23, 2022. doi: 10.1145/3524494.3527623.

[99] Valve. Counter-strike: Global offensive. `https://store.steampowered.com/app/730/CounterStrike_Global_Offensive/`, 2012. Accessed: 2021-08-01.

[100] Roland van der Linden, Ricardo Lopes, and Rafael Bidarra. Procedural generation of dungeons. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(1):78–89, 2014. doi: 10.1109/TCIAIG.2013.2290371.

[101] Ajay Vasal, Sanjeev Vohra, Emmanuelle Payan, and Yusof Seedat. closing data value gap. https://www.accenture.com/$_acnmedia/pdf-108/accenture-closing-data-value-gap-fixed.pdf. Accessed: 2023-01-03$.

[102] Fanny Vaudour and Aleksej Heinze. Software as a service: Lessons from the video game industry. *Global Business and Organizational Excellence*, 39(2):31–40, 2020.

[103] Breno M. F. Viana and Selan R. dos Santos. A Survey of Procedural Dungeon Generation. In *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 29–38, 2019. doi: 10.1109/SBGames.2019.00015.

[104] Dulakshi Vihanga, Michael Barlow, Erandi Lakshika, and Kathryn Kasmarik. Weekly Seasonal Player Population Patterns in Online Games: A Time Series Clustering Ap-

proach. In *2019 IEEE Conference on Games (CoG)*, pages 1–8, 2019. doi: 10.1109/CIG.2019.8848108.

[105] Patrick Walker. Early access popularity growing, but only 25% have released as a full game. https://www.gamesindustry.biz/articles/2014-11-13-early-access-popularity-growing-but-only-25-percent-have-released-as-a-full-2014. Accessed: 2021-01-05.

[106] Dylan Wang, Melody Moh, and Teng-Sheng Moh. Using Deep Learning and Steam User Data for Better Video Game Recommendations. In *Proceedings of the 2020 ACM Southeast Conference*, pages 154–159, 2020.

[107] Yitong Wang, Bai Liu, Xiangyi Zhang, Runhua Shi, and Mingwu Zhang. A Gini Coefficient Measurement Method of Income Distribution Gap Based on Privacy-preservation. In *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, pages 1059–1064, 2021. doi: 10.1109/ICSIP52628.2021.9688722.

[108] Travis W Windleharth, Jacob Jett, Marc Schmalz, and Jin Ha Lee. Full steam ahead: A conceptual analysis of user-supplied tags on steam. *Cataloging & Classification Quarterly*, 54(7):418–441, 2016.

[109] Kang Xiaohan, Mohd Nor Akmal Khalid, and Hiroyuki Iida. Player satisfaction model and its implication to cultural change. *IEEE Access*, 8:184375–184382, 2020.

[110] Georgios N. Yannakakis and Julian Togelius. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing*, 2(3):147–161, 2011. doi: 10.1109/T-AFFC.2011.6.

[111] Haiyan Yin, Linbo Luo, Wentong Cai, Yew-Soon Ong, and Jinghui Zhong. A data-driven approach for online adaptation of game difficulty. In *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 146–153, 2015. doi: 10.1109/CIG.2015.7317918.

[112] Chris Young. 7. unity production: Capturing the everyday game maker market. In *Game Production Studies*, pages 141–158. Amsterdam University Press, 2021.

[113] Andrei Zanescu, Marc Lajeunesse, and Martin French. Speculating on steam: consumption in the gamblified platform ecosystem. *Journal of Consumer Culture*, 21(1):34–51, 2021.

[114] Zeliang Zhang, Kang Xiaohan, Mohd Nor Akmal Khalid, and Hiroyuki Iida. Bridging ride and play comfort. *Information*, 12(3):119, 2021.

[115] Zeliang Zhang, Naying Gao, Siqi Li, Mohd Nor Akmal Khalid, and Hiroyuki Iida. Action games evolution analysis: A case study using the god of war series. *IEEE Access*, 10: 123697–123710, 2022. doi: 10.1109/ACCESS.2022.3224469.

[116] Alexander E. Zook and Mark O. Riedl. A Temporal Data-Driven Player Model for Dynamic Difficulty Adjustment. In *Proceedings of the Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, AIIDE'12, page 93–98. AAAI Press, 2012.

Publications

Journal Papers

1. Rizani, M. N., Khalid, M. N. A., and Iida, H. (2022). Application of Meta-Gaming Concept to The Publishing Platform: Analysis of The Steam Games Platform. Information 2023, 14, 110.

International Conferences (Proceedings)

1. M. N. Rizani, S. Thavamuni, M. N. A. Khalid and H. Iida. (2021). Steam Game Achievement Analysis. The First Artificial Intelligence and Entertainment Science Workshop (AIES 2021), pp. 67-69, 2021.
2. Rizani, M. N.,Liu, C., Abuluaih, S., Khalid, M. N. A., and Iida, H., Motion-in-Mind Apporach Level Generation in FlowFree, the Artificial Intelligence and Entertainment Science Workshop (AIES 2021) Online, Japan.
3. Liu, C., Rizani, M. N., Khalid, M. N. A., and Iida, H., Entertainment Analysis of Single-Agent Game: Case Study in Match-3 puzzle Game, the Artificial Intelligence and Entertainment Science Workshop (AIES 2021) Online, Japan.
4. Muhammad Nazhif Rizani, Xiaohan Kang, Mohd Nor Akmal Khalid, Hiroyuki Iida and Saajid Abuluaih. Puzzle Generation and Analysis in FlowFree. The 10th ASEAN Workshop on Information Science and Technology (AWIST 2022), pp. 173-182, 2022.

Presentations at Academic Conferences (Oral)

1. M. N. Rizani, S. Thavamuni, M. N. A. Khalid and H. Iida. (2021). Steam Game Achievement Analysis. The First Artificial Intelligence and Entertainment Science Workshop (AIES 2021), pp. 67-69, 2021.
2. Rizani, M. N.,Liu, C., Abuluaih, S., Khalid, M. N. A., and Iida, H., Motion-in-Mind Apporach Level Generation in FlowFree, the Artificial Intelligence and Entertainment Science Workshop (AIES 2021) Online, Japan.
3. Liu, C., Rizani, M. N., Khalid, M. N. A., and Iida, H., Entertainment Analysis of Single-Agent Game: Case Study in Match-3 puzzle Game, the Artificial Intelligence and Entertainment Science Workshop (AIES 2021) Online, Japan.
4. Muhammad Nazhif Rizani, Xiaohan Kang, Mohd Nor Akmal Khalid, Hiroyuki Iida and Saajid Abuluaih. Puzzle Generation and Analysis in FlowFree. The 10th ASEAN Workshop on Information Science and Technology (AWIST 2022), pp. 173-182, 2022.