CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

AI-driven Player Experience Modeling in Serious Games for Software Engineering

Master's Thesis in Computer science and engineering

Julia Dahlberg

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2025

Master's Thesis 2025

AI-driven Player Experience Modeling in Serious Games for Software Engineering

Julia Dahlberg

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2025

AI-driven Player Experience Modeling in Serious Games for Software Engineering
Julia Dahlberg

Supervisor: Linda Erlenhov, Department of Computer Science and Engineering
Examiner: Birgit Penzenstadler, Department of Computer Science and Engineering
Examiner-in-Practice: Cristina Martinez Montes, Department of Computer Science and Engineering

Master's Thesis 2025
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2025

Julia Dahlberg
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Serious games offer great potential for enhancing software engineering education, yet the use of AI-driven Player Experience Modelling (PEM) remains under-explored in this context. This study investigates how AI-enhanced adaptive features such as difficulty adjustment and dynamic guidance affect player engagement and skill acquisition in the programming game Elara. By comparing AI-driven and traditional versions of the game, the research highlights the benefits of personalized gameplay for supporting learning and motivation. The study demonstrates the promise of AI in tailoring educational experiences and calls for future work to incorporate automated player profiling, larger datasets, and emotional state recognition to further improve adaptive learning in serious games.

Keywords: Player Experience Modelling, PEM, Serious Games, Educational Games, AI-driven PEM, AI-enhanced learning, Personalized learning, Adaptive gameplay

Acknowledgements

I would like to sincerely thank my supervisor Linda Erlenhov, for her invaluable guidance, support, and encouragement throughout this project. Her expertise and thoughtful feedback were instrumental in shaping the direction and quality of this research.

I also extend my gratitude to my examiner-in-practice Cristina Martinez Montes, for her time and effort in reviewing my work and providing constructive insights that helped improve the final outcome. Additionally, I wish to acknowledge my examiner Birgit Penzenstadler, for her role in the evaluation of this thesis and her contribution to its final assessment.

Finally, I am deeply grateful to all the participants who took part in this study. Their time, openness, and valuable insights made this research possible.

Thank you all for your dedication and support.

Julia Dahlberg, Gothenburg, June 2025

Contents

List of Figures xi

List of Tables xv

Acronyms xvii

1 Introduction 1
1.0.1 Player Experience Modeling (PEM) . . . 1
1.1 Problem Description . . . 2
1.1.1 Scope . . . 2
1.1.2 Research questions . . . 2
1.2 Purpose of the Study . . . 3
1.3 Significance of the Study . . . 3

2 Background 5
2.1 Serious Games and Related Concepts . . . 5
2.2 Personalisation in games . . . 6
2.2.1 Player Modeling . . . 7
2.2.1.1 Player Behavioural Modeling . . . 7
2.2.1.2 Player Experience . . . 8
2.2.1.3 Player Experience Modeling (PEM) . . . 9
2.3 Programming in educational games . . . 10
2.3.1 Educational Goals . . . 10
2.3.2 Features and Limitations . . . 11

3 Related Work 13
3.1 Player Experience Modelling . . . 16
3.1.1 Feedback & Debriefing . . . 16
3.1.2 Game difficulty balancing . . . 16
3.2 Demographics . . . 17
3.2.1 Player Types . . . 17

4 Method 19
4.1 Design Science Research (DSR) Framework . . . 19
4.1.1 Activity 1: Problem identification and motivation. . . . 20
4.1.2 Activity 2: Define the objectives for a solution. . . . 20
4.1.3 Activity 3: Design and development. . . . 20

4.1.3.1 Elara . . . 20
4.1.3.2 Performance Tracking . . . 21
4.1.3.3 Player types . . . 23
4.1.3.4 Dynamic elements . . . 24
4.1.4 Activity 4: Demonstration. . . . 25
4.1.5 Activity 5: Evaluation. . . . 26
4.1.5.1 Participants . . . 26
4.1.5.2 Data Collection . . . 27
4.1.5.3 Data Analysis . . . 28
4.1.6 Activity 6. Communication. . . . 29

5 Results **31**
5.1 Participant Demographics and Background . . . 31
5.2 Quantitative Outcomes . . . 32
5.2.1 Pre- and Post-Assessment Coding Scores . . . 32
5.2.2 Gameplay Metrics . . . 33
5.2.3 Game Experience Feedback Surveys . . . 34
5.2.3.1 Frustration and boredom . . . 34
5.2.3.2 Perceived engagement . . . 34
5.2.3.3 Motivation . . . 35
5.2.3.4 Challenge appropriateness . . . 36
5.2.3.5 Guidance and hints . . . 37
5.3 Qualitative Findings . . . 37
5.3.1 Engagement and Enjoyment . . . 38
5.3.2 Challenge and Difficulty . . . 39
5.3.3 Learning and Knowledge Acquisition . . . 40
5.3.4 Usability and Game Mechanics . . . 41
5.3.5 Emotional Response . . . 42
5.3.6 Preferences and Comparisons . . . 43
5.4 Summary of comparison . . . 44
5.4.1 Gameplay Behaviour and Learning Outcomes . . . 44
5.4.2 Subjective Experience and Qualitative Insights . . . 44
5.4.3 Overall Comparison . . . 45

6 Discussion **47**
6.1 Delimitations . . . 47
6.2 Research Questions . . . 47
6.2.1 RQ1: How AI-driven PEM Improves Engagement and Skill Acquisition . . . 48
6.2.1.1 Engagement . . . 48
6.2.1.2 Skill Acquisition . . . 48
6.2.2 RQ2: Advantages and Disadvantages of AI-driven PEM . . . 49
6.2.2.1 Advantages of AI-driven PEM . . . 49
6.2.2.2 Disadvantages of AI-driven PEM . . . 50
6.3 Broader implications . . . 51
6.4 Limitations . . . 52
6.4.1 Threats to Validity . . . 53

6.5 Future Work 53

7 Conclusion **55**

Bibliography **57**

A Appendix **I**
A.1 Survey Questions I

List of Figures

2.1 Taxonomy of player behavioural modelling by means of indirect measurements of the human player (i.e., utilising actual in-game observations to generate player models) [4]. 8

3.1 Central themes in successful serious games [26, 36]. 14
3.2 Incorporated AI in serious games for software engineering [15] 15

4.1 DSMR Process Model 19
4.2 Screenshot from Elara gameplay [10] 21

5.1 Frustration and Boredom Response for the AI Game Version and Original Game Version 35
5.2 Engagement Ratings for the AI Game Version and Original Game Version 36

List of Tables

5.1 Participant demographics. GE = Gaming Experience, PE = Programming Experience (1–5 scale). 32
5.2 Thematic Grouping of Codes from Qualitative Analysis 38

Acronyms

- AI: Artificial Intelligence
- DSR: Design Science Research
- ES: Educational Simulations
- GBL: Game-Based Learning
- IDN: Interactive Digital Narrative
- IP: Introductory Programming
- NLP: Natural Language Processing
- NPC: Non-Playing Character
- PEM: Player Experience Modelling
- PX: Player Experience
- SE: Software Engineering
- SEG: Serious Educational Game
- SG: Serious Game
- SIDN: Serious Interactive Digital Narrative

1

Introduction

Gaming is a growing field worldwide and introduces an area for interdisciplinary academic research. The gaming market has been rapidly expanding beyond simple entertainment and has become a promising addition to the education sector. Game-Based Learning (GBL) is a term describing approaches using game-components as a method for learning. There are two common GBL approaches: Serious Games (SG), and gamification [15]. Serious games have gained increasing attention as a supplement to teaching and learning materials, and its application shows significant promise, particularly in the field of Software Engineering [8]. However, in a study performed by Kharbouch et al. [15], they emphasize that serious games in software engineering are still at a very early stage. They also state that AI & Adaptability is one of the central themes of a successful SG, but despite its potential it is rarely incorporated into SGs in software engineering, and among the games that reported the use of AI none of the studied SGs in software engineering tackled Player Experience Modeling (PEM) - which is one of the three key areas of AI in educational games [36].

Building on this foundation, this project aims to explore the untapped potential of AI in enhancing serious games for software engineering education, with a specific focus on PEM in games for programming.

Through this project, we seek to address this gap by investigating how the use of AI-enhanced PEM can be effectively utilized to improve engagement, learning retention, and skill acquisition in software engineering. For this purpose we will use the educational programming game Elara [10], where we will implement a PEM using adaptive difficulty and dynamic guidance.

1.0.1 Player Experience Modeling (PEM)

PEM captures player characteristics to create an adaptive and personalized gaming experiences for the player, by utilizing metrics like behavioural data or input from speech, visual tracking, and physiological signals [36]. PEM can be implemented through data-driven or theory-driven approaches, often combined for greater effectiveness [22]. Studies show PEM can optimize difficulty, track engagement, and even predict player emotions to enhance learning in serious games [24].

1.1 Problem Description

AI has been slowly incorporated into the game industry, yet its application in educational games remains an area with significant potential for further research [15, 38, 27, 22]. Within the context of SGs for software engineering, there is a noticeable lack of studies exploring the use of AI, particularly in the domain of PEM [15]. While AI holds significant potential to enhance the educational value and engagement of SGs, it remains unclear how AI-driven PEM can specifically improve learning outcomes in software engineering education and training. There is also limited understanding of which AI techniques - such as adaptive difficulty adjustment, intelligent tutoring systems, or personalized feedback mechanisms - yield the most effective results in fostering skill acquisition, engagement, and overall player satisfaction. This lack of clarity underscores the need for more focused research to explore the role of AI in optimizing the player's experience and its broader impact on software engineering education.

There have been studies highlighting the benefits of SGs and the application of AI in games more broadly. However, there is limited research investigating the actual impact of incorporating AI-driven PEM into educational software engineering games. This study aims to address this gap by examining the effectiveness of AI-driven PEM in fostering engagement and supporting personalized learning of software engineering learners. More specifically, the study will explore this topic with the use of a SG developed for programming education called Elara [10]. The lack of focused research in this area underscores the need to explore how AI can optimize player experiences and advance the role of SGs in software engineering education.

1.1.1 Scope

This research will focus on using a programming game as a testbed to evaluate the effectiveness of AI-driven mechanisms in software engineering games. As programming games are central to software engineering education, the findings will specifically contribute to the understanding of how AI can enhance learning within this subfield. The integration of AI into a programming game will provide valuable insights into how such games can be optimized for teaching complex coding concepts, problem-solving, and software development skills. The study is limited to the use of Elara as the primary game-based learning tool. The study will investigate the effects of PEM by using adaptive difficulty adjustment and personalized timed guidance on players' learning experiences. The research will also assess the impact of AI in serious games vs non-AI serious games, aiming to uncover the specific benefits of AI-driven PEM in the context of software engineering education compared to traditional game models.

1.1.2 Research questions

To guide the research toward developing actionable insights and strategies for leveraging AI-driven PEM to optimize the educational potential of serious games in SE, the following research questions have been identified:

1. How does AI-driven player experience modelling improve engagement and skill acquisition in serious games for SE?
2. What are the advantages and disadvantages of AI-driven PEM in serious games for SE compared to traditional game models that do not incorporate AI?

1.2 Purpose of the Study

The study seeks to fill the gap in existing literature by examining the actual impact of incorporating AI-driven PEM into educational games, as prior research has largely focused on the benefits of serious games or the general application of AI in gaming, without addressing their intersection in the software engineering domain. The findings of this study are anticipated to benefit educators, instructional designers, and game developers by providing actionable insights into the design and implementation of adaptive, AI-enhanced SGs. These results may be used to create more effective and engaging educational tools that support personalized learning and improve outcomes in software engineering training. Furthermore, this research aims to contribute to the broader understanding of how AI can revolutionize the educational potential of SGs, paving the way for future advancements in GBL.

1.3 Significance of the Study

This study will make a significant contribution to the field of SGs in software engineering education, by exploring the role of AI-driven PEM in enhancing gameplay. While existing research has shown the potential of SGs to improve learning across various educational contexts, the application of AI to personalize and adapt the gaming experience in software engineering remains under-explored. By investigating how AI-enhanced gameplay influences engagement and learning outcomes, this research aims to bridge the gap between AI-driven PEM and SGs in the context of programming education. While prior studies have examined the benefits of SGs and the role of AI in gaming, limited research has focused on their intersection, particularly in the domain of software engineering education. The key contributions of this study are:

1. Results of incorporating PEM in Software Engineering Education and Programming Games:

   Investigating how personalized gameplay like adaptive difficulty and selective scaffolding impact player engagement and student learning outcomes.

2. Comparative Evaluation of AI-Enhanced and Non-AI Versions of a Programming Game:

Conducting a comparison between two versions of the game - one incorporating AI-driven PEM and one without - to assess differences in player engagement, perceived difficulty, and learning experiences.

2

Background

Games have been used for educational purposes in different ways throughout the years. They have become a valuable teaching tool, as they allow the use of interactive, engaging and immersive activities to provide a compelling context. As an example, simulation games have allowed for very successful training within the military and aviation industries, making it possible to recreate scenarios that can be either dangerous or costly to produce in real life. These simulations have been effective for allowing users to practise and improve their psychomotor skills, or how to react in specific situations. For instance, flight simulators can help pilots practice things like emergency procedures and navigation in a risk-free and cheaper environment before operating a real aircraft [12, 28, 8].

2.1 Serious Games and Related Concepts

Educational games is a term referring to a software application which utilises games technologies such as game playing and storytelling for an educational purpose [33]. These games are meant to use interactive and appealing activities to capture player's interest in learning, but also to keep them motivated during the game-play. There are many different approaches to this. They can for example reinforce the player's motivation using the sense of challenge, encouraging players to persevere, think critically, and apply problem-solving skills to overcome obstacles. Another approach is to present achievable goals and immediate results to increase the player's self-confidence and promote their decision making skills [20]. When players receive instant responses to their actions—whether through rewards, scores, or progress indicators—they experience a sense of accomplishment, which boosts their self-confidence. This reinforcement encourages players to take initiative, make informed decisions, and improve their skills in a structured yet engaging environment. Research supports the benefits of educational games in learning. A study by Lamb et al. [16] analysed and compared different educational games with each other, as well as to traditional instruction. The conclusion was that the use of educational games in the classroom is advantageous. Reviews of learning outcomes associated with these forms of educational gaming further corroborate this finding. Karkaj [22] supports this claim while emphasising the significance of narrative games as an effective structure for learning and information gathering across multiple application levels.

The field of educational games employs a variety of terms, often used interchangeably, which can lead to ambiguity [16, 15]. Lamb et al. [16] introduce the broad term *Educational Gaming* (similar to educational games) to encompass this area. Within this domain there are multiple sub-groups describing different implementations of educational games. The concept of *Educational Simulations (ES)* is defined by Lamb et al. as electronic representations of real phenomena that serve as practice for real-life tasks (flight simulators etc.). Tang et al. [33] offer a related definition of *simulations* as software applications designed to replicate the behaviour of a real-world object, machine, or system by utilizing precise mathematical models and highly accurate visual representations of its state over time. *Serious Games (SG)* are described by Lamb et al. as games intended to teach a specific set of tasks using real-world examples, potentially including board games. *Serious Educational Games (SEG)*, according to Lamb et al., are similar to serious games but are specifically designed with predefined pedagogical approaches to not only train tasks but also to teach content, as highlighted by Annetta et al. [2]. Tang et al. describes serious games as a synonym to the term educational games, explaining the definition as digital games for non-entertainment purposes. They also state that serious games may also extend to the inclusion of simulation technology, and their applicability across all age groups, contrasting with educational games which they suggest are more child-focused. Kharbouch et al. [15] defines the term serious games as a game built for a specific purpose other than entertainment, suggesting that SGs still utilize typical game mechanics, features, and elements (e.g., rules, challenges, rewards, narrative, visuals, etc.) just like entertainment-focused games. Kharbouch et al. also states that the terms *Serious games* and *Gamification* are used interchangeably when actually referring to the term *Game-based Learning (GBL)*, specifically targeting adults. According to them, GBL is a broad term covering all approaches using game components for educational purposes. Digital GBL is also defined by Tang et al. as the application of educational games within formal learning settings to foster experiential learning.

The concepts discussed above, along with the broadly defined term *educational games*, continue to be used interchangeably in the literature we have reviewed, which can complicate the analysis and comparison of different approaches. The term used most commonly in connection to the software engineering domain has been the term *serious games (SG)*. To ensure clarity throughout this report, we will adhere to the definition of "serious game" offered by Kharbouch et al. with a specific focus on education: a game specifically designed with educational goals as its primary purpose, and which integrates typical game mechanics, features, and elements.

2.2 Personalisation in games

Personalised games are defined by Bakkes et al. [3] as games that use player models in order to tailor the game experience to each individual player. Personalisation within computer games is a fairly new domain. This can be attributed to both the fact that the reason for playing a computer game can vary between such as challenge, exploration, aesthetic experience, social activity or perhaps education. Due

to these different goals of the player, it is also harder to define the player's needs and preferences that the games should adapt to [38].
Apart from this reason computer games are also generally a lot more complex than other digital applications such as websites, both in terms of the content itself and user interaction. As an example, the game-play itself is most commonly multisensory, meaning it includes elements that affects more than one sense such as visual, auditory, and tactile elements. Due to these factors, personalized computer games requires technological advancement as well as new design principles than what is currently available from the classic domains of personalisation. This also allows for this field within computer games to offer a promising domain for further research of personalisation technology [38]. Bakkes et al. [3] underscore several benefits of personalisation in computer games, including the psychological foundation, player satisfaction, contribution to game development, and the requirement for achieving ambitions. This psychological foundation directly aligns with appraisal theory [31], which posits that individuals' emotional responses to situations are not solely determined by the events themselves, but rather by their subjective interpretations or "appraisals" of those events. Personalisation in games therefore, seeks to adapt the game experience to resonate with these unique appraisals. By tailoring content, challenges, and even sensory elements, personalized games aim to evoke more positive emotional responses and enhance engagement by aligning with a player's individual interpretation of the game situation. For instance, a player who appraises a difficult challenge as frustrating might benefit from dynamically adjusted difficulty, while another who appraises it as engaging might have the challenge intensified. Thus, appraisal theory provides a strong theoretical rationale for the potential of personalisation to create more meaningful and enjoyable game experiences.

2.2.1 Player Modeling

When it comes to the application of personalisation within games, the key is the player modeling. Player modeling enables the gaming system to assess the individual player to create a personal experience. The goal is to guide the game toward consistently high player satisfaction using the player model as a foundation. This is possible by establishing models of the player. A player model is typically an abstracted description of the player, or the player's behaviour in the game [3].

2.2.1.1 Player Behavioural Modeling

The addition of AI in games can be used together with different goals. The addition of AI have been used in for example games of chess and Go, with the purpose of defeating the human player. When it comes to the use of AI in commercial games however, the purpose is to entertain the human player. As such, player behavioural modeling becomes increasingly important [3].

Bakkes et al. [4] states that in the domain of video games, there are four different approaches applicable to player behavioural modelling:

- Modelling actions

- Modelling tactics
- Modelling strategies
- Profiling a player

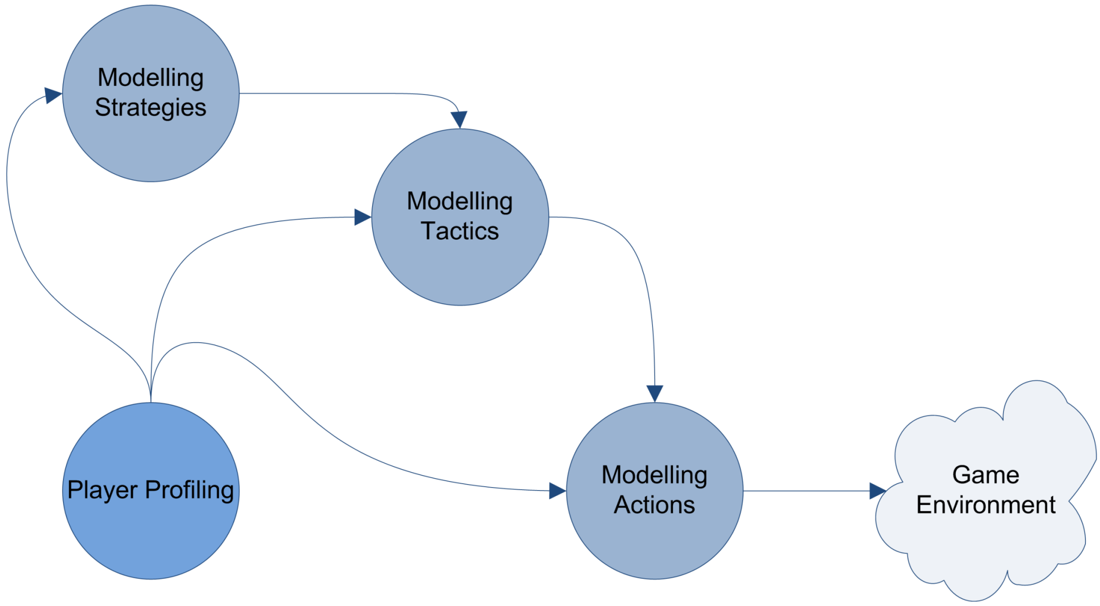

Figure 2.1: Taxonomy of player behavioural modelling by means of indirect measurements of the human player (i.e., utilising actual in-game observations to generate player models) [4].

According to this taxonomy [4], *action models* concern game actions that can either be directly observed or inferred from other observations. *Tactical models* concern local- or short-term game behaviour, meanwhile *strategic models* concern global- or long-term game behaviour. The difference between these two can be described as strategic models can span over the entire game or even across several game iterations, while tactical models only span a series of game actions. *Player profiling* recognizes that a player's choices — such as actions, tactics, and strategies is motivated by the psychological profile of the player. Combining psychological knowledge together with knowledge on player experience and satisfaction, player profiling can have a big impact on the player experience. The taxonomy is illustrated in Figure 2.1, and it is also noted that the defined classes are not mutually exclusive.

2.2.1.2 Player Experience

In a paper written in 2010 by Pedersen et al. [24], they state that there are numerous different theories as to which aspects contribute to different types of player experience such as entertainment, challenge, immersion, and frustration. Despite these multiple theories, there is substantial agreement on some points, like highlighting the significance of challenge and learnability in enhancing players' enjoyment of a game. They also state that none of these theories are quantitative - meaning that they may provide some guidance for general game design but do not offer mathematical models of player experience or address specific aspects of specific games. As there are so many different theories available, the systematic literature review by Revi et al. [27] from 2020 provides a clear overview and compilation of what dimensions that are of interest specifically in the realm of games and Interactive Digital Narrative (IDN). What they found was that player (user) experience can be divided into 47 dimensions, which could serve as useful feedback items covering eight categories: Agency, Affect, Cognition, Dissonance, Drama, Immersion, Motivation and Rewards. *Agency* and *Immersion* are identified as some of the most important player experience targets [29, 22].

2.2.1.3 Player Experience Modeling (PEM)

PEM is a method that allows capturing individual characteristics and traits of the player. It can be derived from various types of player data, including the player's behavioural and performance metrics (ex. speed, score and decisions), as well as various player-related input channels including speech data (ex text or intonation), visual data (ex the tracking of gestures, gaze and body movement), or physiological signals (ex EEG, respiration, pulse) [36]. However, capturing physiological signals has proven challenging, as it requires specialized hardware and sensors that are often intrusive and impractical for long-term use. PEM can be divided into two different types of approaches: Data-driven PEM and Theory-driven PEM. *Data-driven* PEM gather data through direct processes, in the form of sensory- or physiological readings such as heart rate, brain activity or visual or auditory stimuli. Data-driven PEM can also be gathered through indirect processes, by observing and analysing how players interact with the game during gameplay. *Theory-based* PEM on the other hand applies the principles of a theory to their modelling. This can include such as adjusting the challenge level in the game or other theories. As both these approaches have their own functionalities they can benefit in different ways, but hybrid approaches using a combination of them are also often used and may yield an even more comprehensive design [22].

When it comes to the implementation of PEM, there are many ways to realize this concept. In [22] they use telemetric data using player interactions and idleness in the game. This can entail things such as; time spent reading a note, amount of times revisiting a specific object, the number of different objects interacted with etc. The applied player modelling approach for this instance is a combination of both data and theory driven methods. Another use of PEM by [22] is in the application of dynamic difficulty. This is based off on previous player performance, and questions and answers with different difficulty levels. Pedersen et al. [24] writes a study on player experience within the context of a platform game, in which they make use of data-driven computational intelligence techniques to build quantitative PEMs, specifically for a platform game. Their models were able to accurately predict some key emotional states of the player, using a combination of gameplay metrics related to the player actions in the game, as well as parameters of the specific level in the game that was played.

2.3 Programming in educational games

Although students today have grown up in a digital world and are using computers on a daily basis, many students find courses like computer programming to be particularly challenging [20]. Even as programming is a skill in high-demand on the job market, Introductory Programming (IP) courses are reporting high failure and dropout rates [21].

In a study conducted by Mathew et al. [21], they introduced an educational game called PROSOLVE in an IP course, with the goal of enhancing the problem solving skills of the students. Using a survey and interviews, they found that both students and instructors appreciated the addition of the game in the course. The results indicate that the game helped most students enhance their problem-solving skills and better understand programming concepts and structures. Additionally, the game supported students' cognitive engagement, gains, and affective engagement in learning programming. The instructors appreciated the game as a supporting teaching tool, and considered it a good means of attracting students' interest in the programming domain. Malliarakis et al. [20] conducted a review exploring the functionalities required in educational games and how well they are met in educational programming games. The study highlights the value of educational games in teaching computer programming, and presents interesting principles that can explain why educational games can improve teaching and learning of computer programming. They explain that games can facilitate the learning-process during the teaching of specific topics or concepts within programming. In a computer programming course the extent to which games are used can vary, but can be beneficial to increase motivation in students [20]. When it comes to the educational value, they derive that the games provide students with:

- Clear educational goals and learning outputs.
- An immersive environment encouraging students' attention, participation and motivation.
- Interesting scenarios with comprehensive problems.
- Tools that help them communicate and collaborate with their classmates.

As such, evaluations showed that the majority of the users of the games expressed a positive attitude towards the examined environments [20], and the addition of educational games for programming shows good potential as a supporting teaching tool in IP courses.

2.3.1 Educational Goals

The original Taxonomy of Educational Objectives, more commonly known as Bloom's taxonomy [23], presents a classification system used to define and distinguish different levels of the cognitive domain of learning. The taxonomy is widely used in education to structure curriculum, assessments, and teaching methods. The taxonomy outlines six main categories: knowledge, comprehension, application, analysis, synthesis, and evaluation.

In the study by Malliarakis et al. [20], they explored frameworks for educational games. They found that all relevant frameworks included similar requirements as important concepts, but one suggested by Becker [6] encompassed these concepts, and was therefore used as the basis for the educational goals proposed by Malliarakis et al. The educational goals are presented along two axes: the Cognitive axis and the Emotional axis.

Cognitive axis. Educational goals should ensure that information acquired by the students begin from the first category in the Bloom's taxonomy (Knowledge) and concludes in the final category (Evaluation).

Emotional axis. Educational goals should help students navigate specific situations by appealing to their emotions. An example is given by Malliarakis et al. where the desire to free a prisoner in a game setting can be used as motivation to solve an assigned task.

Additionally they discovered that existing educational games specifically for computer programming covered both the cognitive and emotional axes, and the educational goals are explicitly tied to teaching specific concepts, in this case within computer programming.

2.3.2 Features and Limitations

Malliarakis et al. [20] also reviewed existing educational games specifically for computer programming, as well as their limitations. They state that the problems set to solve by the students are consistent with the set educational goals and their cognitive limitations. The games are also found to have students execute and complete quests, teaching them information relevant to the programming concepts set in the educational goals. All games reviewed in the study are based on a scenario with the purpose of motivating and engaging students. Most games also present some form of explanatory messages designed to assist students in understanding and navigating the game, as well as the learning material.

In the study it is noted that none of the reviewed programming games provides *relevant cases* that are used to prepare the user for the required activities. Relevant cases are defined by Malliarakis et al. as a set of pre-solved similar cases, which provide the student with a better insight of the knowledge and skills requirements of the game. They also state that the inclusion of relevant cases can increase the quality of the game, as it is a good way of providing useful tutorials and guidelines that can aid the students.
Another identified limitation, in a course setting, is the limited control the teacher has over how the game is set up or customized according to the learning objectives. The authors also suggests an improved collaboration concept, which can enable

the teaching of more complex programming concepts through team-based learning activities.

3

Related Work

With the big range of commercial games available today, player's have the possibility to chose a game that appeals to them and suits their personal interests. In contrast to this users of educational games typically only have one specific game available to them, regardless if this game fits their learning needs or their gaming-preferences. A solution to this issue can be found in the idea of creating adaptive and personalised educational games [19].

A study by Ravyse et al. [26] identified common characteristics or factors across successful SGs that positively influence learning experiences. In Figure 3.1 we can view the five central themes that emerged from the analysis: Back story and production, Realism, Feedback & Debriefing, AI & Adaptability and Interaction. These factors were deemed crucial for making the games engaging and educationally impactful. AI and Adaptability broadly refer to the game's ability to respond dynamically to player actions in an unscripted manner. This impacts SGs in two key ways: first through in-game adjustments by intelligent agents, and secondly through game-level adjustments achieved through adaptive mechanisms. In an article by Westera et al. [36], they provide a comprehensive overview of the state of AI for SGs. They introduce three key areas of AI in educational games, as can be seen as an extension to the central themes in Figure 3.1: PEM, Natural Language Processing (NLP), and advanced Non-Playing Character (NPC) modelling. The NPC is stated to envision a teaching agent that actively explores and evaluates the learner's cognitive and mental states (PEM) and can also use NLP to engage in a constructive and interactive dialogue with the player. In this project we will focus on adaptation and assessment, which includes factors like behavioural and performance data to adapt features like game difficulty. Westera et al. also finds that AI in games has significantly improved during recent years, but most applications are used to control NPC behaviours. They also state that there have been new opportunities within natural dialogues with NPC with the advances in NLP, which can also be used to facilitate interactive storytelling. Most of these applications of AI in games are specifically developed for a specific game, and the application within the area of serious games have been limited.

Zhu et al. [38] states that computer games provide an ideal research domain for exploring personalised digital applications. One of the main arguments of the paper is that the existing AI research on personalised games is mostly focused on sophis-

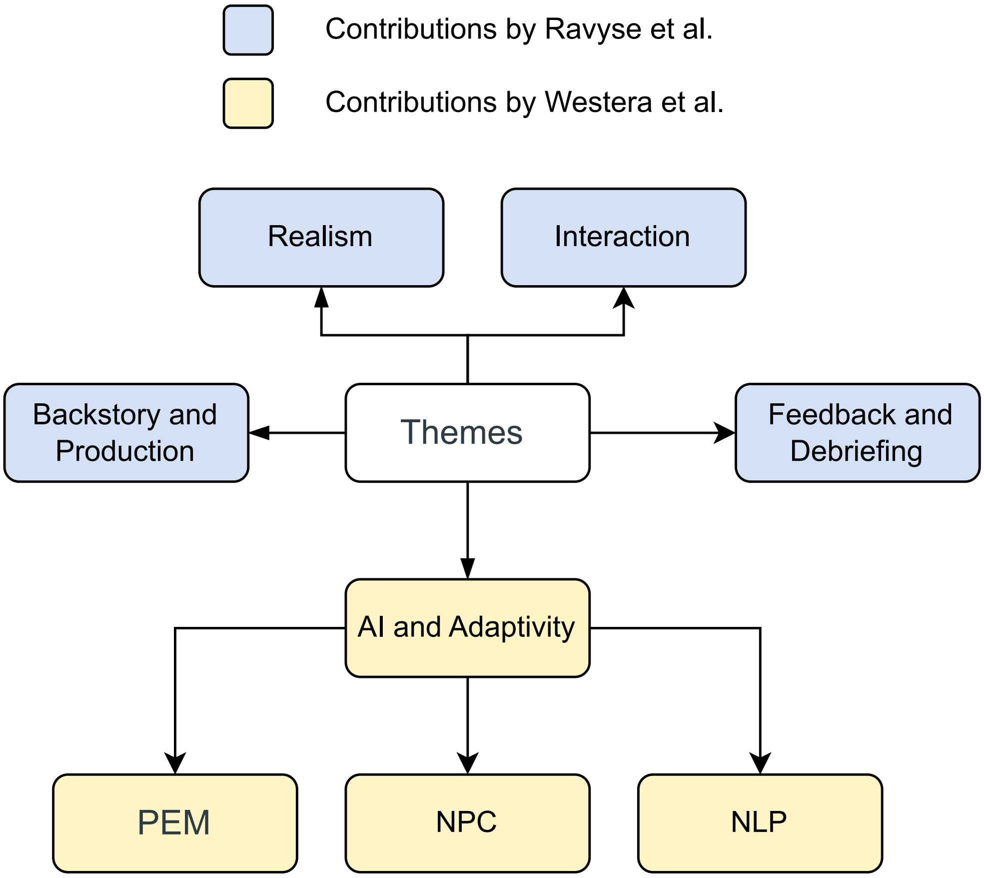

Figure 3.1: Central themes in successful serious games [26, 36].

ticated algorithms and system capabilities, however they can benefit from focusing on more player-centered approaches. This can be done by aligning the technical research with deep models of player needs and behavior, which is exactly the focus of player experience modelling. Revi et al. [27] mention that further investigations into Artificial Intelligence (AI) and NLP techniques can enable the automation of estimating feedback items of player experience, which in turn may facilitate the process of identifying problems as well as the tailoring of player experience, and allowing for more complex narratives. There is also a lack of research identified on combining IDN and SG design from an AI perspective [22].

A comprehensive mapping study specifically focusing on serious games in the realm of software engineering, is introduced by Kharbouch et al. [15]. They state that AI is regarded as a crucial element of game technology, and is often considered the core of SGs. However, despite its potential the study revealed that there is a serious lack of AI in SGs for software engineering. in Figure 3.2, we can see that only 17.56% of SGs in software engineering incorporate AI into their gameplay. We can also see that among the games that reported the use of AI - 76.92% implemented

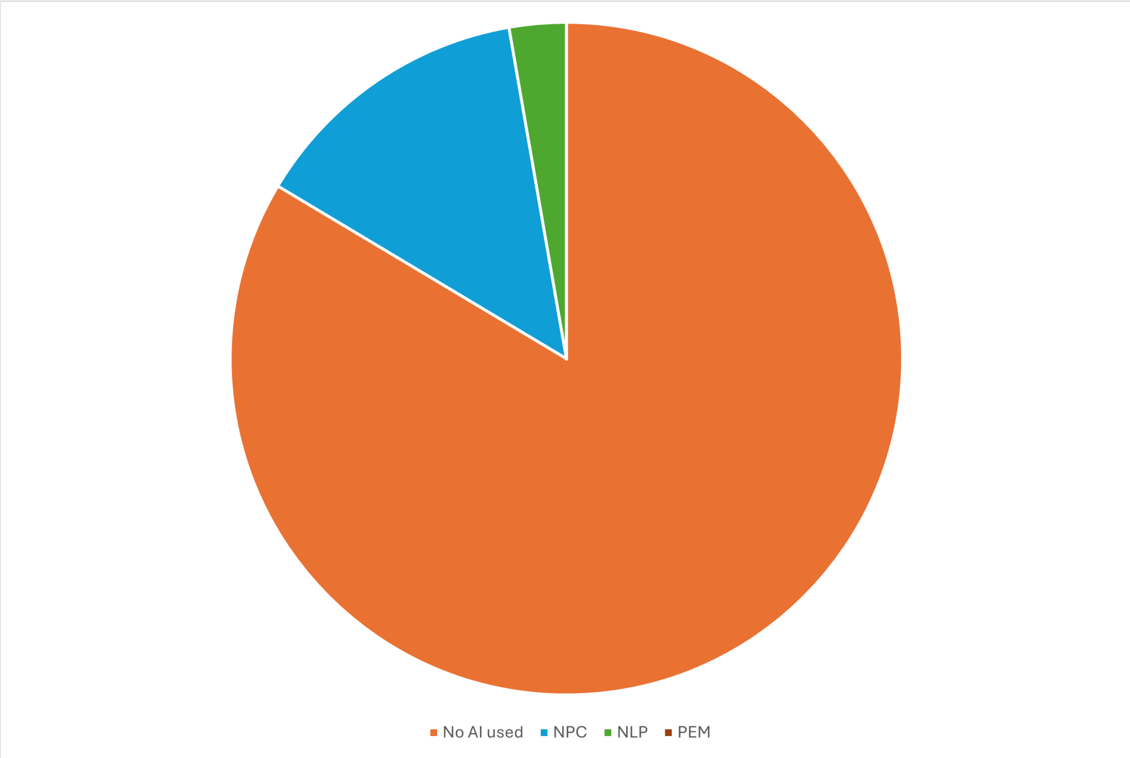

Figure 3.2: Incorporated AI in serious games for software engineering [15]

NPC, 15.38% supported NLP and none of the studied SGs in software engineering tackled PEM. Despite the potential, there are also apparent challenges with AI in serious games to keep in mind. One of these challenges is to assure that the game AI supports educational objectives [14], meaning the artificial intelligence within the game shouldn't just make the game fun or challenging in a generic way. Instead, the AI needs to be designed and programmed to actively contribute to the learning goals of the game. Another challenge can be found in incorporating other educational AI functions in order to maximize learning, while still ensuring it doesn't detract from game design principles [14]. While incorporating AI features that directly support learning is essential, it's equally important not to let these features detract from the core gameplay experience. A game that is overly didactic, boring, or frustrating will likely fail to engage learners, thus undermining its educational potential.

The study by Kharbouch et al. also mentions some game AI concepts and technologies that show high potential such as: Promoting Instructive Gameplay, Managing the Level of Challenge, Providing Scaffolding Selectively and Supporting Reflection and Skill Improvement [15]. AI can be used to promote instructive gameplay by integrating adaptive teaching elements, ensuring that educational content is engaging and effective. It can also be used to manage the level of challenge by dynamically adjusting difficulty based on player performance, keeping the game balanced between engagement and frustration. Additionally, AI may provide scaffolding selectively, offering support only when needed to encourage independent learning. Lastly, it can support reflection and skill improvement by analysing player performance and offering personalized feedback to foster continuous learning. In this project we will focus on incorporating Managing the Level of Challenge and Providing Scaffolding Selectively.

3.1 Player Experience Modelling

In the domain of serious games, ensuring an engaging and educationally effective experience requires an understanding of how players interact with and respond to game mechanics. PEM plays a crucial role in this by leveraging data-driven approaches to assess and adapt the gaming experience to individual players. By analysing various aspects of player behaviour - such as decision-making patterns, performance metrics, and physiological responses - PEM enables games to dynamically adjust elements like difficulty, feedback mechanisms, and narrative progression to better suit the learner's needs. For the purposes of this study, we developed a tailored PEM model, referred to as *DualTypePEM*, to evaluate and compare two versions of the educational programming game Elara: one incorporating the PEM model and one without. This model emphasizes two central areas: feedback & debriefing and game difficulty balancing. These critical areas are aligned with the AI-driven strategies *Managing the Level of Challenge* and *Providing Scaffolding Selectively*, as mentioned by Kharbouch et al. [15]. Feedback & debriefing shape how players interpret and respond to in-game interactions, and game difficulty balancing ensures an optimal challenge level to sustain engagement and learning. These elements collectively contribute to a more adaptive and effective serious gaming experience, aligning educational objectives with player-centered design principles.

3.1.1 Feedback & Debriefing

The feedback and debriefing theme focuses on understanding the communication players receive during and after their gameplay experience. This includes both the in-game feedback provided by the game mechanics (such as scoring, hints, or guidance) and external feedback from the environment, such as from instructors or peers. The theme aims to provide a deeper understanding of how this feedback influences the player's learning experience [26]. A critical component in facilitating effective feedback is how an intelligent educational system, such as a serious game (SG), responds to a player's need for help or guidance. This requires an understanding of the specific solution path that the player is pursuing as well as the player's current level of knowledge on the topic, ensuring that guidance is both relevant and timely [11]. To address these needs, Gertner et al. [11] used a probabilistic student model in their intelligent tutoring system to deal with help request from the users. The model used a Bayesian network assessing the user's general knowledge of the domain, the specific knowledge of the problem, as well as the potential strategies the user may be exploring to solve the problem.

3.1.2 Game difficulty balancing

Game difficulty balancing refers to the process of adjusting the level of challenge in a game to ensure it is neither too easy nor too difficult for the player. This balance is critical for maintaining player engagement because if a game is too hard, players may become frustrated and quit; if it's too easy, players may lose interest due to lack of challenge. Dynamic game difficulty balancing can improve player motivation by providing just the right amount of challenge to keep players interested and motivated to progress. Additionally, this balancing act contributes to an improved overall gameplay experience by ensuring that players feel accomplished when overcoming challenges without feeling overwhelmed. Essentially, it makes the game enjoyable, rewarding, and satisfying, keeping players invested in the game over time. The adaptation mechanism should involve an advanced, self-adjusting optimization algorithm that continually fine-tunes both task difficulty and skill mastery [36].

Learner motivation is a key factor in determining how well someone learns [20], and by maintaining the right balance of difficulty, a game can keep players engaged and invested in their progress. This ultimately helps to enhance and preserve motivation, which is essential for effective learning. As such, dynamic game difficulty balancing ensures that players stay challenged without feeling overwhelmed, leading to a more rewarding and effective learning experience [36].

3.2 Demographics

Serious games have shown promising results in studies, and the majority of student's express a positive reaction to addition of serious games in education. It is however important to keep in mind the different types of player prerequisites or demographics in the design of educational games [30]. Studies of media-use in children [7] have shown that video- and computer games are used more by older children, and used more by boys then girls. The gender difference is also often even more evident with higher age. When taking into account the full scenario of an educational game in a classroom setting, together with the individual player's gender identity and gaming orientation, all these factors contributes to the risk of making classroom educational games a gender-polarizing experience [13]. Heeter et al. [13] points out that many researchers who highlights the benefits of using game based learning, neglect to address these different kinds of players and their demographics in their research.

3.2.1 Player Types

A player's personal goals and motivations within a game can contribute to the categorization of player types. These player types can provide a foundation for understanding and assessing how players differ from each other, and how their motivations may be related to factors such as gender, age, usage patterns, and in-game behaviours. Bartle [5] found two dimensions of playing style, and from this derived four distinctive player types identified as: achievers, explorers, socialisers and killers. Yee [37] found 10 different player motivations, which could be grouped into three overarching types: achievement, social, and immersion.
According to Magerko et al. [19] Bartle's and Yee's achievers/achievement types all likely refer to gameplay driven by extrinsic motivation. They also highlight the links between the explorers and immersion types, likely depicting gameplay driven

by intrinsic motivation. There is also a suggestion that a player's motivation behind playing a game may be related to how a player learn.

Building on these foundations, our study introduces the DualTypePEM model, which simplifies the player type framework into two primary motivational categories: achievers and explorers. Achievers are typically goal-oriented, responding well to structured challenges, measurable progress, and rewards. Explorers, in contrast, are driven by curiosity and tend to enjoy open-ended problem solving and discovery. This binary classification informed the design of adaptive features in the AI-enhanced version of the programming game, guiding how difficulty was balanced and how feedback was personalized.

4

Method

This study explores the impact of AI-driven PEM in SG for Software Engineering (SE) by adopting the solution seeking methodology of Design Science Research proposed by Peffers et al. [25] as its guiding framework. Design Science Research (DSR) is particularly well-suited for this study as it provides a structured approach for developing and evaluating novel artifacts – in this case, the *DualTypePEM* model which is integrated into the SG Elara [10]. This chapter outlines the six key steps of the DSR process and details how each step was applied within the context of this research.

4.1 Design Science Research (DSR) Framework

The DSR process model [25] is shown in Figure 4.1, and consists of the following six iterative steps:

1. Problem identification and motivation
2. Define the objectives for a solution
3. Design and development
4. Demonstration
5. Evaluation
6. Communication

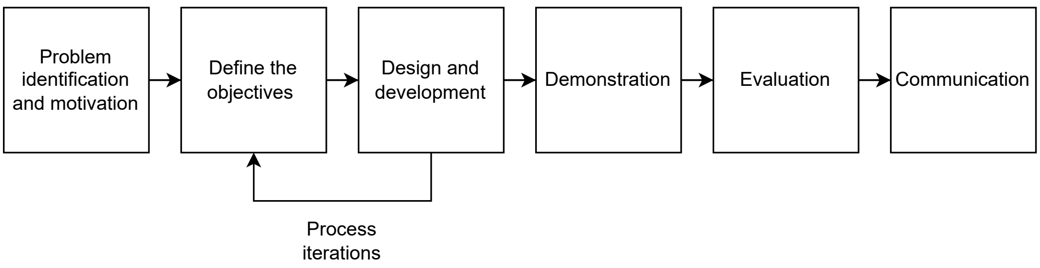

Figure 4.1: DSMR Process Model

As these steps are iterative, some have been revisited during this process. The following sections detail how each of these steps have been enacted in this study, as of the final iteration of this process.

4.1.1 Activity 1: Problem identification and motivation.

The first activity is to define the research problem and justify the value of a solution. The problem description and motivation for this research is outlined in the introduction and can be found in section 1.1-1.3.

4.1.2 Activity 2: Define the objectives for a solution.

Based on the identified challenges and motivations outlined earlier, the DualTypePEM is designed to address key issues in serious games for software engineering education. The following objectives guide the development of the DualTypePEM, focusing on creating a more personalized, adaptive, and engaging learning environment. These objectives aim to improve not only measurable outcomes such as skill acquisition but also the qualitative aspects of player experience, ensuring that the game effectively supports diverse learners and fosters sustained motivation.

1. Personalize gameplay by dynamically adjusting difficulty levels based on ongoing player performance and profile, in order to maintain challenge without inducing boredom or frustration.

2. Support different player types (e.g., achievers vs. explorers) by offering flexible adaptation paths.

3. Deliver in-game, dynamically timed feedback that replaces or supplements static instructional materials, thereby promoting trial-and-error learning and reducing reliance on external guidance (e.g., tutorial videos or text).

4. Enhance the emotional quality of the learning experience by maintaining a balance between challenge and support, reducing the likelihood of cognitive overload or disengagement.

4.1.3 Activity 3: Design and development.

This phase involves creating the research artifact - in this case our AI-driven PEM model *DualTypePEM* and its integration into Elara. The process includes defining the artifact's functionality, outlining its structure, and building it. Successful development draws on relevant theoretical knowledge to translate objectives into a functional solution.

4.1.3.1 Elara

Elara [10] is an SG with the purpose of teaching programming. The game is developed to teach the user the basics of programming and does so using a language called Rhai [1]. Rhai is a scripting language and evaluation engine that integrates with the Rust language [34]. The purpose of Rhai is to provide a dynamic layer on top of code written in the Rust language. As a scripting language Rhai is described as small, fast and easy-to-use. As such, Rhai is considered as a language that is easy to learn, and according to the developer of Elara, the language can be considered to be somewhat similar to JavaScript. Elara have served as the primary tool to examine how AI can enhance engagement, learning retention, and skill acquisition within the context of SE education.

Elara is built using Rust [34] for most of the game logic, and Typescript [35] for a web UI responsible for rendering and user interaction. The game is accessible via Github. The game contains different elements such as story-elements, game levels where the player practise their programming skills, and a guide-book - referred to as the "journal" - where the player can read or watch some YouTube videos about different programming-related concepts that they come across during the gameplay. The game is structured in a way that it begins with a short background of the story, followed by a tutorial about how to play the game where the player is unlocking guide-book chapters to learn the theory behind the programming concepts, and levels where they apply these concepts. As the player completes one level, the next level is unlocked, as well as potential progression in the storyline. When playing the levels you are controlling a small space rover using simple programming with already constructed functions, with a position-based goal that you need to reach with the rover, as can be seen in Figure 4.2.

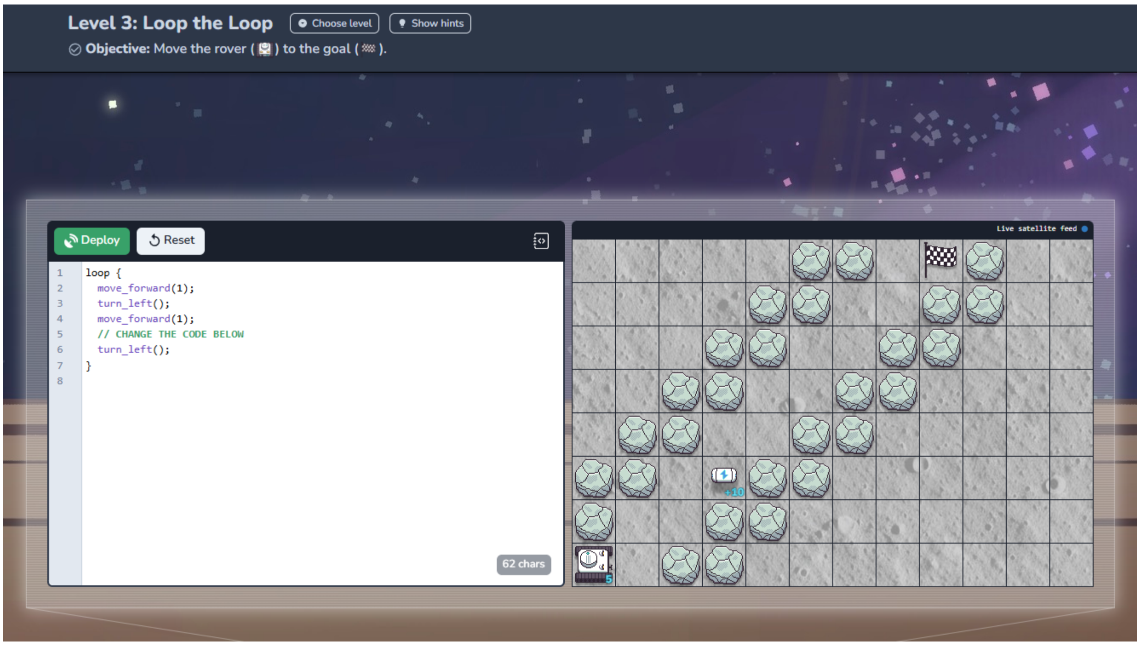

Figure 4.2: Screenshot from Elara gameplay [10]

4.1.3.2 Performance Tracking

When developing the DualTypePEM model, the initial step involved integrating player performance tracking. These metrics was be used both in the model, but also to inform the quantitative analysis in the evaluation phase. To effectively analyze player experience and inform the adaptive mechanisms of the player experience model, we selected three key performance metrics:

- The time spent solving each level
- The number of failed attempts per level
- Problem-solving accuracy: proximity to the desired solution

These metrics provide a balance of temporal, behavioural, and spatial data, offering a well-rounded view of player interaction with the game.

Time Spent Solving Each Level: This metric captures the total time a player takes to complete a level, measured from the start of the level to either a successful completion or an exit. It serves as a basic yet informative indicator of cognitive effort and player engagement. Longer times may suggest that a level is too difficult, unclear, or that the player is experimenting with strategies. Conversely, very short times may indicate either high proficiency or disengagement. This data can help identify whether the game flow is appropriately challenging and whether AI-driven adaptations result in improved pacing.

Number of Failed Attempts Per Level: Tracking the number of failed attempts gives insight into how players are learning and adapting. Each failed attempt is logged when a player restarts a level without reaching the goal. A high number of failures may suggest frustration or a poor fit between the level's difficulty and the player's current skill level. When analysed across multiple players and levels, this metric helps assess the effectiveness of difficulty balancing, particularly in comparing the AI-enhanced version of the game to the original. It also enables the DualTypePEM to determine when to intervene with support or adjust level complexity.

Problem-solving accuracy: Given the position-based nature of the game's goals, often involving intermediate state changes, we define accuracy as the spatial difference between the rover's current coordinates and the coordinates of the target goal. This continuous measure of proximity allows for a granular understanding of the player's progress and informs the adaptive mechanisms of the DualTypePEM.

To quantify this spatial difference, we employed the Euclidean distance formula, which calculates the straight-line distance between two points in a 2D space [18]. This method is well-suited to our game environment, where the player's rover operates on a grid-based plane. The formula is defined as:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

where (x_1, y_1) represents the rover's final position at the end of a level, and (x_2, y_2) denotes the target goal's coordinates. The resulting distance provides a precise and interpretable value for how close a player came to solving the level optimally.

Using Euclidean distance as a continuous performance metric enables us to:

- Track incremental improvements even when levels are not completed successfully
- Differentiate between near-misses and complete failures, offering a more nuanced view of problem-solving attempts
- Support adaptive gameplay by informing the DualTypePEM of when to increase or reduce difficulty based on player proficiency

This metric is particularly valuable in an AI-enhanced context, where subtle variations in performance can guide the system's feedback, level suggestions, or hint generation in real time.

4.1.3.3 Player types

During the development of the DualTypePEM, we considered the use of machine learning techniques for reliably modelling the player's emotional state. However, to implement a machine learning model you need a large amount of data. According to a commonly cited rule of thumb [17, 32], a simple model should be trained on at least ten times as many data points as there are features in the dataset. However, due to time constraints, we were unable to obtain the necessary volume of data for this study. Additionally, although the defined performance metrics provided useful indicators of potential frustration or engagement, they did not offer clear or consistent thresholds that could be used to trigger adaptive changes in gameplay, for example when a player becomes frustrated and might benefit from easier gameplay. To train a reliable data-driven model, a large-scale dataset would be required - including not only gameplay behaviour but also information about players' mental or emotional states. In the absence of such comprehensive data, it was not feasible to implement a machine learning-based solution. As a result, a more flexible and interpretable approach was adopted. Rather than relying on inferred emotional states through predictive modelling, the DualTypePEM model was designed to use a simplified system based on two general player types: the *Achiever* and the *Explorer*, allowing the game to adapt to individual differences in play-style and performance. These categories were inspired by established player motivation models, such as those proposed by Bartle [5] and Yee [37], and adapted to suit the specific context of this educational game.

The Achiever: This type is characterized by a desire for challenge, progression, and mastery. This type of player is typically extrinsically motivated and may thrive on difficult tasks, fast-paced problem-solving, and a sense of accomplishment. They are not as interested in the story-line and avoids excessive amounts of information.

The Explorer: As a contrast, the Explorer is motivated more by enjoyment, discovery, and steady progress. This player type is typically intrinsically motivated and may prefer a lower-difficulty experience that emphasizes comfort, narrative, or creative experimentation.

Although adopting a player-type-based model provided a more flexible alternative to machine learning, a further challenge emerged: there were still no clearly defined thresholds for automatically classifying players as either Achievers or Explorers based solely on the performance metrics. While the metrics offered insight into player behaviour, such as task completion time, number of failed attempts, and accuracy, they did not translate directly into distinct categories without contextual interpretation.

To address this, a manual classification approach was adopted during the evaluation phase. For each participant, the DualTypePEM model was explicitly assigned a player type — either Achiever or Explorer — based on the evaluator's overall impression of the player's mental state in combination with their observed performance. This decision was informed by gameplay patterns, verbal and non-verbal cues (where available), and subjective indicators of engagement or frustration. While this method lacks the objectivity of automated classification, it provided a practical way to explore how the model's adaptations affected players with differing motivations and needs. Furthermore, it laid the groundwork for refining future, more data-driven versions of the model.

4.1.3.4 Dynamic elements

Building on the player types framework, the DualTypePEM model integrates two key AI-driven educational strategies [15] to enhance the effectiveness of the game: Managing the Level of Challenge and Providing Scaffolding Selectively.

Managing the Level of Challenge: The model adjusts the game difficulty based on the inferred player type, by picking the next levels presented to the player. For example, an Achiever, who seeks challenge and mastery, was be presented with progressively harder levels to maintain engagement and motivation. Conversely, an Explorer, who prefers steady progress and enjoyment, may receive a more gradual difficulty curve that reduces frustration and emphasizes discovery. This dynamic adjustment is informed by continuous monitoring of performance metrics such as time spent on tasks, number of failed attempts, and problem-solving accuracy. This was prevent frustration from excessive difficulty or disengagement from tasks that are too easy. This adaptive approach was also help maintain motivation and encourage skill progression.

Providing Scaffolding Selectively: Instead of offering constant or predefined guidance, the DualTypePEM model implements selective scaffolding, where hints are delivered dynamically based on the player's performance and inferred player type. While both the original and AI-enhanced versions of the game include the same set of hints, only the AI-enhanced version uses the DualTypePEM to determine when and if these hints are shown. This approach allows the game to respond intelligently to moments where the player may be struggling. For example, Achievers — who are typically motivated by challenge and mastery - was receive hints more sparingly to preserve the sense of difficulty and progression. In contrast, Explorers - who value steady progress and enjoyment - may receive hints earlier or more frequently to avoid frustration and maintain engagement. The scaffolding adapts over time as the player progresses, becoming less frequent when consistent success is observed. This gradual reduction in support is designed to promote independent problem-solving and reinforce learning. By dynamically controlling the timing of support rather than its content, this approach allows for a controlled and consistent comparison while still offering a personalized gameplay experience.

4.1.4 Activity 4: Demonstration.

The demonstration of the AI-enhanced game was conducted through the "Game Play" phase of the study. This phase introduced 8 participants to interact with both the original and the AI-enhanced versions of Elara, providing a practical setting to observe the functionality of the DualTypePEM model in action. For the demonstration, gameplay was structured as follows:

First, participants were introduced to the game through an in-game story dialogue sequence, which provided narrative context and immersion. Following this, they played the first three levels of the game, which served as a tutorial to teach the core game mechanics. These introductory levels were the same across all versions and were not influenced by AI adaptations, ensuring a consistent learning baseline for all participants.

After completing the tutorial levels, each participant was asked to play four additional levels — two from the AI-enhanced version of the game and two from the original version. To counterbalance potential ordering effects and ensure fairness, participants were randomly assigned to one of two groups:

- **Group A** began with the AI-enhanced version, followed by the original.
- **Group B** began with the original version, followed by the AI-enhanced.

This ordering was carefully chosen because the first set of post-tutorial levels included slightly easier challenges, and alternating the starting condition allowed us to reduce bias in participants' initial impressions or performance. For participants who started with the AI-enhanced version, they were assigned a player type (Achiever or Explorer) based on observed behaviour and performance in the tutorial levels. For those who started with the original version, the player type was determined using insights gathered from both the tutorial levels and the gameplay in the original, non-adaptive levels.

During gameplay of the AI-enhanced version, the system dynamically responded to player behaviour using the following mechanisms:

- Dynamic adjustment of game difficulty, where the next levels presented to the player were selected based on their performance and their assigned player type.
- Selective scaffolding, in which the same hint content available in the original game was delivered in a more adaptive manner — triggered based on observed

need, as interpreted through the DualTypePEM and informed by the participant's assigned player type.

Throughout the session, gameplay data was recorded to allow for post-hoc analysis. These data provided an initial understanding of how users interacted with the AI-enhanced features and how the DualTypePEM model's adaptations shaped the learning experience. This demonstration phase served not only to validate the technical implementation of the AI features but also to observe how players with different types responded to adaptive gameplay in a controlled, comparative setting.

4.1.5 Activity 5: Evaluation.

The evaluation of the AI-enhanced game's effectiveness in achieving the research objectives was be conducted through a combination of data collected during the "Game Play" phase and the subsequent "Assessment" phase. The study includes pre- and post-assessments to evaluate participant performance and experiences. Data was be analysed using both quantitative methods (e.g., performance metrics) and qualitative methods (e.g., thematic analysis of surveys) to provide a comprehensive understanding of the effects of AI-driven adaptations on learning and engagement. The following section outlines the research design, participants, data collection methods, and analytical procedures that was be used to address the research questions.

4.1.5.1 Participants

A diverse demographic was sought to reflect realistic learner variability and to assess how individuals with different backgrounds, preferences, and characteristics engage with both static and adaptive game formats. The only requirement was that the participants should have limited prior experience in programming. A diverse group of eight participants was recruited in the study, ranged in age from 16 to 56 years old, with the majority falling within the 20–30 age group. The participants were of different experience levels in terms of both computer games and programming. Most of the participants had no prior experience in programming, with a few reporting some previous experience. This composition reflects the target audience of Elara, which is designed for beginners seeking an accessible and engaging introduction to programming and software engineering concepts.

Beyond age and experience, the participant pool also demonstrated gender diversity, with representation from both men and women. This balance contributed to the inclusivity of the study, supporting findings that are more representative and applicable to a broader spectrum of learners. Participants also exhibited different player-types, with a relatively even distribution of explorers and achievers. These varied player-types allowed the study to observe how adaptive features in the AI-enhanced version of Elara catered to different motivational profiles and engagement patterns. This diversity in age, experience, gender, and player-type adds richness to the analysis and supports a more comprehensive understanding of how adaptive serious games can accommodate a wide range of learners.

4.1.5.2 Data Collection

The study employed several surveys and logged in-game player performance metrics to gather data at different stages:

1. **Pre-assessment survey:** Assessed baseline programming knowledge, demographics, and prior experience.

2. **Player Performance Tracking:** Logged in-game performance metrics, including time spent on each level, number of failed attempts, hint usage, and solution accuracy.

3. **Game Experience Feedback surveys:** Collected participant reflections after playing each game version.

4. **Post-assessment survey:** Measured learning gains and gathered qualitative feedback comparing the two game versions.

Pre-assessment: Participants completed a pre-assessment survey prior to engaging with the game. This survey served multiple purposes: it gathered information on participants' baseline programming knowledge, prior exposure to computer games, and relevant demographic information such as age and gender. The programming section included basic coding questions designed to assess initial understanding and establish a reference point for later comparison. This data provided a comprehensive overview of each participant's starting point and helped identify variations in background that could influence their engagement with the game and their learning outcomes. By capturing these differences, the study aimed to better interpret individual trajectories during gameplay and understand how initial knowledge and perceptions might shape the effectiveness of both the adaptive and static versions of Elara.

Player Performance Tracking: Throughout the game-play session, the game was instrumented to automatically log key performance metrics during gameplay. These included time spent on each level, number of attempts per level, and problem-solving accuracy. All data were automatically recorded locally in a structured log file by the game, which was later accessed by the researcher for analysis. This enabled an objective, real-time capture of player behaviour without requiring manual input.

Game Experience Feedback surveys: After completing each version of the game, participants completed a Game Experience Feedback survey designed to capture their subjective impressions of the gameplay. These surveys employed a mix of open-ended questions and 5-point Likert scale to evaluate key aspects of user experience, including engagement, motivation, challenge appropriateness, and satisfaction with in-game guidance and feedback. These responses provided qualitative and quantitative insights into how players experienced the two game versions and complemented the objective gameplay data logged by the system.

Post-assessment: After completing the gameplay sessions, participants were asked to complete a post-assessment survey designed to evaluate changes in their knowledge, perceptions, and overall experience. The survey included a set of programming and conceptual questions that were identical to those presented in the pre-assessment. This consistency enabled a direct comparison of pre- and post-performance, offering insight into whether participants experienced any learning gains during the study. In addition to the objective assessment items, the survey included open-ended questions prompting participants to reflect on their experience with both versions of the game. They were specifically asked to compare the AI-enhanced and original versions and explain the differences that they noted, as well as which version they preferred. These qualitative responses provided valuable perspectives on how the adaptive features influenced player experience and perceived learning effectiveness.

4.1.5.3 Data Analysis

The data collected in this study was analysed using both quantitative and qualitative methods to provide a comprehensive understanding of how AI-driven PEM influences player engagement and skill acquisition. The data collected was analysed using the following procedures:

Quantitative Analysis: Descriptive statistics were used to summarize the key performance metrics as well as the pre/post-assessment programming scores. These metrics helped illustrate patterns of gameplay behaviour and learning progression across both the AI-enhanced and original versions of the game. To assess differences between the two game versions, paired sample t-tests was employed to compare within-subject differences in things like engagement, frustration and satisfaction.

Qualitative Analysis: The open-ended questions of the survey data was analysed using a thematic analysis based on the Braun and Clarke framework [9]. This framework contains six steps:

1. Data familiarization
2. Initial code generation
3. Thematic search
4. Thematic review
5. Thematic naming and definition
6. Report

This method was chosen due to its flexibility and suitability for exploratory research. It allows for identification of both common and divergent experiences across participants and can surface unanticipated insights. Additionally, it accommodates both psychological interpretations (e.g., player motivation and frustration) and social aspects (e.g., collaborative or competitive tendencies) of engagement with the AI-driven game.

Example: In the post-assessment survey, the participants were asked - "Which differences did you notice between the two games?".

During initial coding, one participant answered:

> *"In the second game (AI-version) the difficulty was one level higher, but I did not find it too difficult. It was a good difficulty level for me, and for beginners."* - Participant P1 (Table 5.1)

This segment was coded as *Perceived challenge* and *Good difficulty balance*, both of which later contributed to the broader theme labelled *Challenge & Difficulty*.

Together, the quantitative and qualitative analyses aim to triangulate findings and offer a richer understanding of how adaptive systems impact learning and motivation in serious game contexts.

4.1.6 Activity 6. Communication.

The final step of this Design Science Research Methodology involves communicating the findings of this thesis research to relevant audiences. This communication aims to contribute to the body of knowledge in the fields of serious games, artificial intelligence in education, and software engineering pedagogy, as well as to inform future research and practice. The primary channel for communication is this thesis report itself, as it details each step of the research process, from problem identification to evaluation, the design, development, and assessment of the AI-driven Player Experience Model DualTypePEM integrated into the serious game for software engineering education.

5

Results

This chapter presents the findings from the study investigating the impact of adaptive gameplay features in Elara on novice programming learners. The results are organized to provide a comprehensive overview of participant demographics, quantitative measures of learning and gameplay behaviour, subjective player feedback, qualitative insights from open-ended responses, and statistical comparisons between the AI-enhanced and original versions of the game.

5.1 Participant Demographics and Background

A total of eight participants took part in the study. The age of participants ranged from 16 to 56 years old, with the majority falling within the 20–30 age group. The participant pool included five women, three men, and no participants who identified as non-binary or preferred not to disclose their gender, ensuring some degree of gender diversity within the sample.

As can be seen in Table 5.1, regarding prior experience with digital games most participants rated their familiarity in the mid-to-high range on a five-point scale. This indicates that while participants had varied levels of exposure to gaming, most were at least moderately comfortable with interactive game environments.

In contrast, programming experience was considerably lower across the group, rated their familiarity in the low-to-mid range. This distribution aligns with the intended beginner focus of Elara and supports the study's goal of assessing how adaptive gameplay might support novices in learning foundational programming concepts.

In terms of player types, participants were evenly split between explorers (n = 4) and achievers (n = 4). This balance enabled the study to explore how adaptive and static game experiences cater to different motivational profiles and interaction styles, such as goal-driven versus curiosity-driven engagement. One participant was possibly misclassified as an achiever, which is also noted in Table 5.1.

ID	Group	Player Type	Age	Gender	GE	PE
P1	B	Achiever	25	Female	4	1
P2	B	Explorer	56	Male	4	1
P3	B	Achiever (Pos. Explorer)	28	Female	3	3
P4	A	Explorer	19	Female	2	2
P5	A	Achiever	16	Male	5	2
P6	A	Explorer	27	Female	4	1
P7	A	Explorer	52	Female	2	2
P8	B	Achiever	26	Male	1	1

Table 5.1: Participant demographics. GE = Gaming Experience, PE = Programming Experience (1–5 scale).

5.2 Quantitative Outcomes

To evaluate the effectiveness and impact of the AI-enhanced version of Elara compared to the original version, several metrics were compared across conditions, including gameplay performance and survey-based measures of user experience. This section presents the numerical data collected from the study to evaluate participant performance and gameplay experiences. It includes analysis of pre- and post-assessment coding scores to measure learning gains, objective gameplay metrics recorded during play sessions, and quantitative feedback from game experience surveys. To assess whether the AI-enhanced version of Elara significantly affected players' engagement and perception of challenge compared to the original version, paired sample t-tests were conducted on the survey ratings for these dimensions. Together, these data offer insight into how the AI-enhanced and original versions of Elara influenced participants' engagement, motivation, challenge perception, and problem-solving outcomes.

5.2.1 Pre- and Post-Assessment Coding Scores

In the pre-assessment, three out of eight participants answered both questions incorrectly, reflecting limited prior programming knowledge. Only three participants answered at least one question correctly.

In the post-assessment, performance improved: six participants answered both questions correctly, and two answered one correctly. No participant maintained both their incorrect responses. While the sample size limits statistical power, this shift suggests a positive trend in conceptual understanding following the gameplay sessions.

Out of the two participants who answered one wrong question in the post-assessment, one was classified as an achiever and one as an explorer. They were also in different groups so one was introduced to the concept of functions in the AI-enhanced game, and one in the original game version. As such, there was no substantial difference in post-assessment performance between participants' experiences with the AI-enhanced and original game versions, indicating that while overall learning occurred, the adaptive features did not produce a clearly measurable impact on short-term conceptual gains based on this assessment.

5.2.2 Gameplay Metrics

Player interactions during gameplay were automatically logged, including:

Time per level: Participants generally spent slightly more time on the AI-enhanced levels. The mean time over all levels played in the AI-enhanced game was around 4.5 minutes, while the mean for the original game was slightly below 4 minutes. This is likely due to the adaptive elements offering personalized challenges and hints.

Number of attempts per level: The AI-enhanced version also recorded a higher number of attempts per level compared to the original. The mean number of attempts for the original version were 1,9 attempts per level, while the AI version had a mean of 3,3 attempts per level across all participants. This may be attributed to the adaptive version encouraging experimentation by providing a less amount of hints, and adjustments in difficulty. Players seemed more willing to explore different approaches, which could indicate a higher level of engagement and iterative learning. In contrast, the static version may have prompted more cautious or conservative play, with fewer attempts made per level.

Problem-solving accuracy: The problem-solving accuracy - defined as how close participants came to solving the task on unsuccessful attempts - varied greatly between players. The mean problem-solving accuracy for the original game was around 0,73 (where 0 indicates position at the goal), while the mean for the original game was around 0,71. As such, there was not a substantial difference between the two versions. This suggests that while the adaptive features may have enhanced player guidance and engagement, they did not significantly change how close participants typically got to the correct solution. This suggests that while the adaptive features may have improved guidance and engagement, they did not significantly alter the end result in terms of correct solutions.

Additionally if we look at the time per level and number of attempts per level, there was also a notable variation in individual play styles. Some participants spent only a short amount of time per level but had a high number of failed attempts, frequently testing different solutions in quick succession. Others adopted a more cautious approach, spending a longer time planning their solution mentally before making a single, often more successful attempt. This variation in strategy highlights how players engaged with the game mechanics in different ways, regardless of the version being played. Interestingly, most of the participants who followed the rapid, high-attempt strategy were achiever-type players. This may reflect a goal-oriented mindset focused on reaching a correct solution quickly, even if it meant iterating through several failed attempts. In contrast, explorer-type participants tended to engage more methodically, aligning with their preference for thoughtful experimentation.

5.2.3 Game Experience Feedback Surveys

To capture participants' subjective impressions of their gameplay experience, a Game Experience Feedback survey was administered after each version of the game.

The goal of this analysis is to examine how participants rated their experience numerically and to identify patterns in how adaptive features may have influenced their perceptions. The results presented below are limited to these numerical ratings and do not yet incorporate open-ended qualitative feedback, which is addressed in a separate section.

5.2.3.1 Frustration and boredom

In response to the question "Did you ever feel frustrated or bored while playing the game?", three participants reported experiencing less frustration and boredom while playing the AI-enhanced version. Out of these participants, two were from Group B (original version first), and one were from Group A (AI-enhanced version first). They were all achiever-type players, who appeared to appreciate the additional challenge and structured support offered by the adaptive mechanics. The tailored feedback and dynamic difficulty adjustments seemed to align well with their motivation to succeed and progress through the game efficiently. On the other hand, two explorer-type players from Group A reported experiencing less frustration with the original game. As none of the explorer-type players from Group B indicated a difference in frustration or boredom, this may also be credited to the fact they were initially more frustrated while learning the game mechanics, but when playing the second version (the original) they had gotten more exposure to the game and was therefore less frustrated. The remaining three participants did not experience a difference in frustration and boredom between the two games. A summary of the responses can be seen in Figure 5.1, where we can also see that the AI version retrieved more "No" responses overall - indicating a slightly higher satisfaction.

5.2.3.2 Perceived engagement

When asked "Did you find the game to be engaging?", a majority of five participants rated the original, non-adaptive version of Elara as 4, indicating a moderate to high level of engagement. The AI-enhanced version received more mixed ratings. One outlier rated the AI-enhanced version as a 2, indicating lower engagement. This individual had been initially categorized as an Achiever, but their responses and gameplay behaviour did not fully align with that profile, suggesting a possible misclassification.

As can be seen from Figure 5.2, both versions have similar medians around 4, indicating that overall engagement levels were comparable across the two versions. The AI-enhanced version shows greater variability, as reflected in its wider interquartile range and higher standard deviation.

We performed paired sample t-tests on the participants' responses to the question “Did you find the game to be engaging?”. This question was measured using a 5-

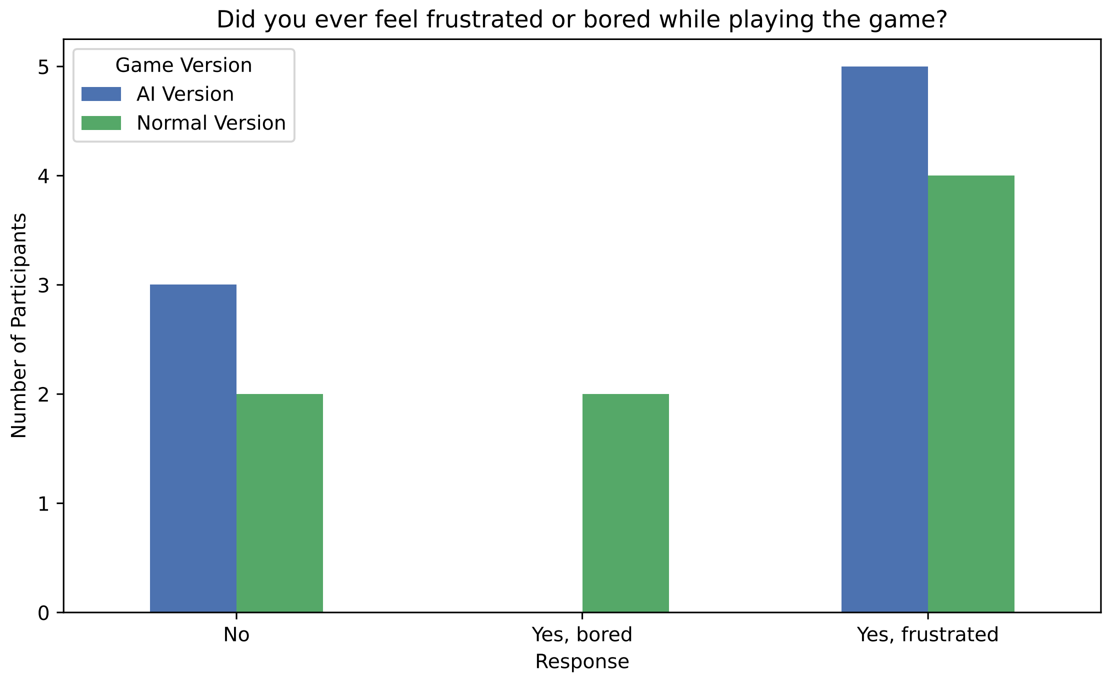

Figure 5.1: Frustration and Boredom Response for the AI Game Version and Original Game Version

point Likert scale, where a score of 5 represented the highest level of engagement, indicating that the participant found the game very engaging. By analysing these responses statistically, we aimed to determine whether there were significant differences in perceived engagement between the AI-enhanced and original versions of the game, providing insight into how adaptive features may influence player involvement and interest. The mean ratings were similar across versions, with no statistically significant difference found ($t(7) = 0.31$, $p = 0.76$). This suggests that participants felt equally engaged when playing both the AI-enhanced and the original game versions. Despite some variability in individual responses, the overall engagement level was consistent regardless of the presence of adaptive features.

5.2.3.3 Motivation

In response to "I felt motivated to continue playing the game," half of the participants rated the original version a 4. The AI-enhanced version received higher ratings overall, with three participants scoring it a 4 and another three scoring it a 5. This suggests that the adaptive features - such as hints and dynamic challenge levels - may have enhanced the sense of progression and motivation, especially for players who enjoyed overcoming increasingly complex tasks.

The t-tests also demonstrated a statistically significant increase in motivation ratings for the AI-enhanced game ($M = 4.38$, $SD = 0.52$) relative to the original version ($M = 3.63$, $SD = 0.52$), $t(7) = 3.42$, $p = 0.011$. This finding indicates that participants felt more motivated to continue engaging with the AI-enhanced version, suggesting that

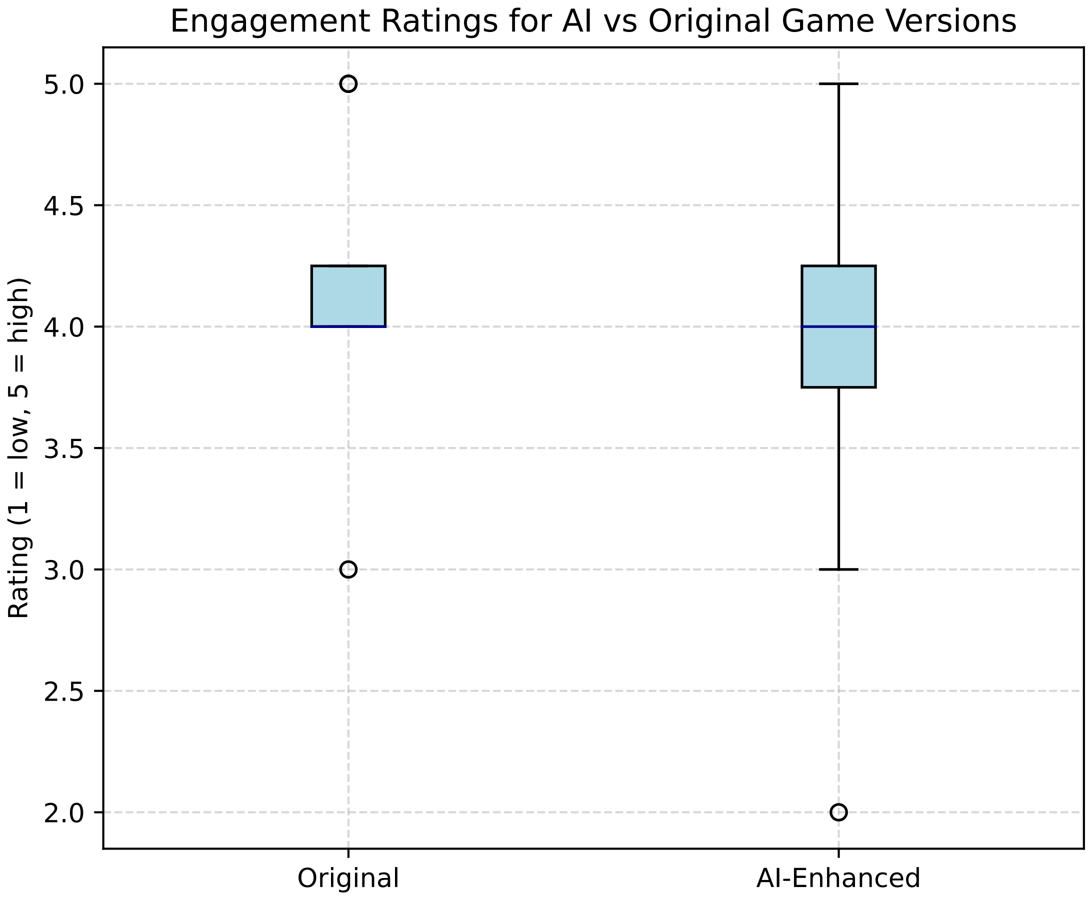

Figure 5.2: Engagement Ratings for the AI Game Version and Original Game Version

the adaptive features have positively influenced their willingness to persist with the game. This increase in motivation is particularly noteworthy given the small sample size (N = 8), which typically reduces statistical power. The fact that the difference reached significance despite this limitation points to a potentially meaningful impact of AI-driven adaptivity on user experience.

5.2.3.4 Challenge appropriateness

When asked to agree with the statement "The challenges in the game were at a good level for me," both versions received mostly positive ratings with a mean slightly above 4, indicating that players generally found the difficulty suitable and the adaptive difficulty felt well-balanced. The AI-enhanced version's challenge level was preferred by 2 participants scoring it higher than the original. However, one participant again gave it a low score (2), suggesting that adaptive difficulty may not benefit all users equally. This case also highlights that the underlying assumptions - such as how it classifies player types or adapts difficulty - may not always align perfectly with individual preferences or behaviours, potentially leading to mismatches in the user experience.

Analysis of the paired sample t-tests to the responses also revealed no statistically significant difference between the AI-enhanced and original versions of the game ($t(7) = 0.36$, $p = 0.73$). Participants generally perceived the difficulty levels of both versions to be comparable, indicating that the adaptive modifications incorporated in the AI-enhanced game did not notably impact their subjective experience of challenge. This finding suggests that, despite the presence of adaptive features designed to tailor difficulty dynamically, players' overall perception of how challenging the game was remained relatively consistent across both versions.

5.2.3.5 Guidance and hints

On the prompt "There should have been more or less guidance (hints)," most participants indicated that the original game offered too much guidance, giving it a mean score around 2,6 (on a scale where lower scores indicated better balance). In contrast, the AI-enhanced version was seen more favourably in this regard. Participants generally gave it lower scores with a mean around 2, implying that the amount and timing of guidance were better calibrated and more acceptable in the adaptive version.

A paired samples t-test was also performed to investigate whether players' perceptions of the amount of guidance (hints) differed significantly between the AI-enhanced and original versions of the game. The analysis showed no statistically significant difference in responses between the AI-enhanced game ($M = 2.00$, $SD = 0.82$) and the original version ($M = 2.63$, $SD = 1.19$), $t(7) = -1.53$, $p = 0.17$. Although the AI-enhanced game received slightly lower scores - suggesting that participants were somewhat more satisfied with the guidance provided - this difference did not reach statistical significance within this small sample. These findings imply that players generally found the amount of guidance to be acceptable in both versions of the game, with a trend toward preferring the more adaptive and context-sensitive hints in the AI-enhanced version. The lack of significant difference could reflect individual variation in players' preferences or the small sample size, which limits statistical power. While the adaptive hints in the AI-enhanced version may have improved user satisfaction with guidance, further research with larger participant groups is necessary to confirm these effects and explore how different player types respond to varying levels of instructional support.

5.3 Qualitative Findings

Open-ended feedback provided rich insights into participants' subjective experiences with both the AI-enhanced and original versions of Elara. Thematic analysis of participants' open-ended responses revealed six key themes as can be seen in Table 5.2: Engagement & Enjoyment, Challenge & Difficulty, Learning & Knowledge Acquisition, Usability & Game Mechanics, Emotional Response, and Preferences & Comparisons. These themes capture how participants perceived and interacted with both the original and AI-enhanced versions of the game.

Theme	Codes
Engagement & Enjoyment	Enjoyment (positive emotional response) Active engagement Atmosphere (affective environment supporting engagement) Aesthetic appreciation Motivation to continue Fun and playfulness
Challenge & Difficulty	Perceived challenge Negative emotion Difficulty with mechanics Stress and overstimulation Frustration due to difficulty Good difficulty balance
Learning & Knowledge Acquisition	Learning through play Cognitive stimulation Adaptiveness Scaffolding and progression newline Preferred streamlined experience Passive vs active learning
Usability & Game Mechanics	Design and usability Intuitive controls Difficulty with mechanics Missing features Pacing and flow Guidance and hint quality Frustration due to interface
Emotional Response	Enjoyment (positive emotional response) Frustration and confusion Stress and overstimulation Negative emotion Comfort and calm
Preferences & Comparisons	Preference for AI game Preference for original game Neutral perception Preferred streamlined experience Passive vs active learning Progression liked

Table 5.2: Thematic Grouping of Codes from Qualitative Analysis

5.3.1 Engagement and Enjoyment

Participants generally reported high levels of engagement and enjoyment in both versions of the game, though the AI-enhanced version elicited a wider range of emotional responses, including both enthusiasm and frustration. For the AI-enhanced version, participants often described the game as fun and mentally stimulating. P1 stated:

"It is fun to move the little thing around. I also enjoyed the music and the calmness of the game. It felt very positive." - Participant P1

Participant P6 referred to the game as a “fun brain activity.” Similarly, P5 remarked, “It was fun, I enjoyed programming,” and P7 described the experience as:

"It was challenging and fun. It felt good when I managed to complete the level." - Participant P7

Participant P4 appreciated the accessibility of the AI-enhanced game, stating:

"It was good, and easy to understand for someone who hasn’t programmed a lot before." - Participant P4

When discussing the most engaging aspects of the AI-enhanced game, participants emphasized autonomy and feedback. To the question "What aspects of the game did you like the most / (or) find most engaging? Why?", P7 described what they valued as:

"to follow instructions, some thinking and testing. Knowing instantly if it worked or not and able to try again." - Participant P7

P6 enjoyed “seeing the robot move,” and P8 noted that the “move functions to use” made it feel more fun.

For the original game version, responses were continuously positive and concise. P2 stated “It was fun.”, P7 said it was “fun and wanted to play more.”, and even P3, who had found the AI version frustrating, called the original “kind of intuitive.” Participants also reflected on what they found most engaging in the game overall. P2 and P3 enjoyed programming the robot, and P6 said they liked “figuring out how to solve the code.”

5.3.2 Challenge and Difficulty

Perceptions of difficulty varied across participants. Many participants (P1, P4, P5, P7, P8) appreciated the AI-enhanced version’s adaptive difficulty, noting that the personalized challenge helped maintain their motivation and engagement. Achiever-type players especially valued the dynamic adjustment of difficulty, allowing for a higher challenge level which matched their desire for a challenging experience. Similarly some achievers also expressed a lack of challenge in the original game-version, with participant P1 expressing:

"I thought the first game (original) was a bit too slow. It was a good difficulty level but at the same time I think it could have been a bit more challenging. I preferred the difficulty level in the second game (AI-version), it was good that it got more

challenging quickly" - Participant P1

However, participant P3 found the adaptive difficulty to be challenging, as it resulted in the participant expressing frustration over the steep curve in difficulty. P3 reported feelings of frustration, admitting: "I was stupid enough to not look at the instruction... it became frustrating", which highlights the potential effects of a mislabelling in player-type. P6 also stated that the ai-enhanced game "seemed hard at first", underscoring the potential difficulty spikes inherent in adaptive gameplay.

However, not all participants perceived a change in difficulty. Explorer-type participants, in particular, tended to report little to no noticeable difference between the two versions, with P2 stating, “I did not notice any difference.” Nonetheless, several participants — regardless of game version — praised the overall balance. P1, P4 and P7 described the levels as being good for beginners, and P4 expressed "It was good, a nice challenge", suggesting that while individual responses varied, the overall challenge level was acceptable for most.

5.3.3 Learning and Knowledge Acquisition

Participants’ responses revealed a range of perspectives on how effectively the games supported learning and knowledge development. The original game, with its more structured use of journals and video tutorials, was seen by some as educational but also slower and more passive. Participants P1, P2, P4 and P7 noted dissatisfaction with the original game’s instructional materials, such as the journal and YouTube videos, with P7 stating that they didn’t like reading or listening to the instructions, and P1 and P2 expressing:

"The youtube videos were too slow for me and in the end I only read the texts as it seemed to be the same info anyway. Maybe would be nicer to have the reading/educating part more fun as well." - Participant P1

"I did not see the journal in the second game, so that was better." - Participant P2

This feedback suggests that static or passive forms of instruction were less engaging, particularly for users who preferred more interactive or concise learning modalities. Conversely, the AI-enhanced version promoted more active learning through experimentation, which was noted by P4, P7 and P8. Participant P7 explained that they liked "thinking and testing different ways to reach the goal", and P8 expressed:

"I liked the try and error, it felt like I learned by doing." - Participant P8

Some participants also appreciated the AI-enhanced game’s ability to streamline educational content, noting that it allowed them to learn new concepts more efficiently. For example, one achiever-type participant highlighted:

"The second game (AI-version) was quicker. We learned a new function, but we did not need to go in the journal and watch the videos." - Participant P1

This reflects a preference for a more integrated learning experience, where instructional elements are embedded within gameplay rather than relying on separate tutorials. P1 also expressed positive feelings regarding the experience of using a serious game by stating "Good! Fun way of learning". Overall, while the adaptive version was often praised for supporting a smoother and more integrated learning process, these responses highlight the importance of balancing instructional clarity with interactivity and user autonomy. Participants generally favoured active, embedded learning experiences facilitating self-directed exploration over passive or disconnected instruction.

5.3.4 Usability and Game Mechanics

Usability and interface feedback emerged frequently. Several participants (P2, P3, P6, P8) highlighted the appealing graphics and the engaging opportunity to see and control a robot, with P2 stating "I liked to program the bot to do what I told it". Positive comments also included the intuitiveness of the controls and the game structure, with P3 stating that they enjoyed "That it was kind of intuitive", and P7 expressing:

"I liked to follow instructions, with some thinking and testing. I liked knowing instantly if it worked or not and able to try again." - Participant P7

Participants also seemed to appreciate the tutorial elements, describing them as a good introduction to the game mechanics and gameplay - highlighting the importance of structured onboarding for novice players. Participant P1 stated:

"It's good to start super basic, so the first levels was at a good level to learn the basics." - Participant P1

However, participant P3 who stated having some previous experience in programming, raised concerns about mechanical complexity, such as not being able to use loops in one version, or having to repeat actions manually which was described as "annoying".

"In the second game (AI-version) you had to type out everything you wanted to do, and for someone with some programming experience it becomes a little annoying since I already know what to write." - Participant P3

P3 also mentioned being unsure how to interact with elements in the AI-enhanced version, contributing to a higher cognitive load.

"For the buttons level I had to run the code 5 times to realize I would stand on the button." - Participant P3

Despite these challenges, many appreciated the flow and pacing of the tasks, particularly when game mechanics aligned well with task goals.

5.3.5 Emotional Response

The emotional reactions to the two game versions revealed a complex interplay of enjoyment, frustration, stress, and calmness. These responses were often rooted in how intuitive the game felt, how well the mechanics were understood, and how the game's design aligned with players' expectations or prior experience. Positive emotions were frequently expressed, particularly in relation to enjoyment and calmness. In both versions, participants described the games as "fun," "positive," and "interesting". The AI-enhanced version was praised by P1 as being "more fun," and "more demanding, which was more fun," suggesting that increased cognitive stimulation did not necessarily undermine enjoyment. However, this positive affect was also contrasted by negative emotions from participant P3, particularly frustration and confusion, most evident in the AI-enhanced version. P3 described their experience playing the ai-enhanced game saying "... it became frustrating".

Stress and overstimulation were also identified, particularly when participants felt rushed or unsupported. Participant P1 noted that the AI version was more stressful compared to the original version, stating:

"The last version (original) was more calm and soothing. So maybe this one was a little bit too stressed" - Participant P1

This contrast illustrates how emotional tone — influenced by pacing, feedback design, and cognitive load — shaped overall player experience. Interestingly, several participants expressed ambivalent emotional responses: acknowledging increased difficulty in the AI version while holding differing preferences based on how that challenge was experienced. For example, participant P1 stated "I preferred the difficulty level in the second game (AI-version)," highlighting an appreciation for a greater challenge. In contrast, P3 commented, "I preferred the first game (original) because I did not get frustrated right away," suggesting that lower initial difficulty supported a more comfortable experience. This indicates that challenge was not uniformly perceived as either positive or negative; for some, it enhanced engagement, while for others, it risked leading to frustration and disengagement. These preferences broadly aligned with the defined player types. Achievers tended to appreciate the increased challenge, whereas explorers favoured a more relaxed experience. One exception was participant P3, who was labelled as an achiever but whose feedback and emotional responses aligned more closely with explorer-type preferences, suggesting a possible mislabelling.

Finally, feelings of comfort and calm were particularly associated with the original game's slower pace and ambient features. Participant P1 described it as cozy, and appreciated being able to play the game at your own pace. These emotional responses indicate that atmosphere and pacing significantly influenced affective reception, especially among more exploratory player types.

"It was a little cozy to play the first game (original) in your own pace." - Participant P1

5.3.6 Preferences and Comparisons

Participant preferences revealed nuanced and sometimes conflicting evaluations of the two game versions, shaped by individual differences in learning style, programming experience, and emotional response.

Several participants expressed a clear preference for the AI-enhanced version, highlighting its quicker pace, reduced reliance on passive instruction (e.g., journals and videos), and a more engaging level of challenge. Participant P1 stated "It got more challenging quickly," which was appreciated by achiever-type players who preferred a faster progression. At the same time P1 also valued the original version for its calmer pacing, while P3 preferred the more deliberate, slower progression:

"I think I preferred the first game (original) because I did not get frustrated right away." - Participant P3

This suggests that the original version better supported conceptual understanding for some users, especially those who leaned toward reflective learning styles. There were also cases of neutral or mixed evaluations, where preferences were not tied to a specific version but rather to particular aspects of game design or user experience. For example, participant P2 stated:

"I did not see the journal in the second game, so that was better. I don't know if this was due to the different versions or only due to the fact that we were playing other levels. If we take these things into account I preferred the second game (AI)" - Participant P2

This indicates that it was sometimes unclear whether perceived differences stemmed from the AI-enhanced version itself or simply from the natural progression and variation between levels in the game. Importantly, certain design choices in the AI-enhanced version provoked frustration or confusion in some users. The "button level," in particular was cited as a pain point by P3. This aligns with coded themes such as frustration due to interface, difficulty with mechanics, and stress from unclear feedback, highlighting the potential trade-offs of reduced instructional scaffolding.

In terms of learning preference, a divide emerged between those who favoured passive learning (e.g., watching videos, reading instructions) and those who preferred active engagement, such as experimenting directly with code. P1 remarked:

"I only read the texts as it seemed to be the same info [as the video]. Maybe it would be nicer to have a little guy popping up and explaining everything." - Participant P1

Thereby suggesting a desire for more engaging instructional delivery, regardless of version. Ultimately, the data suggest that both versions offered valuable but distinct experiences. The AI-enhanced game appealed more to the achievers seeking challenge, speed, and autonomy, while the original version resonated with learners who appreciated clarity, structure, and conceptual guidance. These preferences appear to be influenced not only by learning style and programming experience, but also by emotional response, interface intuitiveness, and the perceived fairness or logic of in-game mechanics. Notably, participants identified as explorers perceived less difference between the AI-enhanced version and the original game compared to achievers, suggesting that the AI enhancements had a smaller impact on the gameplay experience for explorers than for achievers.

5.4 Summary of comparison

This study involved eight participants with diverse ages, genders, and gaming backgrounds, but limited prior programming knowledge, reflecting the beginner focus of the Elara game. While the small sample size limited robust inferential statistics, descriptive patterns and t-tests were analysed to identify meaningful trends.

5.4.1 Gameplay Behaviour and Learning Outcomes

Participants generally made more attempts per level and spent more time engaged with the AI-enhanced version compared to the original. This suggests a higher difficulty level, and together with the open-ended responses to the game-experience survey we can note that the adaptive features encouraged greater experimentation and iterative problem-solving. However, individual strategies varied: achiever-type players tended to exhibit a higher number of rapid attempts, indicating a goal-driven, exploratory approach, while explorer-type players often adopted more methodical, reflective strategies. Despite these behavioural differences, problem-solving accuracy - measured by the closeness of failed attempts to correct solutions - was similar across versions. Although the AI-enhanced game showed a slight higher accuracy, this difference was not substantial.

Both game versions led to improved understanding of programming concepts, as indicated by pre- and post-assessment coding scores. However, no significant differences were observed in short-term knowledge gains between the AI-enhanced and original versions, suggesting that we can not conclude whether adaptive features markedly impact immediate conceptual learning.

5.4.2 Subjective Experience and Qualitative Insights

Survey responses indicated that the AI-enhanced version was often perceived as more engaging, motivating, and appropriately challenging - particularly by achiever-type players who appreciated the dynamic difficulty and tailored support. Nonetheless, ratings of engagement and perceived challenge did not differ significantly across versions, reflecting diverse individual preferences. One participant experienced frustration with the adaptive difficulty, while most valued its integrated learning approach. Open-ended feedback reinforced these findings: the AI-enhanced version was praised for promoting active, self-directed learning and a more demanding yet enjoyable challenge. Conversely, players who found the AI game too challenging and frustrating tended to prefer the original version, appreciating its structured, clear and calm gameplay.

5.4.3 Overall Comparison

While the AI-enhanced Elara fostered greater motivation and encouraged exploratory play through its adaptive features, it did not yield statistically significant improvements in engagement or perceived challenge compared to the original version. Player type strongly influenced responses, with achievers generally favouring the adaptive environment, whereas explorers showed no strong preference. These findings highlight the complexity of designing adaptive learning games that effectively balance challenge and support for diverse learner profiles, emphasizing the need for tuning adaptivity to accommodate varying player preferences and avoid potential frustration.

6

Discussion

This study investigated how AI-driven player experience modelling (PEM) influences engagement and skill acquisition in serious games for software engineering education, as well as the advantages and disadvantages of AI-driven PEM compared to traditional game models. The following discussion interprets the quantitative and qualitative results in light of these research questions, highlighting key insights, implications, and limitations.

6.1 Delimitations

This study is delimited to the investigation of AI-driven Player Experience Modelling (PEM) within the context of serious games (SGs) designed for software engineering education. The scope is intentionally focused on the integration and evaluation of the AI-driven DualTypePEM model developed for the Elara programming game. While this provides a controlled environment to explore the impact of PEM on engagement and skill acquisition, the results may not generalize to other serious games in software engineering or to educational games in other domains. These delimitations were necessary to ensure a focused and in-depth analysis of AI-driven PEM within a concrete and relevant use case.

6.2 Research Questions

This section interprets the findings of the study in relation to the two central research questions:

- RQ1: How does AI-driven player experience modelling improve engagement and skill acquisition in serious games for software engineering education?

- RQ2: What are the advantages and disadvantages of AI-driven PEM in serious games for SE compared to traditional game models that do not incorporate AI?

Drawing on both quantitative data (e.g., engagement metrics, survey responses, and learning assessments) and qualitative insights from participant feedback, we analyse how the DualTypePEM model influenced player experience and educational outcomes in the Elara programming game. The discussion also considers how different player types responded to adaptivity, and what this reveals about the broader potential and limitations of integrating AI into serious games for technical education. By unpacking each research question, we aim to understand not only whether AI-enhanced PEM makes a difference, but how and for whom it makes that difference - laying the groundwork for future improvements in adaptive learning design.

6.2.1 RQ1: How AI-driven PEM Improves Engagement and Skill Acquisition

The analysis of player engagement and skill acquisition indicates a nuanced impact of AI-driven PEM in serious games for software engineering education.

6.2.1.1 Engagement

Quantitative survey results showed no statistically significant difference in overall engagement between the AI-enhanced game and the original version. However, qualitative feedback and gameplay metrics suggest that AI-driven PEM positively influences engagement in specific ways. For instance, participants spent more time on levels and made more attempts in the AI-enhanced version, implying increased experimentation and deeper interaction. Players described the AI-driven game as more demanding but also more fun and motivating. This was particularly evident among "achiever" type players, who appreciated the adaptive difficulty and personalized feedback, which kept the challenge aligned with their skill level. The results also align with previous literature suggesting that adaptive game elements can enhance player engagement by tailoring difficulty and support to individual needs, thereby maintaining interest and reducing frustration. However, it is important to interpret these findings cautiously and to verify them with larger samples and more diverse player populations.

6.2.1.2 Skill Acquisition

In terms of learning outcomes, both versions of the game led to improved knowledge of software engineering concepts as evidenced by pre- and post-assessments. However, there was no significant difference in short-term conceptual knowledge improvement between the AI-driven and original versions. This may be due to the limited exposure participants had to the game during the study. Given more extended interaction and repeated play sessions, the adaptive nature of the AI-driven PEM could have a more pronounced effect on skill acquisition and knowledge retention. A more extensive study might reveal differences in long-term learning gains, as players adapt more deeply to the personalized challenges and feedback provided by the AI system. Despite this, the qualitative data reveal that AI-driven PEM promotes a more active, exploratory learning process. Unlike the original version, which relied on static tutorials and a structured journal, the AI-enhanced game encouraged players to learn through trial, error, and immediate adaptive feedback.

This shift toward self-directed learning aligns with educational theories emphasizing active engagement for deeper understanding and skill development.

Furthermore, the immediate post-test assessments may not fully capture the depth of learning promoted by the AI-driven approach. Skill acquisition in complex domains like software engineering often requires iterative practice and reflection, which AI-driven PEM facilitates through ongoing adjustment of difficulty and tailored support. Therefore, while short-term measurable knowledge gains appear similar, the AI-enhanced experience may better support sustained engagement and gradual mastery over time. In real-world settings, where continued voluntary use of educational games is critical, the personalized feedback and adaptive challenges provided by the AI can sustain learners' interest and encourage ongoing interaction. This sustained engagement may ultimately result in more meaningful and lasting skill development than what short-term assessments alone can capture.

The AI-driven PEM appeared to increase players' motivation to continue playing compared to the original game version. This heightened motivation can positively influence both engagement and skill acquisition outside of the controlled study environment. When learners are intrinsically motivated to engage with the material repeatedly, they have more opportunities to practice and reinforce their skills, leading to deeper understanding and improved performance over time.

6.2.2 RQ2: Advantages and Disadvantages of AI-driven PEM

This study explored how AI-driven player experience modelling compares to traditional, non-adaptive serious games in software engineering education, focusing on its benefits and drawbacks as experienced by participants. In this section we will discuss the advantages and disadvantages of AI-driven PEM.

6.2.2.1 Advantages of AI-driven PEM

Enhanced Engagement and Motivation for Specific Player Types. The AI-enhanced version of the game was perceived by many participants - especially achiever-type players - as more engaging and motivating. Quantitative data showed a statistically significant increase in motivation ratings for the AI-driven game for both player-types. Participants appreciated the dynamic difficulty adjustments and tailored feedback, which seemed to sustain their interest and foster a stronger drive to continue playing. This aligns with qualitative reports describing the AI-enhanced game as "more demanding, which was more fun," indicating that adaptive challenge can positively influence player involvement.

Personalized Challenge and Support. AI-driven PEM enabled the game to adjust difficulty levels and hint availability in response to player performance and style. Participants noted that the AI-enhanced version provided a better balance of challenge, with fewer instances of boredom or frustration reported by achievers. The adaptive features allowed for a more streamlined and active learning experience, where instructional content was embedded within gameplay rather than relegated to separate tutorials. This facilitated quicker learning and reduced reliance on external educational materials, as some participants preferred the AI version for being "quicker" and more conducive to self-directed exploration.

Encouragement of Experimentation and Iterative Learning. Gameplay metrics showed that participants generally made more attempts per level and spent slightly more time on AI-enhanced levels. This suggests that the adaptive elements may have encouraged experimentation and iterative problem-solving for some users by adjusting challenge and feedback. However, subjective responses indicated that this effect was not uniform - while achiever-type players often appreciated the increased difficulty and reduced scaffolding, others experienced frustration or cognitive overload. These findings highlight the potential of adaptive gameplay to support active learning, but also underscore the importance of tailoring difficulty and support to diverse learner preferences.

6.2.2.2 Disadvantages of AI-driven PEM

Inconsistent Impact on Learning Outcomes. Despite positive trends in motivation and engagement for most players, the adaptive features did not produce a statistically significant improvement in post-assessment programming scores compared to the original game. This indicates that while AI-driven PEM may enhance motivation, its impact on immediate learning gains can be limited or vary across individuals. However, it is also important to consider limitations in the assessment design when interpreting these results. The uniformly high performance across participants suggests that the programming assessment may have been too easy, or the gameplay sessions were too short to detect meaningful differences in learning outcomes. A potential ceiling effect could have obscured variations in player learning that a more challenging or comprehensive evaluation might have revealed. Future studies using more robust and differentiated assessments could offer better insight into how adaptive mechanics influence programming competence over time.

Potential for Increased Frustration and Cognitive Load. One participant reported frustration and overstimulation when interacting with the AI-enhanced version, particularly due to sudden spikes in difficulty and less intuitive mechanics (e.g., confusion around button-level tasks). The participant explicitly described the experience as "frustrating" because of unclear guidance and interface issues. This highlights a risk that adaptive systems, if not carefully calibrated, can introduce challenges that exceed some players' current capabilities or preferences, potentially reducing enjoyment and hindering learning.

Mismatch with Player Types and Preferences. While achievers generally benefited from and appreciated the AI-driven adaptivity, explorer-type participants often perceived little difference between the two game versions. Explorers preferred a more relaxed pace and reflective learning style, which the traditional game also supported. One participant reported frustration when the adaptive game increased the difficulty. This reaction may have stemmed from a misclassification of the participant as an Achiever type, leading the system to emphasize challenge progression when the player might have preferred consistency or narrative immersion instead. This highlights a key risk of adaptive gaming: if the system incorrectly classifies a player's type or preferences, the adaptations can backfire - disrupting engagement rather than enhancing it.

6.3 Broader implications

The findings of this study hold several broader implications for the design of AI-driven PEM in serious games for software engineering education. While the sample size was limited, the emerging patterns provide valuable insight into how adaptive systems may shape learning experiences, engagement, and educational outcomes in novice programming contexts.

First, the integration of AI-driven adaptivity demonstrates potential for tailoring learning experiences to individual player profiles, particularly for those with goal-oriented (achiever) motivations. These players generally responded positively to dynamic challenge levels, personalized feedback, and reduced reliance on static instructional materials. This suggests that AI-enhanced PEM systems can support more autonomous, explorative learning pathways by responding to gameplay behaviour and adjusting support accordingly. In practical terms, such systems may help reduce cognitive overload by scaffolding learning just in time, especially for users who are motivated by rapid progression and iterative problem-solving.

Second, the variability in player responses underscores the importance of designing adaptive systems that are not just reactive but flexible and interpretable. For some users - particularly those who preferred more reflective or exploratory engagement styles (explorers) - the AI-enhanced version was either not noticeably different from the non-adaptive version or, in at least one case, perceived as confusing and frustrating. This suggests that the adaptivity may not have been sufficiently transparent or impactful for all players, and highlights the need for clearer signalling of adaptive behaviours and better alignment with user needs. It also underscores the importance of designing adaptive systems that are intuitive and accessible, especially for users who may not immediately recognize or understand the nature of in-game adjustments. This illustrates a central design tension: adaptivity can enhance motivation and engagement for some users, while alienating others if misaligned with their cognitive or emotional needs. Therefore, successful PEM implementation in serious games must go beyond simple player categorization and instead support hybrid models that accommodate a broader spectrum of learner preferences and play styles.

Third, the feedback related to instructional design highlights a critical shift in user expectations: learners increasingly prefer embedded, interactive guidance over passive, externalized instruction. The AI-enhanced version's ability to deliver feedback within the gameplay loop rather than through detached videos or journals was more positively received. This points toward the broader potential for AI-driven games to redefine how pedagogical content is delivered, shifting the focus from didactic explanation to experiential, trial-and-error learning. In turn, this may facilitate deeper conceptual engagement for novice learners, especially when reinforced through meaningful feedback and personalized progression.

Finally, the emotional dimension of gameplay - frustration, calmness, stress, enjoyment - emerged as a significant mediator of learning and preference. Emotional response is often overlooked in serious game design but plays a vital role in user retention and efficacy. The findings suggest that adaptive systems must be calibrated not just for cognitive fit but also for emotional resonance, ensuring that gameplay remains within a player's optimal zone of challenge without tipping into stress or boredom.

As such, these results suggest that AI-driven PEM has promising but nuanced advantages over static serious games in software engineering education. The challenge moving forward lies in designing adaptive systems that are both personalized and inclusive, capable of adjusting to individual needs while maintaining transparency, fairness, and emotional balance.

6.4 Limitations

The study may have limitations, which should be acknowledged when interpreting the results.

A key limitation of this study lies in the manual classification of participants into player types (e.g., achievers and explorers), which was based on the researcher's interpretation of observed behaviours and performance metrics rather than a data-driven or psychometrically validated model. This subjective approach introduces the potential for bias or misclassification, affecting the accuracy of conclusions drawn about differences in engagement and performance between player types. Since the adaptive system relied on this initial classification to tailor gameplay experiences, any inaccuracies could have led to suboptimal adaptation strategies. This issue was reflected in reports from one participant who expressed a different experience with the AI-enhanced version that did not align with the response from the rest of the participants in the same player-type, suggesting a mismatch. These limitations highlight the need for more nuanced, dynamic, and user-informed models for identifying and adapting to learner profiles in future implementations of AI-driven PEM.

Learning outcomes were measured through immediate post-game assessments, which may not fully reflect the long-term educational impact of the AI-driven PEM. Skill acquisition in complex domains such as software engineering often requires sustained practice and reinforcement over time. The limited duration of gameplay in this study may have constrained the extent to which adaptive features could influence deeper conceptual understanding or retention. The controlled nature of this study may also limit the extent to which the findings generalize to real-world learning environments, and certain limitations in the research instruments may have affected the depth and accuracy of the results. For example, the game performance data collected from Elara—such as time spent on tasks and number of attempts offers useful indicators of gameplay behaviour, but may not capture the full complexity of the learning experience. These metrics provide only a surface-level view and may overlook important aspects such as conceptual understanding, cognitive engagement, or long-term knowledge retention. As a result, deeper learning outcomes may not have been fully assessed through the available data.

Another limitation is the nature of self-report data, which is used in the surveys. Participants' responses may have been influenced by social desirability or recall bias, potentially impacting the accuracy and consistency of their reported experiences. For example, one participant indicated in an open-ended response that they enjoyed the more challenging obstacles in the "first" game, but later referred to those same obstacles as being part of the "second" game. Such inconsistencies suggest that some responses may reflect retrospective confusion or a desire to provide favourable answers, which could compromise the reliability of subjective engagement and learning assessments.

The relatively small number of participants also limited the statistical power of quantitative analyses. While qualitative and gameplay data revealed meaningful patterns, the small sample restricts the generalizability of findings and may potentially obscure real differences in engagement or learning outcomes between the AI-driven and original game versions.

6.4.1 Threats to Validity

This study faces threats to validity common in exploratory research.

Internal validity may be affected by the manual and subjective classification of player types, potentially introducing bias or misclassification that impacts the adaptation and participant responses.

External validity is restricted by the small, relatively homogeneous sample and controlled laboratory environment, which may limit the generalizability of findings to broader populations or real-world educational settings.

Addressing these threats in future work will help strengthen the reliability and applicability of findings.

6.5 Future Work

This study highlights several opportunities for expanding and deepening research on AI-driven PEM in serious games for software engineering education.

One key area for future work is improving the identification of player types. In this study, player profiles (e.g., achievers and explorers) were assigned manually based on observed gameplay behaviour and performance, introducing a degree of subjectivity. Future studies could benefit from employing validated player-type questionnaires or leveraging automated behavioural clustering methods. These data-driven approaches can increase the objectivity and reliability of player classification, allowing adaptive systems to more accurately align gameplay experiences with individual preferences and motivational styles. In addition, future research should investigate the development of more refined and flexible adaptive mechanisms. While this project employed a heuristic-based PEM, future work should explore the use of machine learning to create more dynamic and responsive models. One of the current study's limitations was the lack of a large, diverse dataset necessary for training robust machine learning models. With access to more extensive gameplay and learning outcome data, machine learning techniques could uncover complex behavioural patterns that rule-based systems may overlook.

An important yet currently overlooked factor in the PEM framework is the emotional state of the player. As noted by Westera et al. [36], emotion plays a critical role in memory, decision-making, and learning. However, the current system does not account for players' emotional states, largely due to the difficulty of detecting emotion in real time. As emotion recognition technologies (e.g., facial analysis, voice sentiment, physiological sensors) become more accessible and reliable, integrating emotional data into PEM could greatly enhance the effectiveness of adaptive feedback and personalized support. Including emotional awareness in serious game design could help maintain players in their optimal zone of challenge—avoiding states of frustration or boredom—and further improve both engagement and learning outcomes.

Finally, future studies should involve larger and more diverse samples to improve statistical power and generalizability. Incorporating learners from different backgrounds, with varied experience levels and educational contexts, would offer deeper insights into how different groups respond to adaptive systems. Longitudinal studies are also needed to assess whether short-term gains in motivation and engagement translate into sustained learning benefits and skill retention over time.

In sum, advancing PEM research will require not only more sophisticated technical approaches but also a deeper integration of cognitive, emotional, and behavioural factors. Addressing these challenges will be key to unlocking the full potential of adaptive serious games in educational contexts.

7

Conclusion

This study set out to explore the integration of AI-driven Player Experience Modelling (PEM) in serious games for software engineering education, addressing a gap in existing research where AI's potential in adaptive game-based learning remains underutilized. By implementing adaptive difficulty and personalized guidance within the programming game Elara using the DualTypePEM model, the research examined how AI-enhanced gameplay can influence engagement and skill acquisition in programming education.

The results offer promising insights into the role of AI in fostering personalized learning experiences. Participants exposed to AI-driven PEM generally reported higher motivation and engagement, although the findings were influenced by individual player-types. The study also highlighted important challenges, such as the difficulty of accurately identifying player types and the limitations of short-term assessments in capturing deeper learning effects.

Beyond empirical findings, this project contributes to the field by presenting a practical example of PEM implementation in a serious game context, offering both methodological and design-based takeaways for future work. It emphasizes the importance of adaptive systems that respond to player behaviour, while also acknowledging the limitations of rule-based models and the need for more scalable, data-driven approaches.

Looking forward, future research should aim to refine adaptive mechanisms through robust player profiling techniques, incorporating emotional state recognition and leveraging machine learning for dynamic PEM could significantly enhance personalization. Expanding the study to larger and more diverse learner populations will also be critical for validating and generalizing these findings.

In conclusion, while this study represents an early step into the intersection of AI-enhanced PEM, serious games, and software engineering education, it underscores the transformative potential of AI-driven personalization in educational games. As technology and methodologies advance, such systems can evolve into powerful tools for delivering meaningful, engaging, and effective learning experiences.

Bibliography

[1]

[2] Leonard A. Annetta. The "i's" have it: A framework for serious educational game design. *Review of General Psychology*, 14(2):105 – 112, 2010. Cited by: 354.

[3] Sander Bakkes, Chek Tien Tan, and Yusuf Pisan. Personalised gaming: A motivation and overview of literature. *Journal. Creative Technologies*, 07 2012.

[4] Sander C.J. Bakkes, Pieter H.M. Spronck, and Giel van Lankveld. Player behavioural modelling for video games. *Entertainment Computing*, 3(3):71–79, 2012. Games and AI.

[5] Richard Bartle. Hearts, clubs, diamonds, spades: Players who suit muds, 1996.

[6] Till Becker. The character of successful trainings with serious games. *International Journal of Emerging Technologies in Learning (iJET)*, 5(SI3), Nov 2010.

[7] David S. Bickham, Elizabeth A. Vandewater, Aletha C. Huston, June H. Lee, Allison Gilman Caplovitz, and John C. Wright and. Predictors of children's electronic media use: An examination of three ethnic groups. *Media Psychology*, 5(2):107–137, 2003.

[8] Keivan Borna and Hesam Maghami Rad. Serious games in computer science learning goals. *2018 2nd National and 1st International Digital Games Research Conference: Trends, Technologies, and Applications (DGRC)*, Nov 2018.

[9] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[10] Alex Browne. Elara: A free coding game, 2024.

[11] Abigail Gertner, Cristina Conati, and Kurt Vanlehn. Procedural help in andes: Generating hints using a bayesian network student model. 09 1998.

[12] Glenda A. Gunter, Robert F. Kenny, and Erik H. Vick. Taking educational

games seriously: Using the retain model to design endogenous fantasy into standalone educational games. *Educational Technology Research and Development*, 56(5-6):511–537, Oct 2007.

[13] Carrie Heeter and Brian Winn. *Gender Identity, Play Style, and the Design of Games for Classroom Learning*, pages 281–300. 09 2008.

[14] W. Lewis Johnson, Hannes Vilhjalmsson, and Stacy Marsella. Serious games for language learning: How much game, how much ai? *Frontiers in Artificial Intelligence and Applications*, 125:306 – 313, 2005. Cited by: 114.

[15] Manal Kharbouch, Aurora Vizcaino, José Alberto García-Berná, Félix García, Ambrosio Toval, Oscar Pedreira, Ali Idri, and José Luis Fernández-Alemán. Uncharted dimensions, gaps, and future trends of serious games in software engineering. *Computer Standards Interfaces*, 92:103915, Aug 2024.

[16] Richard L. Lamb, Leonard Annetta, Jonah Firestone, and Elisabeth Etopio. A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. *Computers in Human Behavior*, 80:158–167, 2018.

[17] Cristina Leza. So you think you don't have enough data to do machine learning, Oct 2020.

[18] Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino. Euclidean distance geometry and applications. *SIAM Review*, 56, 05 2012.

[19] Brian Magerko, Carrie Heeter, Joe Fitzgerald, and Ben Medler. Intelligent adaptation of digital game-based learning. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, Future Play '08, page 200–203, New York, NY, USA, 2008. Association for Computing Machinery.

[20] Christos Malliarakis, Maya Satratzemi, and Stelios Xinogalos. *Educational Games for Teaching Computer Programming*, pages pp 87–98. 06 2014.

[21] Roy Mathew, Sohail Iqbal Malik, and Ragad Moufaq Tawafak. Teaching problem solving skills using an educational game in a computer programming course. *Informatics in Education*, 18(2):359–373, 2019.

[22] Arash Moradi-Karkaj. Serious interactive digital narrative: Explorations in personalization and player experience enrichment. In *2021 International Serious Games Symposium (ISGS)*, pages 35–42, 2021.

[23] University of Florida. Bloom's taxonomy, 2023.

[24] Christopher Pedersen, Julian Togelius, and Georgios Yannakakis. Modeling player experience for content creation. *Computational Intelligence and AI in Games, IEEE Transactions on*, 2:54 – 67, 04 2010.

[25] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee and. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3):45–77, 2007.

[26] Werner Siegfried Ravyse, A. Seugnet Blignaut, Verona Leendertz, and Alex Woolner. Success factors for serious games to enhance learning: a systematic review. *Virtual Reality*, 21(1):31–58, Sep 2016.

[27] Ashwathy T. Revi, David E. Millard, and Stuart E. Middleton. A systematic analysis of user experience dimensions for interactive digital narratives. In Anne-Gwenn Bosser, David E. Millard, and Charlie Hargood, editors, *Interactive Storytelling*, pages 58–74, Cham, 2020. Springer International Publishing.

[28] Syed A. Q. Rizvi, Umair Rehman, Shi Cao, and Bradley Moncion. Exploring technology acceptance of flight simulation training devices and augmented reality in general aviation pilot training. *Scientific Reports*, 15(1), 2025. Cited by: 0; All Open Access, Gold Open Access.

[29] Christian Roth and Hartmut Koenitz. Evaluating the user experience of interactive digital narrative. In *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, AltMM '16, page 31–36, New York, NY, USA, 2016. Association for Computing Machinery.

[30] David Williamson Shaffer. *How Computer Games Help Children Learn*. Palgrave Macmillan US, New York, 2006.

[31] Craig Smith and Richard Lazarus. *Emotion and Adaptation*, volume 21, pages 609–637. 01 1990.

[32] Hrvoje Smolic. How much data is needed for machine learning? | graphite note, Dec 2022.

[33] Stephen Tang, Martin Hanneghan, and A. El Rhalibi. Describing games for learning: Terms, scope and learning approaches. 01 2007.

[34] Rust Team. Rust programming language, 2018.

[35] TypeScript. Typescript - javascript that scales.

[36] Wim Westera, Rui Prada, Samuel Mascarenhas, Pedro A. Santos, João Dias, Manuel Guimarães, Konstantinos Georgiadis, Enkhbold Nyamsuren, Kiavash Bahreini, Zerrin Yumak, Chris Christyowidiasmoro, Mihai Dascalu, Gabriel Gutu-Robu, and Stefan Ruseti. Artificial intelligence moving serious gaming: Presenting reusable game ai components. *Education and Information Technologies*, 25(1):351–380, Jul 2019.

[37] Nick Yee. Motivations for play in online games. *CyberPsychology & Behavior*, 9(6):772–775, 2006. PMID: 17201605.

[38] Jichen Zhu and Santiago Ontañón. Player-centered ai for automatic game personalization: Open problems, 2021.

A

Appendix

A.1 Survey Questions

Pre-Assessment Questions

- Age:
- Gender:
- How would you rate your experience with computer games? (Likert scale: 1 = No experience to 5 = Expert)
- How would you rate your experience with programming? (Likert scale: 1 = No experience to 5 = Expert)
- What is the purpose of a "loop" in programming? (Multiple choice)
  - To define a block of code that can be reused
  - To store and retrieve values from memory
  - To repeat a set of instructions until a condition is met
  - I don't know
- What does a "function" generally do in programming? (Multiple choice)
  - It defines a reusable block of code that can be called when needed
  - It organizes code into sections for better readability
  - It is used to store variables in memory
  - I don't know

Game Experience Feedback

Likert Scale Questions

Participants responded on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) to the following statements:

- I found the game to be engaging.
- I felt motivated to continue playing the game.
- The challenges in the game were at a good level for me.
- There should have been more or less guidance (hints).

Open-Ended and Multiple-Choice Questions

- What was your experience playing the game?
- What aspects of the game did you like the most / (or) find most engaging? Why?
- What aspects of the game did you like the least / (or) find least engaging? Why?
- Did you ever feel frustrated or bored while playing the game? (Multiple choice)
  - Yes, bored
  - Yes, frustrated
  - Yes, both
  - No
  - Other (please specify)

Post-Assessment Questions

- What is the purpose of a "loop" in programming? (Multiple choice)
  - To define a block of code that can be reused
  - To store and retrieve values from memory
  - To repeat a set of instructions until a condition is met
  - I don't know

- What does a "function" generally do in programming? (Multiple choice)
  - It defines a reusable block of code that can be called when needed
  - It organizes code into sections for better readability
  - It is used to store variables in memory
  - I don't know
- Which differences did you notice between the two games?
- Which game did you prefer? Why?