Player model analysis for adaptive content delivery in an educational game

By

Huinan Jiang

Player model analysis for adaptive content delivery in an educational game

Submitted in fulfilment of the requirements for the degree of
Master of Science
In
Computer Science

Committee member:

Supervisor:	Dr. ir. Rafael Bidarra
Mentor:	Dr. Anna Vilanova
Member:	Dr. ir. Geertje Bekebrede

Computer Graphics and Visualizations Group
Faculty of
Electrical Engineering, Mathematics and Computer Science
Delft, the Netherlands

Abstract

Gamification and game-based learning have received wide attention in the past few decades. By blending game design and mechanics into traditional learning environment, they enhance students' participation, motivation and engagement. Squla is such a gamified learning platform where we can find game components like coins, virtual shop, competitive activities and collectables. Besides the above gamified elements, Squla has also transformed standard questions into fun games such as shooting catapults and clicking the popping bubbles. These games are designed to further engage the students and improve their learning.

In this project, we analysed player game type preference based on their game log data, and measure the impact of customised game type delivery. We targeted education group 4 and 5 students users, and focus on catapult games and bubble popper games as they are the most played. A set of features that could reflect students' preference and emotion states are selected and analysed, including correct ratio, playtime, quitting possibility, etc. Using data clustering, we group students who have similar behaviour and predict their preferred game types. We identified three group of students, one shows high completion rate on all forms of questions, another shows rather low overall completion rate, and the last group has rather high completion rate on bubble popper games and lower completion rate on the catapult shooting games. Based on such findings, we conducted experiment on them to look into different gaming contents' impact on their learning and engagement. A final experiment consists of a short math quiz and a follow-up questionnaire. The two-week online experiment receives 91 valid responses. Post-play questionnaire, as well as the game log suggest different contents could affect students' engagement. In particular, preferred contents can elevate a sense of happiness and enhance perceived learning.

Preface

This project is a part of my Master thesis work in pursuing a degree in Computer Science in TU Delft. For a long time, I've been having a strong interest in game industry. Current game studies on player behavioral analysis is no doubt both inspiring and challenging.

The work of this project is done in Squla, a game-based learning platform in the Netherlands. I would like to thank my thesis supervisor Dr. Rafael Bidarra for suggesting this topic to me, and for his great help in directing my work and writings. I would like to thank Joey Van Der Kaaij, my company mentor for helping me solve the problems I encountered and conduct the online experiments. I would also like to thank Peter Hofstede, Eric Bouwers and Francette Broekman in Squla for their valuable advice on data analyse and online experiment design.

Finally, I would like to thank my friend Hongping Feng and Liam Mac an Bhaird for giving me great advice on thesis report writings.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Background

For students, learning is not always an easy process. Some of the contents they need to learn may be considered boring and dry. Educators, researchers and parents have been seeking ways to better motivate the students, and game is no doubt one of the most effective medium. Games have the power of motivating and engaging people, keeping them sitting there for hours trying to complete the tasks. In traditional classroom setting, for example, we see history teachers asking students to perform a short play to reproduce historical events, geography teachers give students world map puzzle to enhance their memory. In today's e-learning environment, digital games have been applied for teaching. Currently, there exist quite a few game-based learning applications, especially for teaching mathematics, spelling and logic. As the use of digital devices continues to grow, the game-based learning market keeps expanding.

Meanwhile, the highly interactive nature of video games makes them easy to be adapted to different players. In entertainment games, contents such as game scenes, narratives, rules, levels and objects can all be adaptable. For educational games, they can further be adapted to cater to students' skill, knowledge and learning objectives. Difficulty adaptation, for example, is one of the most extensively researched topics (Ricardo Lopes 2011). Having considered students' initial knowledge and learning goal, the amount and difficulty of questions can be adjusted to provide better service.

While entertainment game players have the freedom to decide on which game to play and which not to, students playing educational games have rather limited choices. Because the design of educational games is strongly influenced by the educational target itself, and students have to put learning at a high priority. If a game happens to suit a student's taste, he may be highly engaged and eager to continue. If not, we are at the risk of causing a negative influence on his learning and eventually lose him as a user.

1.2 Overview of Squla

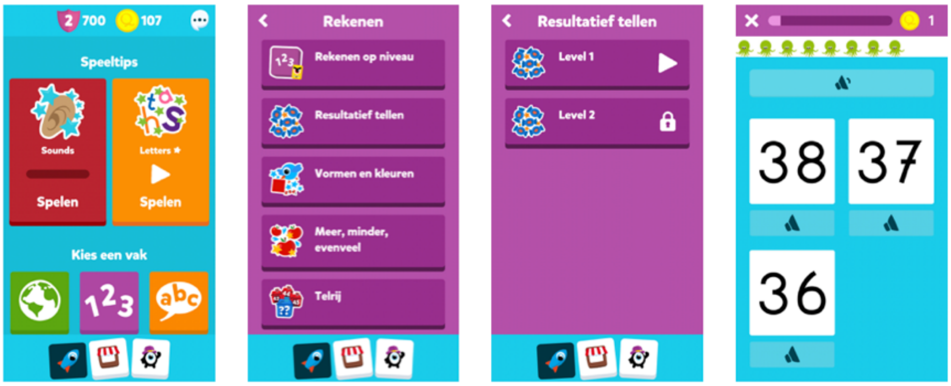


Figure 1. Interface of Squla Application

Our target game-based learning platform, Squla (Squla n.d.), offers kids over a dozen subjects such as Math, Language, Spelling, History, Music and Geography. Each subject has several sub-topics, and each of the sub-topics will lead to a set of questions assigned to different levels. Squla targets young kids, from peddlers to primary school students. Students can order a monthly, half-year or yearly membership that will allow full access to regular training. Recently Squla collaborates with other educational services and introduces extra curriculums such as the Bitsbox X Squla for computer programming. Squla now has over 600,000 users and still growing rapidly.

Gamification has been applied on Squla platform. Upon answering each question correctly, a student will gain certain experience points and virtual coins and receives trophies once meeting a certain requirement. Students can also purchase virtual avatar or actual product with the coins. Like leader board commonly seen in video games, there is a weekly scoreboard recording the top-performing students. Students can also add friends and compete with each other.

Other than the above gamified components, casual games have also been designed to replace the regular form of questions. Such as games with clickable floating bubbles containing answers replacing regular selection questions.

1.3 Motivation

In the traditional classroom settings, one-to-one tutoring has been proved to be beneficial for learning because it offers personalized teaching. Due to the lack of educational resources, it cannot be done for everyone. However, in e-learning, we are able to do so since online teaching can be fully automated.

Squla has already attempted adaptive learning by dynamic difficulty adjustment in special sets of quizzes. The questions are assigned difficulty and confidence value. When playing the quiz, the choice of each following question will rely on students' previous performance. However, Squla has not yet explored the adaptivity in other game components such as game types. It's interesting for them to know how we can learn about students' preference and deliver games they enjoy the most.

Preferred contents should lead to higher engagement, which is an essential part of both learning and gameplay. Higher engagement then leads to better learning outcome – our final objective.

Our goal in this thesis project is to analyse player behavior to understand their gaming preference, then find out the impact on students engagement when they are exposed to different games. The first step is done through player modelling. The development of data science provides the technology for constructing player models. We then measure the impact on the engagement of different contents, proves the benefit from offering preferred contents in learning.

1.4 Research questions

This thesis project aims at answering the following research questions:

1. How to model player preference in an education game environment?

2. How can such model be applied in customized game content delivery?

3. What’s the impact of customized content delivery?

The first question is about player preference analysis, in actual application, we target game type preference. We first collect data, then extract and select features related to our research question. We conduct correlation analysis to discover relations among the data and identify the potential impacting factors. Then, player clustering is conducted to understand their behavioral patterns that lead us to distinguish their preference.

In answering the second and third questions, we initiate a simple game type adaptation by giving the most preferred contents to players and measure their level of engagement. It is shown that customized content has a positive effect. We further discuss the possibility of adaptive contents based on a literature review.

1.5 Thesis structure

The thesis report is structured into 5 Chapters. Having briefly introduced the main background and research questions in the current chapter, the rest chapters are organized as follows: In Chapter 2 we introduce work related to our project, including player modelling, game adaptation and the engagement assessment. Chapter 3 shows the first part of the work of this project, including how we select the data, extract the features, conduct analysis and construct the player model. In Chapter 4, we describe how we conduct the experiment, exploring different game type’s impact on student engagement. Finally, in Chapter 5, we conclude our work and give recommendation for future work in Squla player modelling and adaptivity.

Chapter 2

RELATED WORK

In this chapter, we describe related work on player model analysis, game adaptations and assessments. There have been many successful cases of player modelling in digital games such as Tomb Raider (Tobias Mahlmann 2010) , Super Mario Bros (Chris Pedersen 2009) and Civilization (Pieter Spronck 2010), and in educational games as well. These models allow people to understand player preference, play style, gameplay proficiency etc. and have proved to be helpful in adapting games to improve player experience.

2.1 Adaptivity in games

Video game occupies a large portion of today's entertainment industry. Compared to movies and TV shows, games provide various ways of interaction by which we can explore the virtual world, such interaction is crucial in enhancing player engagement (Georgios . N. Yannakakis 2013).

In order to attract a larger audience and improve retention, game designers have made many attempts to make games more flexible and apply the player-centric game design. Game adaptation is one such way. Most of the video games today are relying on fixed rules such as scaling the difficulty up once a player gains certain experience points or finishes a specific quest. However, in today's video games who has such complex settings, static adaptation rules cannot suit everyone. Meanwhile, because of such complexity, more opportunities for improving player experience are brought to game developers.

In entertainment games where the fundamental goal is to make players happy, game adaptivity is typically applied on game challenge. As lots of research into the well-established flow theory points out, players are at their best when reaching a flow state (Mihaly Csikszentmihalyi 1990) (Jonova Chen 2007), games should then find the balance between player skill and gameplay challenge. Unbalanced design cause players to get bored or increase anxiety as shown in Figure 2. In the work of (Mark Prensky 2011), he concluded 12 structural game elements that make video games engaging. Among them, adaptivity serves for ensuring flow experience. He identified high adaptivity as one of the key components of successful video games.

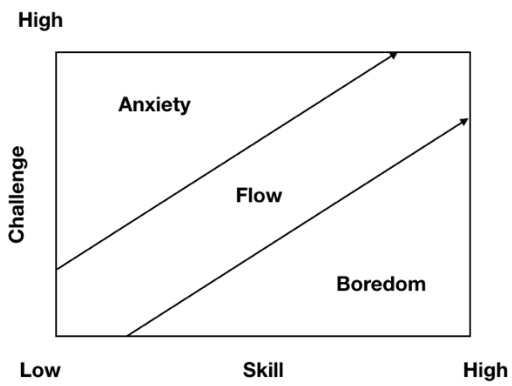


Figure 2. The Flow states of video game players.

In serious games, current research on adaptivity in serious games also focuses on challenge adjustments. Several adaptive learning systems have been developed. One such system is called AutoTutor (Arthur Graesser 2005), developed since 1994. The system adapts the responses to the questions students ask, which makes them feel like having personal tutoring. Another system called ALIGN (Neil Peirce 2008) focuses on adapting challenge level in a flow design mindset, although a significant impact on learning is not yet found, others' work focusing on balancing challenge and student skill has shown some positive results. For example, in the adaptivity research on a Spanish learning game (Sandra Sampayo-Vargas 2013), they applied adaptive difficulty. The difficulty adjustment results in a significant statistical difference in students' learning, yet no there is no evidence found supporting any effect on students' motivation, which is consistent with the previous findings on a similar topic by (Karin A. Orvis 2008). Some point out that the effect on motivation and learning can be rather content-dependent, the development of today's game technology and device bring more opportunities and worth exploring.

We usually assess the impact of adaptivity at player game experience, motivation and engagement perspective. For educational games, there is also learning involved, sometimes weighs more than just fun and joy. We can learn about player experience by direct player feedback such as self-reported level of joy and immersion, or with the help of psychological models such as Koster's theory of fun (Raph Koster 2013) and Foster's flow theory mentioned above. People have designed questionnaires evaluating player engagement, flow and skill progression. The popular Game Experience Questionnaire (Karolien Poels 2013), for example, measures player game feel through flow, immersion, competence, tension, challenge, positive effects and negative effects. The Game Engagement Questionnaire by (Jeanne H.Brockmyer 2009) sometimes considered as a part of game experience assessment, emphasizes on the flow and immersion aspect of game experience. Typical statements for measuring engagement are 'I lost track of time' and 'I play longer than I meant to'. Statements and questions regarding game feel and competence are sometimes also put into consideration because they indirectly reveal engagement level.

The assessment survey and questionnaire can come in-game or after play. In (Chris Pedersen 2009) for modelling player affective states during gameplay on Super Mario games, they put small in-game surveys from time to time to measure how players feel. The surveys give immediate feedback on how one feels during each stage of playing. However, the risk is that these surveys are being intrusive and affecting actual gameplay experience. A post-play questionnaire, on the other hand, mitigates such a problem, but lack the power of discovering in-game feelings in detail. In practice, one should decide on the choice of assessment method based on game context, research purpose and target.

2.2 Player modelling

Player modelling is the study of using computational techniques such as data classification and clustering for constructing computational models that can describe a player (Georgios N. Yannakakis 2013). Similar to the study of user experience and behavioural analysis, player modelling leads us to detect, understand and predict players' future behaviour based on their current gameplay data or survey responses, or both. Figure 3 shows the core components of player modelling.

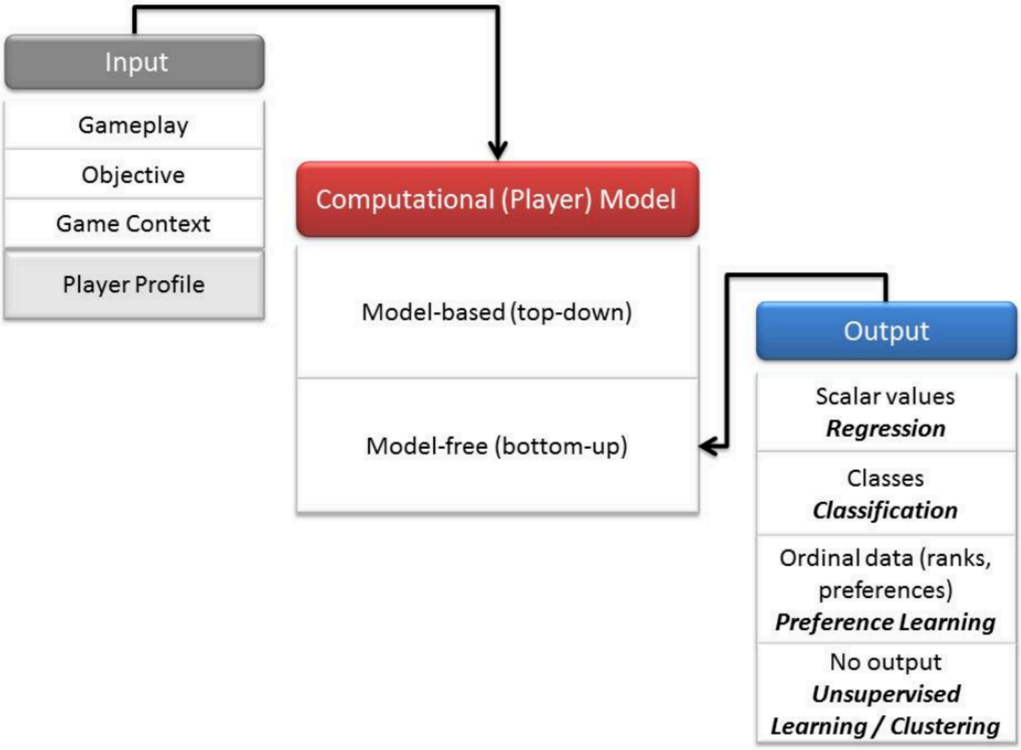


Figure 3. Player modelling: the core components. (Georgios N. Yannakakis 2013)

2.2.1 Goal

In (Georgios N. Yannakakis 2013)'s work, they identified the main goal of player modelling is to discover and specify player's behavioral, affective and cognitive patterns. Practical speaking, we are expecting to utilize the models into refining game design. With sufficient data, one may classify or cluster players using typical algorithms in this field and make assumptions about what they are expecting to experience in games. Consider playing style, for example, someone who's keen on winning may exhibit strong willing to conquer others in the game, while those who play games casually may enjoy exploring the world or interacting with non-player characters. Consider playing skill, we may divide players into novice and expert, and provide a more friendly and tolerant game setting for the former. It depends on the game of investigation which model we need, and usually, there can be multiple choices. For example, if we hope to adapt the difficulty in a puzzle game to make it more enjoyable, we can try to record player' emotional states, or understand their learning style better.

For educational games, responding to students' individual need will help with their learning. Player modelling for educational games opens the possibility for the adaptivity practice in e-learning environments.

2.2.2 Modelling target

The target of modelling is game-dependent. In general, there are player preference, player performance, player experience and player behavior. Player behavior modelling focuses on the specific action, tactics and strategy a player takes (Sander Bakkes 2012), sometimes regardless of the reason behind until it's put into game adaptation. When applying the findings into game design, there comes interpretation based on game engagement, experience and design theories. Player performance is usually directly linked to difficulty adaptation, or for player matching in multi-player games.

Player preference is another important modelling target. It is a rather large topic, as it involves play style, learning style as well as the preference on specific objects in game. In the work of (Julian Togelius 2007), the authors modelled player's driving style in a racing game, and analyse the players' fondness of different level of control. As they have found in their study, one key element relating to 'fun' is the feeling of 'almost lose control', which differs in person. They then adapt the game by inserting a different proportion of straightforward, random walk and radial mutation of the track. The challenge often lies in the proof of model accuracy and the model's correlation with player in-game behaviour.

Among the many modelling targets, the conclusiveness of player type makes it a powerful tool in directing game design. Player type model is widely applied on game world adaptation as well as difficulty scaling. It is a direct indication of player in-game preference. Similar to player personality traits, player type tells static individual differences. Yet personality directed game adaptation is rarely studied because we do not always draw the same conclusions from their self-reported personality to their in-game decisions (Sander C.J. Bakkes 2012), as people may display a completely different personality in the virtual world than in reality.

Dias et. al, for example, adapted difficulty according to the identified player type (Rodrigo Dias, 2011). The author found the key characteristics of different types of players among Conquerors, Managers, Wanderers and Participants. Winning is the uttermost important thing for Conquerors, they do not care about the storytelling such as why it is his responsibility to kill a monster. A Manager type player cares about the world's orientation and the big picture, thus storytelling matters. They can be attracted to a grand story beginning with "After the war between elves and dwarves". Wanderers and Participants care more about character story as they feel more often connected to the main character. Based on such knowledge, they developed a story manager that will select story elements from their pre-scripted story bag labelled with either 'world' or 'character'. The difficulty adaptation of the game works as follows: the hard setting is programmed for Conquerors, the normal setting for Managers and easy setting for Wanderers and Participants as they just want to explore and discover the world. (Gärtner 2009) did it in a simpler fashion, they classify their players into beginners, averagely skilled and experts, and instead of letting users choose their preferred difficulty, they perform support vector machine (SVM) to decide the difficulty level for them. Their work requires an offline and an online phase. In the offline phase, where game difficulty can be chosen, player choices and performance are recorded. The K-means clustering is performed on those data. Later in the online phase, when a new player is present, SVM is run to predict their preferred difficulty level. Similar word has been done to model player as either Hardcore player and Casual player and adapts the difficulty accordingly.

Player experience model offers another direction of game adaptation. Player experience refers to their emotional and cognitive states such as joy, boredom and frustration. Not surprisingly, lots of modelling work in this area are related to dynamic difficulty. Player's states of boredom and anxiety are both great indications of undesired challenge level (Guillaume Chanel 2008), (Hallam 2009). Fun, frustration and challenge are identified by many (Pasquier 2010), (Morelli 2012) and (Noor Shaker 2012) as the three most impacting experience states. In (Noor Shaker 2012)'s work, for example, they analysed these three different states of player and designed an automatic level generation in Super Mario Bro games. They successfully found game features that relate to the level of frustration, challenge and increase the fun. They've also pointed out some promising direction ahead, such as isolating the three factors more accurately, to allows for a method of improving engagement while reducing frustration. They also noted that expert players are easier to model in terms of engagement and challenge, suggesting modelling players experience after an initial performance modelling may be beneficial.

2.2.3 Statistical approach

To construct a player model, one needs to have sufficient input and select suitable statistical approach. The input of a player model includes static player profile, gameplay data and game context. Static profile of player contains information like gender, age, geometric data and cultural background. Gameplay preference difference has been discovered to be related to gender and age. It finds in (Bruce D. Homer 2012) that female players are more interested in puzzle games and virtual life simulation games while male players prefer FPS games and fighting games. It also finds that the females' self-confidence is related to an increased likelihood of enjoying FPS games. A static player profile like this is usually accessible to game developers. Upon registration or purchase, developers can obtain such data by letting players filling out basic information.

There are also games asking for player preferences before they start playing, these questions may include difficulty level choice, character occupation or skills. Some games also allow players to select map size and environment complexity, which gives insight about the player's proficiency and expectation at the very beginning. This kind of data is also valuable.

Game context describes the game itself. There're both static and run-time data. We do like setting a baseline on certain aspects such as the average time spent on a task and expected win rate for the current level, etc. For example, game designers may hope to set a 60% percent win rate based on psychological theories and let players spend at least 10 minutes here to complete the tasks. Some of these baselines can be updated base on the model outcome, others may be concerned with the financial expectations of the game studio thus less changeable. The real-time data is represented by a series of parameterized states, it could be about NPC responses, the number of enemies spawned, and their power / maximum health.

Gameplay data is the basis of player modelling. We state the hypothesis that the way players interact with the game reflects their gaming experience, games affect the players' cognitive processing patterns and vice versa (Georgios N. Yannakakis 2013). The original data can be rather detailed, such as all the routes taken by the player in a maze game. We need to derive attributes describing the detailed player behaviors. For example, the gameplay data in a maze game could be characterized by travel distance, the number of turns, how many times a player go back to the same path, etc. Another way is to describe player route using a heatmap, the more a certain path is visited, the hotter this area will be. The heatmap tells game designers exactly where is the most confusing, in this case, challenging part, and which area should be redesigned because it is rarely visited. It is easy to place attributes for player actions, such as describing player aiming performance in an FPS game with both firearm accuracy and headshot accuracy, counting average grenade used per game, longest shots and average heal. However, a single player action may fail to be successfully linked to the enjoyment, anxiety or engagement of a player. Thus, we need to build player strategy models. Player strategy consists of a set of game actions that leads to a certain outcome, e.g. winning a round or completing the bonus task. Player strategy is considered a high-level feature.

Both supervised and unsupervised learning can all be used on player analysis. Supervised learning, such as Markov models, decision trees, Support Vector Machine and neural networks, is frequently employed in behavioral modelling. Unsupervised learning such as k-means clustering, Linear Discriminant Analysis (LDA) can also be employed and sometimes combined with the supervised approach.

Supervised learning receives labelled training data and learns to classify future input. These learning algorithms are popular in adaptive game research. Among them, Recursive Neural Network is capable of learning complex structural features, making it ideal for player strategy modelling. The training input for this network can be, for example, a decision tree structure describing the strategy or plan of a player. The network learns from all the input plan strictures of a player, and predict future tactics when a player is handling similar tasks.

Support Vector Machine (SVM) and Factor Analysis are also proved to be effective. SVM provides data classification while Factor Analysis reveals hidden unobservable features based on observable facts. Like Principal Component Analysis (PCA), Factor Analysis can also reduce feature dimensionality, relieving us from handling numerous amounts of behavioral input, but Factor Analysis is designed to find the unobservable, while PCA's task is to list out the key factors from the feature we defined.

Unlike supervised learning, unsupervised learning receives no known class labels as input. Cluster analysis offers a way of exploring player behavioral data to derive insights. Well-known algorithms such as k-means clustering, and spectral clustering group player together based on the geometric or statistical distribution of the input. Once clusters are found, reasonable assumptions can be made according to observation on the cluster characteristics. Clustering locates the most relevant dimensions thus reduces dimensionality.

A comparison of different clustering techniques in the game World of WarCraft has been done by (Anders Drachen 2014). They selected two main behavioural features: playtime and levelling speed, and conduct k-means, non-negative matrix factorization and Principal Component Analysis to see the difference.

2.3 Conclusions

Player modelling has been applied for both entertainment games and educational games. Most of the application focuses on adapting game challenge according to player skill or playing style. To construct a player model, it is important to select the right features for analysing. These features, including player behavioral data, together with game context and player profile will all serve as input. Statistical methods such as data classification and clustering can be used to discover player behavioral or emotional patterns. Once we understand how players feel, we may adapt the game accordingly to improve their experience.

Chapter 3

PLAYER MODEL ANALYSIS

In this chapter, we discuss how we collect data, select features and process the features for building the player model.

3.1 Research Elaboration

The goal of our player model it to returns a set of value representing students' preference on different games in Squla. We believe that students with the same preference will exhibit similar behaviors, they may or may not be directly visible, but should be statistically observable.

Modelling player game type preference in Squla is challenging. First of all, though there is theoretical support that a person's personality trait has a certain definitive impact on game style and preference, in game-based learning, things are more complicated. A student may love a game if it is purely for entertainment purpose, but once a learning goal is combined, he may feel much less willing to play. The learning performance is also causing them to behave differently than playing entertainment games. The learning subject, student's skill on each of them, parental pressure, peer pressure and self-motivation, for example, are all affecting students' behavior. As students grow and learn, they may even develop a constant change in preference. Not because they are less interested in, for example, playing catapult games, but because they are putting learning on the higher priority and start considering game elements to be redundant. This is part of the reason why the current game-based learning market targets young kids more often than older students.

Therefore, we select a set of features that are most likely to reflect affection feelings towards the games. The player model analysis can be divided into two sections. It first involves selecting features related to our analytical work, then in the second phase, we analyse the data. We conduct a correlation analysis to find out the most relevant features and perform clustering to discover general game preference states of the students.

3.2 Game types

In Squla there are the following 6 major games applied in different settings. The interfaces of the games are displayed in Figure 4. Some games are subject-exclusive, others are limited to a certain education group. Puzzles and match card games appear only in lower group students. Games such as these usually serve to enhance students' memory on shapes, figures or letters, there is no fail condition designed for such games. On the other hand, catapult games and bubble games only appear for higher group students. These games are more difficult to control and more appropriate for older students who have enough experience in using digital devices.

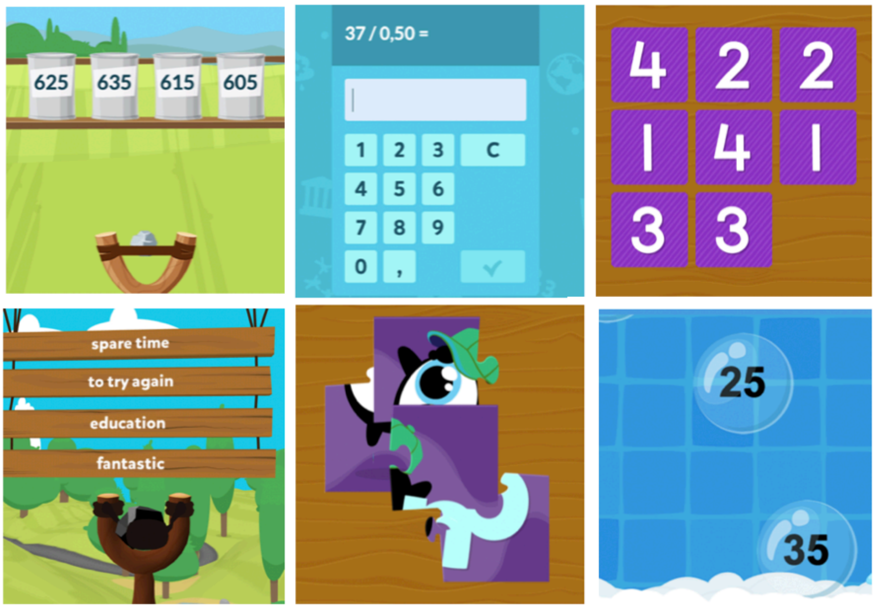


Figure 4. Different games in Squla

Below describes the major active games in Squla.

1. Match card. A few cards appear at the beginning, showing their contents. After a short while, the cards are flipped over. Player click on one card, then another. If they match, the cards disappear,

otherwise they are flipped back. Game is won after all cards are gone. Match card games are memory questions, they are not directly transformable to other games.

2. Bubble clicker games. After the sign of begin, a few bubbles containing different answers will start floating above at a relatively steady speed. They appear recursively after leaving the screen. Player click on the bubble. Game is won is the clicked bubble contains the correct answer. They can be transformed into standard single-answer selection questions.

3. Catapult board games. Player controls the catapult to shoot at the answers on the board. Game is won if he or she hit the correct one. They can be transformed into standard single-answer selection questions.

4. Catapult can games. Similar to the above, except the cans are replaced by boards.

5. Puzzles. Player drags the puzzle pieces to complete the figures. Puzzle games are not directly transformable to other games.

6. Calculator game. Player enters the answer using the calculator. This game is exclusive to math calculations. They can be transformed into open questions.

Besides these there have also been other games appeared in the past but no longer exist in platform. Some games are not yet widely applied in Squla application. There is little point in looking into them because there is not enough data for analysis.

3.3 Data collection

In this section, we discuss how we collect our data for analysis, including the decision on user scope, initial data filtering and reasons behind. We first elaborate on a few terms regarding the application and raw data that would be used later.

1. Question: a question is just by its literate meaning. In Squla, questions are displayed in a single page, upon completion, the student will go to the next one. If answered correctly, the student will gain coins and experience points. If now, this question will reappear later until at some point the completion requirement is met.

2. Quiz: a quiz is a set of questions. It's under the same subject. There are different completion rate requirement. For example, if the rate is set to 100%, students will need to complete all questions correctly. There is no fail condition of a quiz. If stopped half-way, students will begin at where they left when re-entering.

3. Play event: a play event is a database record corresponding to the completion of a question. Therefore, in a play event, we can learn about student's answer time on a single question, whether it is correct and when it is completed.

There exists irrelevant data in the database that need to be excluded, this mainly involves developer account data. We can find them by recognizing registration information such as email address.

Meanwhile, since Squla allows users to enter different education groups at their wish, we discover a small number of students accessing more than one group. Students are not forced to validate their education group upon registration. Therefore, we considered it ok if they only spend a very limited amount of time on a less-visited group. Because it's normal that they are curious and would like to explore. But if there are lots of game session records on multiple groups, we consider this to be abnormal and simply remove this user from our data set. One possibility of that happening is account sharing. We should not attempt modelling single player using such information.

Squla has a huge user base. Not all of them are put into our analysis. For an obvious reason, the old users from a few years ago, who most likely have left the app, do not provide information that is useful enough for our work. Learning contents in Squla have gone through many updates over the past few years. A single question may keep its question text, but the number appears vary. On the other hand, the game types we just discussed are also being updated over the years. One question may be presented as a standard selection question last year and transformed into a catapult game later. From the database, there is record on the last edit time of questions, but it does not state what is being modified. We could not know if it was the game type or there is a change in text and number. Luckily, the gamification of the many questions happens within a certain time range, it allows us to decide on the start date of the data we would want to gather. Most games are being updated during the first half of last year. We narrow down our target time range of researched audience to those who registered between 1st August and 30th September 2018. Data was collected until 26th March 2019. These students have access to a sufficient amount of games that are necessary for carrying on our work.

As mentioned above, not all education group students have access to the same amount of games. There is an obvious difference among different group of students. Age and learning objective are two major concerns when deciding on what game to display to students. Young kids who are still learning about letters, shapes and colours can enhance their memory with the help of completing puzzles and match card games. While older kids dealing with school test and assessment are given more selection and open questions. The education groups in Squla are the same as described in Dutch primary school education system. There are 8 groups in total, kids usually start at around 5 or 6 years old. There is no catapult games for group 1 and 2 students, and only a very small amount of bubble and puzzle games for group 1~3 and 6~8 students. The distribution of games on these groups are shown in Table 1.

Another thing to consider is the subject. Squla has over a dozen subjects, the games are not yet applied to all of them. They appear only in some subjects. Since our main goal is to understand game preference, there is no point in taking the irrelevant records into consideration. Therefore, we eventually narrow down our target group to group 4 and 5 students, focusing on subject Math, Language and English. Our target game types are catapult board game and bubble clicker games because they are the most played. And calculator games are excluded because they are exclusive to basic math open calculations questions only, and are not replaceable by other forms unless we manually transform them into, for example, a selection question. That would mean too much interference with original data.

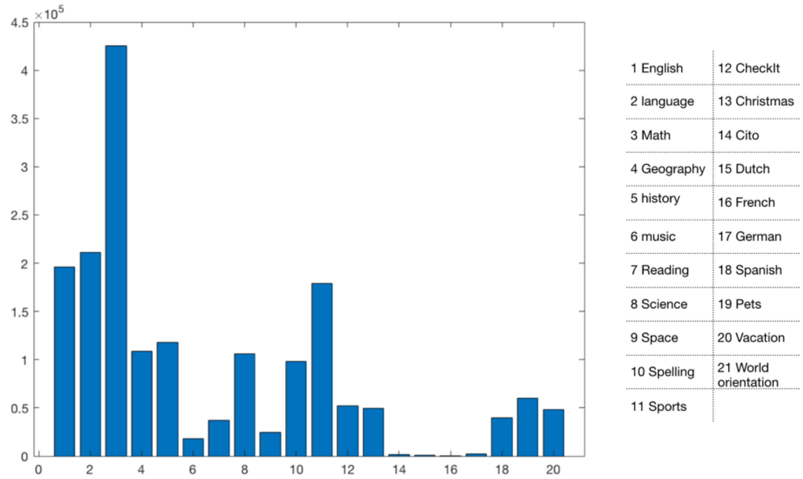


Figure 5. Questions distribution on different subjects.

We also narrow down users to those who hold a 12-month membership. In total 5,246 students are chosen, with 1,741,447 (unfiltered) game session items and over 10 million question records.

3.4 FEATURE SELECTION

In this section, we discuss which features we selected and why they are chosen. Feature selection process aims at finding the most related features and discard the irrelevant ones. The challenge here is how to decide which features are irrelevant and contribute nothing to characterizing the player. One way is to iteratively conduct clustering and remove those that do no good to final cluster results, but it will be extremely time-consuming because the dimensionality of the feature set is always high.

As introduced in Chapter 2, static player profile, in-game data and game context can all serve as input (Georgios N. Yannakakis 2013). Students' static profile includes education group, gender, preference, play style, learning style and skill. Preference here not only involves their fondness of games, but also about the current subject, difficulty of questions, etc. They are not directly observable, and there is no direct student feedback on this subject. Therefore gender and education group are the two main features.

Table 1. Game distribution in on different education groups.

Edu. Group / type	cans	boards	bubble	puzzle	memory
1	0	0	444	1157	399
2	0	0	274	701	299
3	4	2	308	1432	535
4	1885	1995	1632	27	137
5	847	1222	2346	102	129
6	172	192	43	1	20
7	75	164	27	0	3
8	54	73	21	1	8

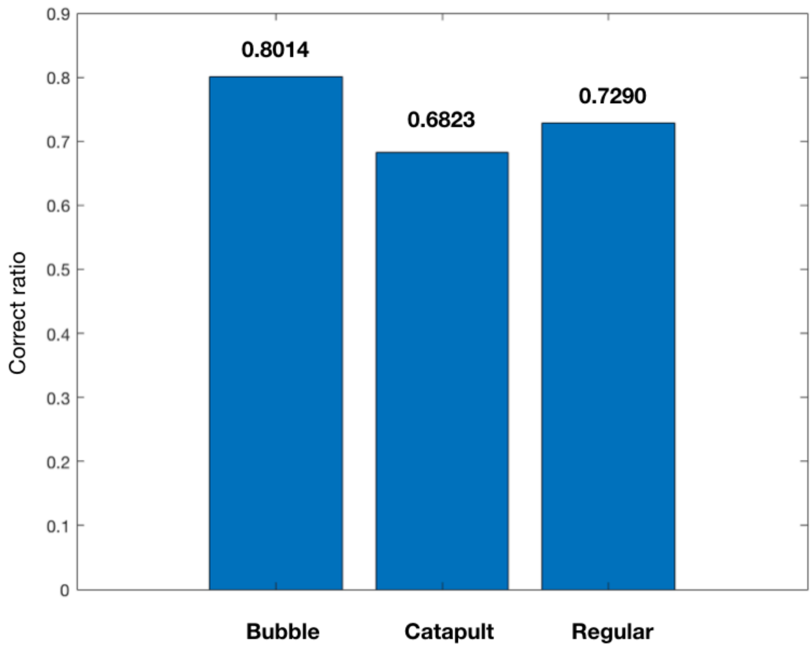


Figure 6. An initial observation of different type of quizzes' mean completion ratio.

The in-game data includes student performance and actions. The difficulty of quizzes are carefully designed to avoid boredom or frustration, at least for the majority, it is not considered in our setting. But rather, we would like to extract player skill and performance as important metrics. These features are described in Table 2. There are various external dynamics affecting student behavior as well. Such as time factor, incentives and goals from parents and teachers, peer pressure and the platform used. There are, of course, neither predictable nor controllable. Among the quiz playtime, question playtime and login playtime, the first two are too dependent on the learning content and too detailed to form a useable individual feature. If a student is not fond of a particular game, we may not see any obvious change in playtime on the current quiz, but he may be more likely to finish the learning session early than usual. Therefore, we eventually put login playtime and correct ratio as the key student performance features in our analysis.

Table 2. Base performance features.

Feature	Range	Description
Quiz playtime	/	Seconds spent on finishing a quiz.
Question playtime	/	Seconds spent on finishing a question.
Login playtime	/	Playtime for current login session
Quiz correct ratio	0 ~ 100%	Correct ratio on a quiz.

Further, we consider the following dimensions of player behavior that are most likely affected by different preference.

1. Quiz interruption. When a student is not interested in the learning content or the game itself, he or she may quit early before finishing. In a previous study on Squla users' affective states and the impacting factors, they found 54.4% students report as feeling happy who completed a quiz, while for those who didn't complete the current quiz, the ratio is only at 28.5% (Roger Smeets 2019). Therefore, we consider quiz interruption as an important metric regarding to their preference.
2. Correct ratio. As described in Chapter 2, one goal of adaptive educational game is to improve learning outcome. Correct ratio can be affected by games delivered.
3. Login Playtime. Quiz playtime on each login session.
4. Quiz repetition. When encountering a preferred game, students may want to return and play for more than once.

5. Games played. The number of games played may also affect player preference as player become more proficient in gameplay.
6. Content switching. The frequency of content switching also matters in our analysis. Students may stick to what they like for a long time, but if they are less interested, they may be annoyed and switch to another subject.

Base features need to be expanded and modified to serve our needs. In Squla, quiz playtime is not specifically designed to be at a fixed amount. It depends on the subject and the subtopics how long a quiz generally takes. Average playtime on each question is also not a good metric considering the diversity of questions. One solution is to compare playtime differences of the same quiz among different students and assign skill level based on their efficiency. Another metric regarding time is total playtime on each game session. This one is easier to handle as it doesn't concern subject difference. Correct ratio is a great measure on student learning. Since each quiz is designed to have a similar correct ratio, comparing correct ratio of quizzes in the same subject is applicable.

Eventually, our feature table is as follows:

Table 3. Final features for each student.

	Features	Range
1	Gender	0, 1
2	Educational group (group 4 and 5)	0, 1
3	Average login session playtime on regular quizzes	/
4	Average login session playtime when having bubble games	/
5	Average login session playtime when having catapult games	/
6	Average correct ratio on regular quizzes	/
7	Average correct ratio on bubble game quizzes	0~100%
8	Average correct ratio on catapult game quizzes	0~100%
9	Quiz completion ratio on regular quizzes	0~100%
10	Quiz completion ratio on bubble game quizzes	0~100%
11	Quiz completion ratio on catapult game quizzes	0~100%
12	Content switch possibility upon finishing regular quizzes	0~100%
13	Content switch possibility upon finishing bubble game quizzes	0~100%
14	Content switch possibility upon finishing catapult game quizzes	0~100%

3.5 Correlating data

From an initial mean analysis, we could already discover certain behavioral data being affected under different game delivered. Table 4 shows the average value of the features.

Table 4. Mean value of behavioural features playtime, content switch and completion ratio.

	Regular	Catapult	Bubble
Login session playtime (s)	1009	1286	1487
Content switch	0.5144	0.4874	0.3397
Completion ratio	71.31%	66.92%	87.48%

While there is no significant difference between different gender and education group, we compute the correlation between students' behavioral change with the encountered game types. For Feature 1: average correct ratio, Feature 2: completion ratio, Feature 3: content switch possibility, Feature 4: average playtime, the correlation coefficient between Regular type and Catapult games are 0.22, 0.53, 0.12, 0.25. The coefficient of Regular type / Bubble games are 0.03, 0.54, 0.51, 0.32. Finally, the correlation value between Fun type / Regular type are 0.02, 0.52, 0.41, 0.27, respectively. Completion ratio is the most impacting factor as they have a moderate correlation indicted by a coefficient between 0.5 ~ 0.7 in all cases, followed by content switching and total playtime. There is no obvious difference between different gender or educational groups.

3.6 Cluster analysis

Clustering analysis in one of the most common methods in unsupervised learning. It does not require labelling on initial data. The core idea of many clustering algorithms is to minimize the in-cluster differences while maximizing the between-class differences. In our analysis, we choose exclusive clustering. That is, each player can only belong to one group. There would be no such case where one player appears in multiple clusters. K-means clustering is a popular algorithm in player clustering. The execution steps of k-means are as follows:

1. Designate k clusters. Initially, data is randomly assigned to these clusters.

2. Calculate the mean of each cluster.

3. Calculate the distance between each data and each cluster, and reassign this data to the closest cluster

4. Repeat 2 and 3 until the clusters become stabilized.

A k-means cluster result on 2-dimensional data is shown in Figure 7.

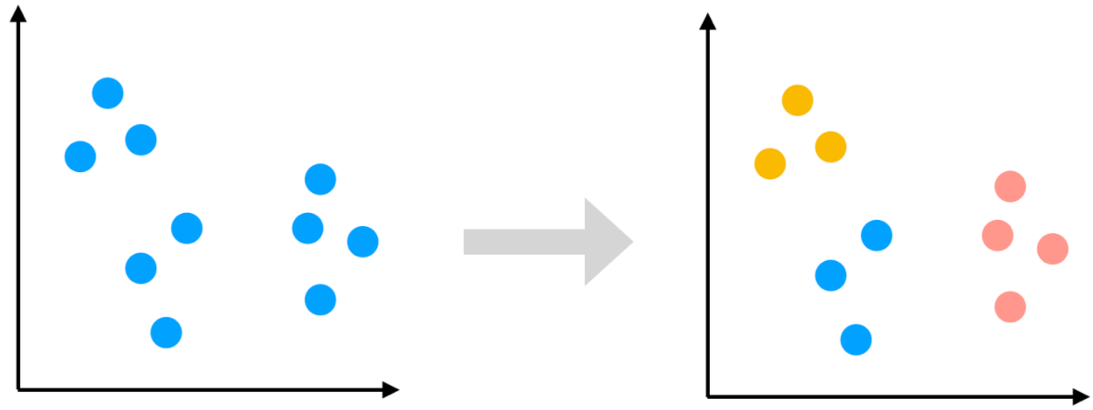


Figure 7. Clustering on 2-dimensional data

Now that we have the above most relevant feature decided, which is completion ratio, we evaluate students' behavioural patterns using cluster analysis. In our k-means analysis, we set the K value to 3, because we expect students could prefer one game over the other or display no obvious difference. The cluster centre variable values are shown in Table 5. Students assigned to cluster 3 occupies 50.92% of the analysed user group, followed by cluster 1 at 33.36% and cluster 2 at 15.72%.

We found that cluster one are students who have strong preference on bubble games and less on the other two. Cluster 2 students show rather moderate completion ratio on all games while Cluster 3 students are relatively high and still demonstrates preference on bubble games over catapults. Unlike originally expected, there is no cluster found indicating a preference on catapult games. Although such students do exist in database, they are much less compared to other groups.

Table 5. The cluster centres on feature completion ratio.

Clusters \ completion ratio	Regular games	Catapult games	Bubble games
Cluster 1	0.5564	0.5214	0.9404
Cluster 2	0.6181	0.6763	0.5115
Cluster 3	0.7592	0.8476	0.9449

3.7 Predicting preferences

Both the correlation analysis and clustering reveals behavioral difference and patterns. We identify completion ratio, content switching and playtime per session to be correlated to game type. Among them, the completion ratio shows the most significant difference, thus used in our final clustering. Looking at the first cluster, students display a strong willing to finish bubble games, but not so interested in catapult games; students in cluster 2 shows similar preference on all type of games, and on average have a moderate completion ratio. Students in cluster 3 are the have high completion ratio on all contents, with higher completion ratio on bubble games followed by catapult and then regular quizzes. The difference between game types is less compared to students in cluster 1.

3.8 Conclusions

In conclusion, we found students tend to prefer bubble games over catapult games in Squla. Factors such as content switching, playtime and quizzes played also contributes. We may quantify students' preference using completion ratio alone or compute a normalized average of multiple factors mentioned above. Here we utilize the clustering results and report a scale-based preference level, which is sufficient for our following experiment.

Chapter 4

EXPERIMENT AND RESULTS

In this chapter, we describe how we design and conduct the experiment to measure different game type's impact on engagement. We then discuss our experiment results and findings.

4.1 Overview

Since we managed to find features correlating to game type and cluster students accordingly, before any attempts on adaptivity, we conduct the following experiment to see the impact of giving a preferred game.

As mentioned in Chapter 2, engagement is one of the most important assessing targets. Besides measuring engagement, we are also interested in how students perceive their learning. Also, previous research on adaptive serious games also shown that adaptive difficulty cause students feel they are doing more moderate questions even they're not. Perhaps adaptive game type could do the same, meaning they have the ability to improve confidence and self-competence in learning.

Compared with offline experiment, where we invite students over for testing, online experiment stands out for several reasons: a) a change to access more users. b) students are in their usual learning settings, and their responses are more genuine because guest students attending an offline experiment may feel obliged to report positively, especially young kids. We conduct an online experiment. Although we cannot control the experiment environment, we believe the overall statistics are not affected by one or a few special cases.

4.1.1 Goal

Our goal for the experiment is to investigate whether offering the preferred game types based on students' preference could improve their learning experience, self-perceived competence and engagement, and to what extent is such difference if there is any. Answering these questions further opens opportunities for content adaptivity in Squla.

4.1.2 Hypothesis

Following the experiment goal, we come up with the following two hypotheses.

Hypothesis 1:

Students doing preferred games are more engaged into the learning.

Hypothesis 2:

Students doing quizzes containing preferred type questions will have better self-perceived competence.

4.1.3 The questionnaire

In testifying the hypothesis, we designed a post-play questionnaire. Following the guidance of flow theory, game engagement and experience study, we assessed students' feeling and engagement by a set of 5-point Linkert scale questions and statements. In Jeanne H. Brockmyer's game engagement questionnaire study, they provide the most direct evaluation of game engagement, including factors such as absorption, presence, flow and immersion (Jeanne H.Brockmyer 2009). There are statements like 'I lose track of where I am', 'Playing seems automatic' and 'I lose track of time'. While these allow for a much better understanding of participants' mental states, they are not suitable for all age groups as young kids participating in our experiment would have lots of problem understanding them. Therefore, we looked into other's work such as Whitton's engagement analysis that divides the measures into interest, purpose, immersion and control as well as the Game Experience Questionnaire (Karolien Poels 2013).

In the end, 11 questions regarding positive experience, negative experience, immersion, boredom, fun, tension and competence are designed. The full question list in Dutch is in Appendix A. The response on each of the questions is transformed as a 1 ~ 5 scale point. For example, the questions regarding agreement will have the following answers provided - ‘Strongly agree’, ‘Agree’, ‘Normal’, ‘Disagree’ and ‘Strongly disagree’. Except the first question asking for current feeling will receive smiley face feedback with four picture answers standing for ‘happy’, ‘confused’, ‘bored’ and ‘frustrated’. This question is displayed right after completing the math questions, giving us the most instance feedback on students’ general feelings. Squla has proved those pictures to be highly recognizable for higher education group students (Squla n.d.), thus they are accurate in telling how students feel.

Table 6. Questionnaire questions and their measuring target.

Measuring	Question
Joy	Did these quizzes make you feel happy or unhappy?
Immersion	I wish I could have played longer.
Fun	Did you find these Math questions interesting or not?
Tension / challenge	I have to put a lot of effort into answering the Math questions.
Boredom	Did you feel bored while doing the Math questions?
Annoyance	Did you feel annoyed while playing?
Challenge	Do you wish the questions would have been easier or more difficult?
Competence / challenge	Did you find the Math questions easy or difficult for you?
Competence	Did you find your performance in the Math quiz good or bad?
Competence / immersion	Did you finish the Math quiz fast or slow?

4.2 Participants

The experiment is conducted online, quiz and questionnaire are put into the app as usual learning contents through a banner channel. The students are not forced to enter or finish it, therefore we need a larger candidate group to ensure we have enough responses. Power analysis with an assumed effect size of 0,3 and power value of 0,8 is conducted. Theoretically, we would need around 111 responses. In total, we delivered the experiment to 2,000 candidates from our user group in Chapter 3. The candidates all meet the following requirements:

1. They have played those ‘fun games’ and have enough behavioral data for constructing their preference model.
2. They have played the target subject Math for a considerable time before.
3. They are still active in playing. To be exact, all participants should have at least access the platform within 4 weeks before the experiment is delivered.
4. They are from education group 5.
5. They display preference on bubble popper games over catapult games.

The candidates are randomly assigned to three different groups In the following, Group A stands for students receiving regular questions, Group B are those who have catapult games and Group C are students receiving bubble popper games.

4.3 Procedure

The experiment entry is an event banner saying ‘Complete the quiz to help us improve our app and also earn 100 coins.’ The 100 coins are used as an incentive to attract more participants, as the experiment includes a total of 26 questions, exceeding the usual question amount in regular quizzes.

Each participant starts by clicking the banner, then the math questions appear. After completing the math questions, they will see a sentence saying ‘Your 100 coins are almost there. We just need to know how you feel now’. Then the engagement questionnaire begins.

In total 16 math questions of various difficulty are selected. The original Math questions are shown in Appendix B. 6 questions are randomly selected to marked as adaptable. The math questions and selected from the database, with some minor change such as shorten the answer to fit into bubbles in our bubble clicker games. Before being put online, the questions and questionnaire items gone through several reviewing and testing stage to make sure there is no error in display, translation and functioning. All original text is in Dutch, including the math quiz, banner information and questionnaire. The experiment took two weeks starting at 13th May.

4.4 Data collection

There are two types of data for collection. One is the questionnaire feedback, the most valuable and direct response from students. The second is behavioral data, including the completion ratio, playtime and correct ratio for each group.

In total, 138 students of education group 5 in the Dutch school system responded to our experiment, 91 completed the entire task, the other 47 finished partial. Among the valid records, 32 of them are from group A, 25 from group B, 34 are from group C. In total there were 7 students attempted the task more than once, we preserve the data of their first attempt only.

On average, completing the whole task takes about 10 minutes, including 7 minutes for the Math quiz and 3 minutes for answering the questionnaire. After the Math question, students will first give emotion response on a smiley faces question, the four identified emotions are joy, confusion, boredom and frustration. After that, the 5-point Likert scale questionnaire begins.

4.5 Statistical approach

For the questionnaire, we analyse the result using Kruskal-Wallis one-way analysis of variance (or the one-way ANOVA on ranks). This is a rank-based non-parametric method often used to determine whether there is a significant difference between two or more groups of an independent variable, and it does not require the data to be normally distributed. As the distribution of data is not following the same shape, we compare the mean ranks of the 5-point Likert scale answers. The calculated H statistic is compared with critical Chi-square table under where the degree of freedom equals 2, and yield a p-value, which is the probability of H being greater than chi-square. To calculate this, *Matlab* function *kruskalwallis* is used, which outputs the H and p-value and the mean ranks.

A null hypothesis is proposed beforehand, stating there is no significant difference among the three groups regarding the variable under investigation. When calculated p-value is small enough, we say it is safe to reject the null hypothesis and states there is a significant difference among the groups. The relationship between p-value and significance levels are as follows:

When $p < 0.01$ significantly strong evidence to reject the null hypothesis
When $0.01 < p < 0.05$ strong evidence to reject the null hypothesis
When $0.05 < p < 0.10$ weak evidence to reject the null hypothesis

We then compare the correlation among data. For each item, we also calculate the mean and standard deviation.

4.6 Results

Here we discuss our results of student's responses. We refer 'regular question' to those of standard selection forms in the quiz.

Emotional feedback
Students report their emotion states immediately after the math quiz is finished, the result is shown in Figure 6.

Overall performance
Our participants are at a similar skill level, it can be validated from their past performance as well as average performance on regular questions in the quiz. Yet the correct ratio of Group B students is slightly lower than that of Group A. Most likely due to the challenge in controlling catapults. For average math skill, we guarantee fairness by randomly distribute students into the three groups, and validate this by measuring the eight standard form math questions to rule out any other affecting factors. As for correct ratio on all questions, Group A students have a correct ratio of 77.06%, while Group B students are at 64.75%. The performance here is Group C > Group A > Group B.

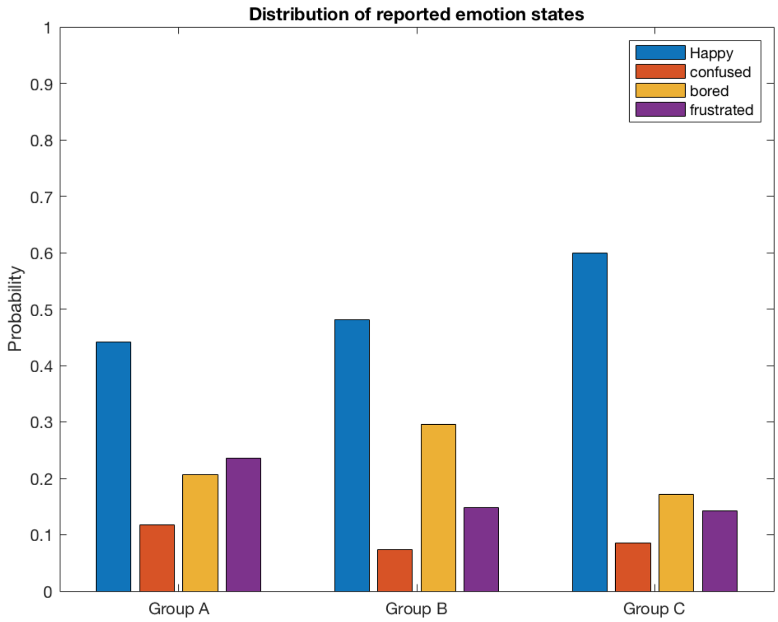


Figure 8. Results of reported emotion states.

Regarding playtime, there is a difference among the groups on those questions answered. For the regular math questions, Group A students spent on average 314.68 seconds, Group B students spent 333.52 secs and Group C students spent on average 230.38 seconds. For the entire math quiz, the difference became smaller. Group A is at an average of 517.41, Group B is at 524.11 and Group C at 481.27, displayed in Figure 9. Test shows no significant difference among the groups.

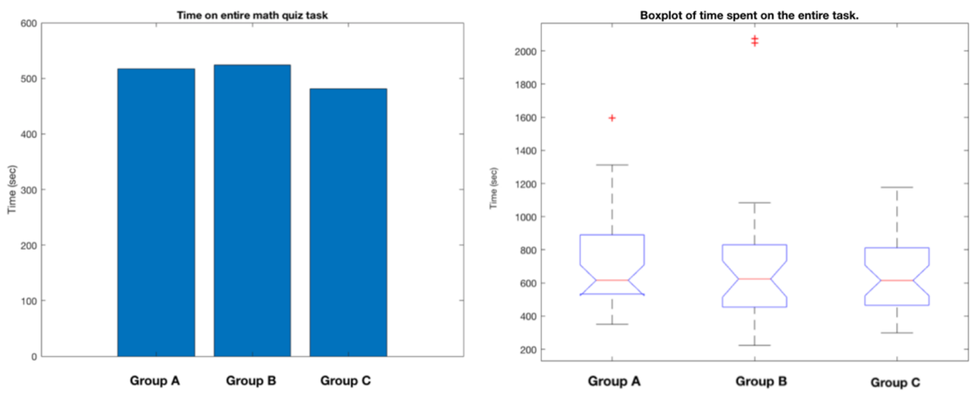


Figure 9. Students' time spent on the entire task (left) and the boxplot (right).

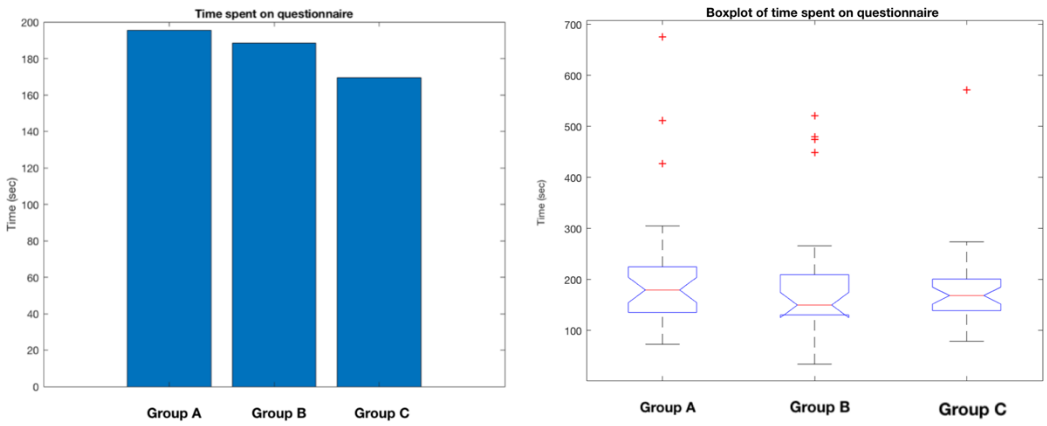


Figure 10. Students' time spent on the questionnaire (left) and the boxplot (right).

Self-reported performance

Table 7. Statistics of questionnaire item Competence-1 concerning correct ratio.

Chi-square	p	Mean-A	Mean-B	Mean-C
10.02	0.0067	3.26	2.81	3.71
		0.90	1.24	1.10

The *p* equals 0.0067, suggests a significant difference on reported performance. Judging by the mean value, we can see students playing bubble games are feeling they've done a much better job.

Table 8. Statistics of questionnaire item Competence-2 concerning efficiency.

Chi-square	p	Mean-A	Mean-B	Mean-C
4.25	0.1197	4.09	3.59	4.17
		0.93	1.19	0.82

The p equals 0.1197, suggests no significant difference on reported efficiency.

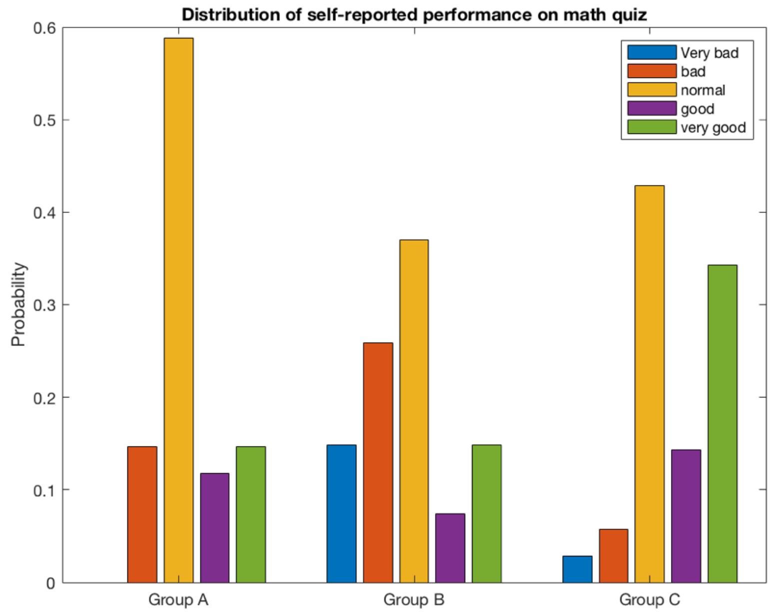


Figure 11. Students' response distribution on self-reported performance.

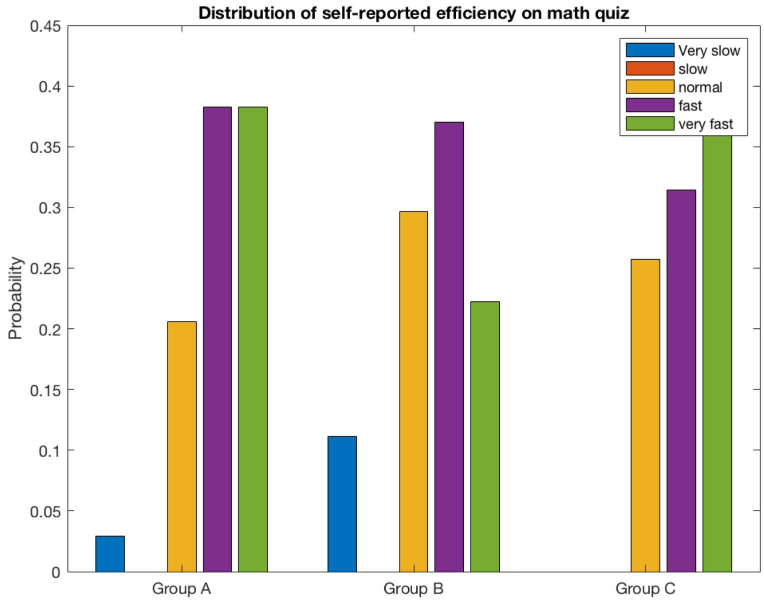


Figure 12. Students' response distribution on self-reported efficiency.

Challenge and tension

Table 9. Statistics of questionnaire item Challenge-1: math difficulty.

Chi-square	p	Mean-A	Mean-B	Mean-C
6.86	0.0324	3.71	2.89	3.63
		1.19	1.31	1.06

The p equals 0.0324, suggests a significant difference on reported math difficulty. Judging by the mean values, students playing regular games and bubble games are feeling the math to be easier, while students playing catapult games are feeling it's much harder. What's interesting here is that students in Group C did not give a higher rating than Group A like they did on performance evaluation.

Table 10. Statistics of questionnaire item Positive-1: it makes me happy.

Chi-square	p	Mean-A	Mean-B	Mean-C
6.34	0.0421	3.26	3.15	3.91
		1.29	1.46	1.04

The *p* equals 0.0421, suggests a significant difference on reported joy.

From table 7, we learnt that there is a significant difference in three groups on the subject of perceived performance. Calculated Chi-square value is 10.02, the corresponding *p*-value equals $0.0067 < 0.01$, suggesting there is a significant difference in perceived performance. This is true to their actual performance. Meanwhile, Table 9 suggests also no significant difference in efficiency, it is in accordance with total playtime, but if we break it down to regular math playtime, we see a big difference in average mean playtime. Table 10 shows the statistics of 'making me happy' questionnaire item, there is also a significant difference among the groups. The rest of the questionnaire items have some observable difference in mean value but differences are not considered siginificant in our Kruskal-Wallis test. They are shown in Appendix E.

To further investigate the relationship between the data, we calculate the correlation among the results. *Matlab* function *corrcoef* is used. It generates the correlation coefficient and the p value. The full correlation table is in Appendix D.

4.7 Discussion

Emotional feedback

Figure 11 shows the emotional feedback of students among the three groups. Although given the contents they do not enjoy, there are still more in Group B students report being happier than the control group. Since a sudden appearance of gamified contents may still surprise and entertain the students, this result is understandable. But Group B students do show a higher chance of feeling bored (the yellow bars). Group C students are generally happier, and less likely to have negative emotional feelings. This could suggest a more positive effect of the most favoured contents on students' general learning experience.

The positivity received in this question may be higher than the regular case since we are giving out 100 coins as an incentive upon completion. But this should not affect the comparison among groups.

Self-perceived competence

From students' perceived efficiency score (Table 2) we found a rather big average score difference: 3.59 in group B 4.09 in Group A and 4.17 in Group C. While their actual playtime in Group B suggests otherwise. Only a 6.7 second (+1.29%) difference was observed between A and B. Group B students feel they're much slower in answering the questions.

Correlations

Through the correlation table, we can see the high correlation between difficulty, performance and efficiency. A reduced level of tension and boredom is then correlated with a higher chance of finding the math interesting.

An unexpected result is about Question 8 asking about the feeling of happiness. Although finishing the quiz fast and good, the participants are also having a higher chance of feeling bored and annoyed during the process. As argued in emotion feedback analysis, feeling happy maybe because of the fast completion of a rewarding task, it doesn't necessarily correlate to the reduction of negative feelings during the process.

This leads to a conclusion that although completing tasks fast and well makes students feel competent and happy, a reduced challenge level may cause students to lose interest, get bored or annoyed, and not motivated to play any longer.
Challenge

The Kruskal-Wallis test shows a significant difference in challenge as well (Table 9). Both Group B and C students feel the math to be more difficult than the control group, with group B students experiencing much more challenge.

There is no significant difference among groups regarding tension, boredom and annoyance. The mean scores of boredom and annoyance for all groups are below

3, meaning students are slightly bored and annoyed while answering. We conclude that this is because our experiment is conducted in May when students are near the end of an education year, and more skilled at Math than when the year started. Also, the questions are randomly selected based on a desired correct ratio from the database with some numbers altered, some of them may seem familiar to a few students if they happened to have met them right before conducting the experiment. Both these reasons could be why they feel slightly bored and annoyed, and not interested in playing longer.

There is a significant difference among groups regarding the sense of happiness (Table 6). Group A, B and C gives a score of 3.26, 3.15 and 3.91 respectively. Students receiving favoured content are generally feeling happier. Similar conclusions can be drawn from the emotion feedback above as well. Because of the same reason in negative feeling feedback (reported boredom and annoyance), they are less motivated in playing longer and find the quiz not so interesting.

4.8 Conclusions

In this experiment, we attempted to find to what extent the desired gamified content can affect students' learning experience. In conclusion, we find different gaming contents do have a certain influence. Bubble popper receivers are generally much happier than the other two groups, while catapult gamers have much lower perceived answer efficiency. With the help of correlation analysis, we identify that gameplay challenge to be one key factor affecting player preference in this game-based learning environment.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

In this thesis project, we conduct player model analysis in the game-based learning platform Squla, targeting on game type preference students in education group 4 and 5, end explore the impact on student engagement given different games.

Going back to the research questions in Chapter 1:

- How to build a player preference model in an education game environment?
- How can such model be applied in customized game content delivery?
- What's the impact?

We build the player preference model by analysing their gameplay behavior. We found statistical differences in player general emotional feedback, and also in their actions. Putting game types as the key variable, we find the completion ration to have the highest correlation value, followed by content switching and playtime. Based on such findings, we further conduct cluster analysis to discover three main clusters. One cluster contains students who are highly focused on learning. They have a high completion ratio on all games. The second cluster is on the contrary, contains students who are less willing to play any quiz. Finally, the last cluster has students that are much more willing to finish bubble popper games than catapult games. There exists in data that certain students prefer the other way around, but they are too few in number, and not distinguishable in cluster results.

The second question asks how we can apply the model in actual use. The main idea is to deliver preferred games rather than others to students based on our model output.

As for the final question, while we do not yet have a fully functioning adaptation system online, we conduct experiment to see on which aspects the preferred games are beneficial. Bubble games, most likely preferred by most due to its convenience in control compared with catapult games, are giving students more sense of control. Based on our experience, the slight drop in catapult game correct ratio most likely suggest some students may know the answer, yet got it wrong because of the gameplay challenge. We see students quitting more often when having such games, it tells the importance of concerning gameplay challenge while design educational games. Classic player type theory and game engagement study offer valuable insight in understanding what' s attracting players and what's putting them away. Though students may be more tolerant and willing to practice if it was simply an entertaining catapult game, he may feel frustrated when it's combined with learning, especially when he could have got it right.

An interesting thing we observed is how students feel about time while given different games. Even they all have similar answer time, theirs is a big difference in their reported time feeling. This is the same as we expected. Answers to this question not only tells about students' self-perceived competence, but it is also an indication of game flow experience and immersion. In many game studies, questions such as 'I feel lost track of time' or 'I feel lost connection with the external world' are asked to measure immersion. In our case, we have to be careful about what to ask, to make sure young kids would understand, hence, no such questions are stated. But this question we have is strong evidence on students' engagement.

5.2 Recommendations

There are still much to be explored in game-based learning. Platforms like Squla has gathered a great deal of user data, there should be no doubt that valuable information about students, games and learning are hiding within it. This thesis work focuses on game type preference, we narrow down our target group students to education group 4 and 5, and found no significant difference among these two groups. However, if we could have sufficient data on the same games in, for example, group 1 and group 8 students, we may discover their change in preference. Certain external factors are also important, such as the big assessment that happens to group 7 Dutch students, may cause students in this group extremely engaged in learning, regardless of what games are being presented. The timing also matters. Understanding student's preference difference across all education groups could provide valuable information on how different games can be applied.

On the other hand, we may combine student skill, question difficulty and game type delivery all together. Squla design each quiz to be at on average of around 70% correct ratio, but students' performance varies, their expectation on their own performance also differs. One interesting game type adaptation is to put preferred games when detecting below-average performance. A possible way of application is to first find when the student play the longest and seem the happiest. For example, if a student usually spend 30 minutes using the app every day with an average correct ratio of 70%, and drop down to 10 minutes when correct ratio gets 60%. It suggests the possibility of frustration and annoyance. Inserting preferred games at this moment could help mitigate the problem. if higher correct ratio also leads to less time playing and higher quitting frequency, it's most likely suggesting boredom. We may also insert their favourite games to keep them happy. In our work, we discover gameplay challenge and sense of control as a key factor affecting game preference. Thus, we can also combine the challenge of gameplay challenge with the challenge of learning. For example, put challenging games, or increase game difficulty while detecting the possibility of boredom, such as increasing the bubbles' floating speed in bubble clicker games.

We could further combine the player type theory in our learning environment, to understand what game traits serve each student better. We mentioned that some students are exploring other education groups. There are also those who open tons of quizzes after registration and never miss any event banners. There are also students who never spend time in virtual shops, changing avatars or visiting the trophy page, only focus on learning. The former students show explorer characteristics in player typology, while the latter can be considered as goal achievers. In the future, there will be more games adding into the application, if we discover a great number of one or two type gamers in our users, we may put game features satisfying these students at a higher priority. Also, what we did not explore in this project is the game diversity. There are a small number of quizzes containing multiple games, but they are not enough for analysing individual differences. We only had an overall statistics showing a much higher completion rate on such quizzes. They are not enough for drawing solid conclusions, yet it's worth exploring how they actually affect students and its relations with player type.

In our work, we conduct experiment without considering long-term impact. Obviously, we do not yet have enough time for it. Games are proved to enhance engagement and motivation, most research conduct short experiment on games teaching certain skill or knowledge, there hasn't been much work done in long-term impact measuring. Game-based learning is about finding a balance between learning and entertainment. Ideally, we want games to be beneficial in actual learning. But as some people have brought up, emphasizing too much on entertainment may lead to worse learning performance despite having high engagement. The challenge of measuring long-term influence in a commercial game-based learning platform is that we are at the risk of having an undesired impact on students learning and losing customers. Another challenge is how we define and quantify students' learning, and also eliminate the impact of initial knowledge, it can be a rather subject-depend.

Bibliography

Anders Drachen, Christian Thurau, Rafet Sifa, Christian Bauckhage. 2014. “A Comparison of Methods for Player Clustering via Behavioral Telemetry.”

Arthur Graesser, Patrick Chipman, Brian C. Haynes, Andrew Olney. 2005. “AutoTutor: an intelligent tutoring system with mixed-initiative dialogue.” *IEEE Transactions on Education* 48(4):612–618.

Richard Bartle. 1996. “Hearts, clubs, diamonds, spades: Players who suit muds.” 06.

Brian Magerko, Carrie Heeter, Ben Medler, Joe Fitzgerald. 2008. “Intelligent Adaptation of Digital Game-Based Learning.” *Proceeding Future Play '08 Proceedings of the 2008 Conference on Future Play: Research, Play, Share* 200-203.

Bruce D. Homer, Elizabeth O. Hayward, Jonathan Frye, Jan L. Plass. 2012. “Gender and player characteristics in video game play of preadolescents.” *Computers in human behavior* 1782-1789.

Chris Bateman, Rebecca Lowenhaupt, Lennart Nacke. 2012. “Player typology in theory and practice.” *Proceedings of DiGRA 2011 Conference: Think Design Play* 05.

Chris Pedersen, Julian Togelius, Georgios N. Yannakakis. 2009. “Modeling Player Experience in Super Mario Bros.” *IEEE Symposium on Computational Intelligence and Games* 132-139.

David Plans, Davide Morelli. 2012. “Experience-driven procedural music generation for games.” *Com- putational Intelligence and AI in Games, IEEE Transactions* 192-198.

Enrica Loria, Annapaola Marconi. 2018. “Player types and player behaviors: Analyzing correlations in an on-the-field gamified system.” *CHI PLAY* 531–538.

Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, Elisabeth Andre. 2013. “Player modeling.” *Artificial and Computational Intelligence in Games* 45-55.

Georgios N. Yannakakis, John Hallam. 2009. “Real-time game adaptation for optimizing player satisfaction.” *IEEE Transactions on Computational Intelligence and AI in Games* 121–133.

Guillaume Chanel, Cyril Rebetez, Mireille Btrancourt, Thierry Pun. 2008. “Boredom, engagement and anxiety as indicators for adaptation to difficulty in games.” *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era* 13-17.

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

Jeanne H. Brockmyer, Christine M. Fox, Kathleen A.Curtiss, Evan McBroom, Kimberly M.Burkhart, Jacquelyn N. Pidruzny. 2009. “The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing.” *Journal of Experimental Social Psychology* 624-63.

Juho Hamari, Janne Tuunanen. 2014. “Player types: A meta-synthesis.” *Transactions of the Digital Games Research Association* 29–53.

Julian Togelius, Renzo de nardi, Simon Lucas. 2007. “Towards automatic personalised content creation for racing games.” *Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Games, CIG 2007* 252-259.

Karolien Poels, Yvonne A.W. de Kort, Wijnand A. IJsselsteijn, . 2013. “The Game Experience Questionnaire.” *FUGA The fun of gaming: Measuring the human experience of media enjoyment.*

Karin A. Orvis, Daniel B. Horn, James Belanich 2008. “The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames.” *Computers in Human Behavior* 24(5), 2415–2433.

Kent Norman. 2013. “(game engagement/experience questionnaire): A review of two papers.” *Interacting with Computers* 25:278–283, 06.

Kristian Kiili, Sara de Freitas, Sylvester Arnab, Timo Lainem. 2012. “The Design Principles for Flow Experience in Educational Games.” *Procedia Computer Science 1* 78-91.

Marc Prensky. 2001. “Fun, Play and Games: What Makes Games Engaging.” *Digital Game-Based Learning .*

Mihaly Csikszentmihalyi. 2008. *Flow: The Psychology of Optimal Experience.* Harper Perennial Modern Classics.

Nathan Sorenson, Philippe Pasquier. 2010. “Towards a generic framework for automated video game level creation.” 131-140.

Neil Peirce, Owen Conlan, Vincent Wade. 2008. “Adaptive Educational Games: Providing Non-invasive Personalised Learn- ing Experiences.” *Second IEEE International Conference on Digital Games and Intelligent J.*

Sykes and S. Brown. Affective Gaming: Measuring Emotion Through the Gamepad. In CHI '03 Extended Abstracts on Human factors in Computing Systems, pages 732–733. ACM, 2003. Toys Based 28-35.

Nicola Whitton. 2007. "An investigation into the potential of collaborative computer game-based learning in higher education. PhD thesis. Napier University."

Nick Yee. 2007. "Motivations for play in online games." *Cyberpsychology behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* 9:772–5, 01.

Noor Shaker, Georgios Yannakakis, Julian Togelius, Miguel Nicolau, Michael ONeill. 2012. "Evolving personalized content for super mario bros using grammatical evolution." *Proceedings of the 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 01.

Olana Missura, Thomas Gärtner. 2009. "Player Modeling for Intelligent Difficulty Adjustment." *International Conference on Discovery Science* 197-211.

Raph Koster. 2013. *Theory of Fun for Game Design.* O'reilly Media, Inc., Usa.

Ricardo Lopes, Rafael Bidarra. 2011. "Adaptivity challenges in games and simulations: A survey." *Computational Intelligence and AI in Games, IEEE Transactions* 85 – 99.

Rodrigo Dias, Carlos Martinho. 2011. "Adapting content presentation and control to player personality in videogames." *ACM International Conference Proceeding Series* 18.

Roger Smeets, Francette Broekman, Eric Bouwers. 2019. "Affect detection in home-based educational software for young children." *Educational Data Mining.*

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

Sander C.J. Bakkes, Pieter Spronck, Giel Lankveld. 2012. "Player behavioural modelling for video games." *Entertainment Computing* 3:7179, 08.

Sandra Sampayo-Vargas, Chris J. Cope, Zhen He, Graeme J. Byrne. 2013. "The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game." *Computers & Education* 452-462.

Sehar Shahzad Farooq, Jong-Woong Baek, KyungJoong Kim. 2015. "Interpreting Behaviors of Mobile Game Players from In-Game Data and

Context Logs." *IEEE Conference on Computational Intelligence and Games (CIG)* 548-549.

n.d. *Squla.* https://www.squla.nl/.

Penelope Sweetser, Peta Wyeth. 2005. "Gameflow: A model for evaluating player enjoyment in games." *Computers in Entertainment* 3:3, 07.

Pieter Spronck, Freek den Teuling. 2010. "Player modeling in civilization iv." *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment.*

Thonas Grtner, Olana Missura. 2009. "Player modeling for intelligent difficulty adjustment." *International Conference on Discovery Science* 197-211.

Tobias Mahlmann, Anders Drachen, Julian Togelius, Alessandro Canossa, Georgios Yannakakis. 2010. "Predicting player behavior in tomb raider: Underworld." *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, CIG2010* 178-185.

Veronica Sundstedt, Efstathios Stavrakis, Michael Wimmer, Erik Reinhard. 2008. "A psychophysical study of fixation behavior in a computer game." *APGV 2008 - Proceedings of the Symposium on Applied Perception in Graphics and Visualization* 08.

APPENDIX A: THE ENGAGEMENT QUESTIONNAIRE

1. Hoe voel je je?
   [picture: happy]
   [picture: confused]
   [picture: bored]
   [picture: frustrated]

2. Ik wou dat ik langer mocht spelen.
   Helemaal mee oneens
   Mee oneens
   Niet mee eens of oneens
   Mee eens
   Helemaal mee eens

3. Werd je blij of niet blij van deze quiz?
   Heel niet blij
   Niet blij
   Normaal
   Blij
   Heel blij

4. Vind je de rekenvragen makkelijk of moeilijk?
   Erg moeilijk
   Moeilijk
   Normaal
   Makkelijk
   Erg makkelijk

5. Vind je dat het goed of niet goed hebt gedaan?
   Heel slecht
   Slecht
   Normaal
   Goed
   Heel goed

6. Heb je de wiskundetoets snel of langzaam gemaakt?
   Erg snel
   Snel
   Normaal
   Langzaam
   Zeer langzaam

7. Had je liever gehad dat de rekenvragen makkelijker of moeilijker waren?
   Veel moeilijker
   Moeilijker
   Normaal
   Makkelijker
   Veel makkelijker

8. De wiskunde vragen waren vervelend.
   Helemaal mee oneens
   Mee oneens
   Niet mee eens of oneens
   Mee eens
   Helemaal mee eens

9. De wiskunde vragen waren interessant.
   Helemaal mee oneens
   Mee oneens
   Niet mee eens of oneens
   Mee eens
   Helemaal mee eens

10. Je moet heel veel moeite in stoppen om de vragen te beantwoorden.
   Helemaal mee oneens
   Mee oneens
   Niet mee eens of oneens
   Mee eens
   Helemaal mee eens

11. De wiskundevragen waren saai.

Helemaal mee oneens
Mee oneens
Niet mee eens of oneens
Mee eens
Helemaal mee eens

APPENDIX B: THE MATH QUIZ

1. Welk getal kun je delen door 7?
   A. 35 B. 25 C. 17 D. 27

2. Reken de som uit: 647 - ... = 257
   A. 390 B. 360 C. 420 D. 410

3. Kijk naar de klok. Hoe schrijf je dit als digitale tijd?

   A. 14:35 B. 13:50 C. 14:25 D. 13:25

4. Je koopt 2 ijsjes van 3 euro. Dat kost 6 euro. Daarbij koop je nog 3 ijsjes van 2 euro. Hoeveel moet je totaal betalen?
   A. 6 euro B. 10 euro C. 12 euro D. 14 euro

5. Reken de deelsom uit. 80 : 9 = ... rest ...
   A. 8 rest 8 B. 6 rest 7 C. 9 rest 5 D. 7 rest 2

6. Kijk naar het plaatje. Los de som op. Wat is het goede antwoord?

   A. 25 B. 35 C. 15 D. 30

7. Hoe schrijf je 3 minuten na middernacht op een digitale manier?
   A. 00:30 B. 00:03 C. 03:00

8. Je koopt iets voor 7,70 euro en moet gepast betalen. Je hebt twee zakjes geld bij je. In welk zakje zit precies genoeg geld? Je mag ook niets over houden.

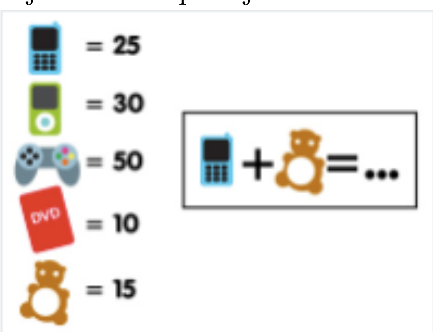
   A. in zakje 1 B. in zakje 2

9. 32 kinderen gaan varen op een meer. Er kunnen 4 kinderen in een bootje. Hoeveel bootjes zijn er nodig?
   A. 6 B. 12 C. 8 D. 9

10. Is het vroeger of later dan 9 uur?

   A. Later B. Vroeger

11. Kijk naar het plaatje en reken de som uit. Welk antwoord hoort bij de som?



A. 30 B. 40 C. 50 D. 60

12. 24 kinderen willen in de achtbaan. Er gaan 2 kinderen in 1 karretje. Hoeveel karretjes zijn er nodig?
A. 12 B. 6 C. 18 D. 10

13. Is het op deze klok overdag of 's nachts?


A. Overdag B. 's nachts

14. Bobby heeft 11 euro. Hij koopt twee ijsjes en moet 5,80 euro betalen. Hoeveel euro houdt hij over?
A. 4,80 B. 5,80 C. 4,20 D. 5,20

15. Schat de uitkomst. 380 - 301 - 10 = ... In welke doos past het antwoord?


A. doos A B. doos B C. doos C

16. Een bakker bakt 10 broden. Hij verkoopt 3 hele broden. Hij verkoopt 2 halve broden. Hoeveel broden houdt hij over?
A. 4 broden B. 5 broden C. 6 broden D. 7 broden

APPENDIX C: QUESTIONNAIRE UI

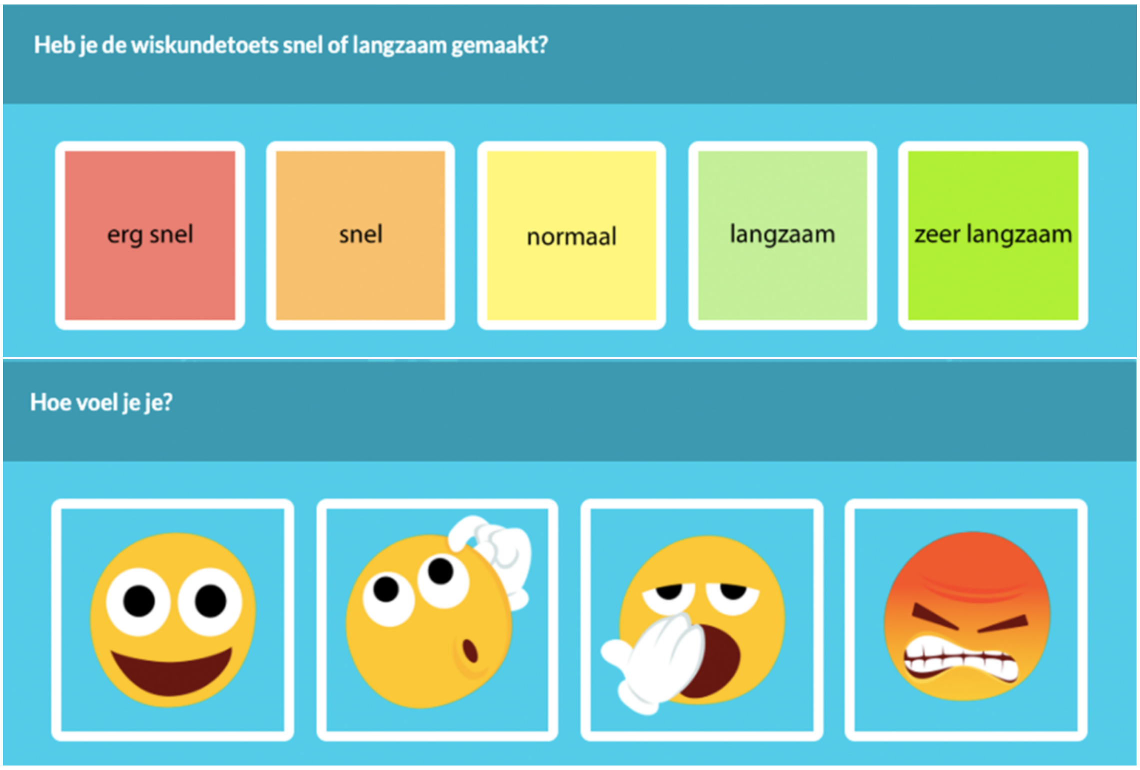


Figure 13. Example of questionnaire items with 5-point Linkert scale answers and the emotion feedback question.

APPENDIX D: CORRELATION RESULTS OF ALL QUESTIONNAIRE ITEMS

The scoring rule of each question:

Q1: I feel annoyed. (agree – low score)
Q2: I feel bored. (agree – low score)
Q3: Do you want the math quiz to be easier or more difficult? (easier – higher score)
Q4: Do you find the math quiz easy or not? (easy – higher score)
Q5: I have to put a lot of effort in the math quiz. (agree – low score)
Q6: Do you think you completed the math quiz fast or slow? (fast – high score)
Q7: Do you think you did good or bad in the math quiz? (good – high score)
Q8: Does the math quiz make you feel happy or not? (happy – high score)
Q9: Do you find the math questions interesting or not? (interesting – high score)
Q10: I wish to play longer. (agree – high score)

Table 11. Full correlation table of questionnaire results.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Q1	1.000	0.324**	-0.148	0.005	0.302**	-0.032	-0.149	-0.234*	0.157	-0.022
Q2		1.000	0.000	0.080	0.091	-0.054	-0.076	-0.349**	0.299**	0.327**
Q3			1.000	0.190	-0.140	0.109	0.138	0.089	-0.062	0.032
Q4				1.000	-0.180	0.484**	0.459**	0.179	-0.237*	-0.307**
Q5					1.000	0.038	-0.163	-0.093	-0.226*	-0.039
Q6						1.000	0.504**	0.368**	-0.238*	-0.172
Q7							1.000	0.351**	-0.326**	-0.187
Q8								1.000	-0.454**	-0.523**
Q9									1.000	0.513**
Q10										1.000

* There is a moderate correlation.
** There is a high correlation.

APPENDIX E: ADDITIONAL RESULTS AND FIGURES

Table 12. Statistics of questionnaire item Challenge-2: desired math difficulty.

Chi-square	p	Mean-A	Mean-B	Mean-C
2.44	0.2945	2.53	2.96	2.97
		1.19	1.48	1.29

Table 13. Statistics of questionnaire item Tension-1: takes considerable effort answering.

Chi-square	p	Mean-A	Mean-B	Mean-C
2.11	0.3483	3.35	3.26	2.86
		1.39	1.38	1.54

Table 14. Statistics of questionnaire item Positive-2: the math quiz is interesting.

Chi-square	p	Mean-A	Mean-B	Mean-C
0.42	0.8121	2.41	2.40	2.17
		1.35	1.42	1.18

Table 15. Statistics of questionnaire item Positive-3: I want to play longer.

Chi-square	p	Mean-A	Mean-B	Mean-C
0.71	0.7028	2.50	2.70	2.34
		1.46	1.59	1.41

Table 16. Statistics of questionnaire item Negative-1: it makes me bored.

Chi-square	p	Mean-A	Mean-B	Mean-C
2.68	0.2621	2.68	2.40	2.11
		1.43	1.50	1.25

Table 17. Statistics of questionnaire item Negative-2: it annoyed me.

Chi-square	p	Mean-A	Mean-B	Mean-C
1.17	0.5574	2.53	2.15	2.14
		1.58	1.32	1.42