Radboud University Nijmegen

Bachelor thesis in Artificial Intelligence

Testing the Effectiveness of Adaptive Educational Video Games Teaching Basic Mendelian Genetics

Author:
Maran Koolen
Student number:
S4405323

Supervisors:
Franc Grootjen
George Kachergis

July 28, 2018

Radboud Universiteit Nijmegen

Abstract

The video game industry has grown over the past few years and its growth does not show any signs of stopping any time soon. This growth is also the case for educational video games. However, most of these educational games are not based upon the great amount of research that has been done on learning techniques and which of these techniques are most effective. Furthermore, even less research has been done on how to effectively adapt the game to the performance of the player in this specific context. Consequently, a game was made based on results from research on active learning, feedback scheduling, and extrinsic and intrinsic motivation. The linear version of this game was then compared to the adaptive version of this game to see whether an adaptive, evidence-based, educational video game teaching Mendelian genetics improved learning compared to its linear counterpart. From the results it became clear that playing the adaptive version did improve the score of the players, however, it remains undecided whether this version actually improves learning.

Contents

1 Introduction

The video game industry has grown over the past few years and its growth does not show any sign of stopping any time soon ("Global games market", 2018). This is also the case for educational video games.

Many children would rather play video games than work on their education. This is why educators are interested combining the two in order to create a product that not only entertains, but also educates (Susi Johannesson, 2007).

Educational games strive for a balance between the intrinsically motivating mechanics that make video games so much fun and being useful for learning. However, most current educational games do not take into account the learning techniques we already know about. Furthermore, they have yet to be tested for their learning efficiency (Hirsh-Pasek et al., 2015). Even less research has been done on effective adaptation to player performance and knowledge (Conlan et al., 2008).

1.1 Active Learning

Active learning describes the process of having students engage in some activity that forces them to reflect upon ideas and how they are using those ideas (Revans, 1982). In our current educational system, we mainly make use of passive learning. That is, a teacher stands in front of a classroom filled with students and talks about a subject, for example Mendelian genetics. A list from a meta research by Michael (2006) included a list with some student-centred, active learning approaches. Among these approaches listed was technology-enhanced learning, an example of which are educational video games.

One of the benefits of active learning, especially its interactive factor, is that, through the use of having the player perform actions, mental processes may be activated. Especially when playing educational video games, people are forced to process material actively, practice retrieval and use metamemory (Ke; Hirsh-Pasek et al., 2015; Cairncross and Mannion, 2001). Video games also allow for distributed practice, for example, through the use of levels. Processing material actively, practicing retrieval, distributed practice, and using metamemory are all part of the four principles of memory improvement.

On the contrary, during passive learning, people must rapidly organise that is presented at a rate they cannot change. This negatively affects the learning process as, during this process, people do not process material actively, practice retrieval, but especially because this method does not allow for distributed practice. Evidence of the effectivity of active learning, especially in the form of video games, comes from research spanning multiple fields (Michael, 2006). There are suggestions (Randell et al., 1992) that some domains are more suited to be taught through the use of video games than others, but there is no evidence of this pattern in current research (Ke).

However, despite the previously mentioned benefits, active learning increases the number of actions that need to be planned and need to be held in memory. This increase in actions also causes an increases in the amount of decisions that need to be made. The combination of the increases in actions and amount of decisions introduces the possible danger of increasing the cognitive load, instead of freeing cognitive resources. Especially if the difficulty can not be adjusted (Schnotz et al., 1999). This is where the adaptive aspect of the video game would come into play.

1.2 Feedback Scheduling

One thing one has to take into account when designing an educational game is the feedback. Specifically, the type and timing of the feedback. Epstein et al. (2002) claim that immediate feedback improves learning, especially in comparison with no feedback at all. During this study, participants who were given immediate feedback were compared to participants with delayed feedback. The results were similar enough. However, the participants who had been evaluated using immediate feedback performed better on the second test, which they had to take after a delay of one day or one week. They argue that this difference exists because immediate feedback allows someone to be actively involved in the assessment process.

However, Metcalfe et al. (2009) found contradictory results during their research. They concluded that delayed feedback worked better than immediate feedback in terms of performance. It should be noted, however, that this was only the case when tested on Grade 6 children. College students showed no difference, whether they were given immediate or delayed feedback. Because of these contradictory findings, I will be using a mix of both immediate and delayed feedback in my video game.

1.3 Intrinsic Motivation

Attractive video games are video games that are intrinsically motivating. This means that, when a game is intrinsically motivating, it is more attractive to the player than a game that is less, or not, intrinsically motivating. Making the game intrinsically motivating also improves engagement of the player.

The problem with many educational games is that they are not actually fun, instead using "sugar coating" to seem fun. In short, this means that the player is motivated using gameplay for completing content. This could be a high score or praise. But this does not fall under intrinsic motivation. Instead, this would fall under the category of extrinsic motivation. And extrinsic motivation is said to undermine intrinsic motivation (Benabou Tirole, 2003). This will be discussed further in the next section.

According to the paper by Habgood and Ainsworth (2011), intrinsically motivating games embody the content within the external structure of the game, providing an external representation of the content that is taught. Furthermore, intrinsically motivating games integrate this previously mentioned content into the gameplay, making it the most fun part to play. According to this same paper, people who played the intrinsically motivating version of the game Habgood and Ainsworth provided had a higher score and more learning gains compared to the extrinsic version of the same game. Thus, intrinsic motivation appears to be favoured over extrinsic motivation.

1.4 Extrinsic Motivation

Habgood and Ainsworth (2011) argue that extrinsic motivation negatively influences the enjoyment people get from video games. Bénabou and Tirole (2003) even go as far as to say that extrinsic motivation completely undermines intrinsic motivation.

This is in disagreement with results from Hirsh-Pasek et al. (2015), as they claim that extrinsic motivation could actually provide engagement. Hirsh-Pasek et al. also claimed this would not intervene with the beneficial effect of intrinsic motivation on engagement. In fact, most structured games that work with a question-and-answer format make use of extrinsic motivation to create engagement. Examples of extrinsic motivation are motivational messages and para-social displays like, for example, earning in-game money. Players tend to be more engaged when either of these rewards are given to them during play.

1.5 Differentiation Between Learning and Fun

Hirsh-Pasek et al. (2015) are not the only ones that disagree with the conclusion of Habgood and Ainsworth (2011) that only intrinsic motivation improves learning. Kam et al. also disagree with Habgood and Ainsworth on this particular subject. They argue that it is important to keep a distinction between the educational aspects and the fun aspects of the game. Kam et al. (2008) found that certain features, like some animations, might only confuse and distract the player from the learning task. Another thing they found was that learning goals were less obvious when those goals were integrated too well into the gameplay. Lastly, Kam et al. also mentioned that these findings might not only apply to mathematics, which is the domain they used in their research, but possibly to other domains as well.

1.6 Research Question

The aim of my thesis is to answer the following question:

"Will a video game teaching the basic concepts of Mendelian genetics with an adaptive learning algorithm improve learning in comparison to the same game with a linear algorithm?"

The answer to this question can provide insight into how video games can be used to improve education, especially ones that make use of an adaptive algorithm. Its social impact can be found in the use of video games in our educational system, both as a way to reduce the workload of teachers and as a way to get students who are less motivated, be it because they do not like the subject or simply struggle with it, to practice with what they are taught during their classes anyway. As can be read in the question, the video game will teach people about the basic concepts of Mendelian genetics. This game could be used for both people who have yet to learn about Mendelian genetics and people who have learned about it a long time ago. I expect that the adaptive version of the video game improves learning in comparison to the linear version of the game.

2 Game Implementation

The goal of this game is to teach people about the basic concepts involved in Mendelian genetics. The main focus lies on homozygosity vs. heterozygosity, but the player will also come into contact with recessiveness and dominance of genes and how this, in combination with the genotype, will affect the phenotype, or looks, of the creatures.

2.1 Materials

The game was programmed using JavaScript in combination with the free open-source Phaser game framework, HTML and CSS. The game could be played at either http://maran.xymion.nl/A/ or http://maran.xymion.nl/B/, depending on which version you want to play.

2.2 Game Design

During the process of designing the game, attention was paid to story, controls and assets. In this part, these three aspects will be described and explained.

2.2.1 Story

As mentioned previously, the game is supposed to teach the player about the basic concepts of Mendelian genetics. In order to accomplish this goal, players are working with a population of fictional creatures who are on the brink of extinction of which only a certain combination of genes can save them. This is the goal players are supposed to reach. Exactly what this combination of genes is, differs between levels.

The objective of the player is always kept visible at the top of the screen, which can be seen in figure 1. This was done in accordance to results found by Kam et. al (2008). To the players, the game was introduced as a puzzle game. The reason it could be seen as such, is that finding the right combination of creatures in order to get a desired outcome could be seen as a puzzle. This takes active mental effort, as it involves strategies, thinking ahead and picking the right combinations of creatures. And, as mentioned previously, active mental effort increases learning.

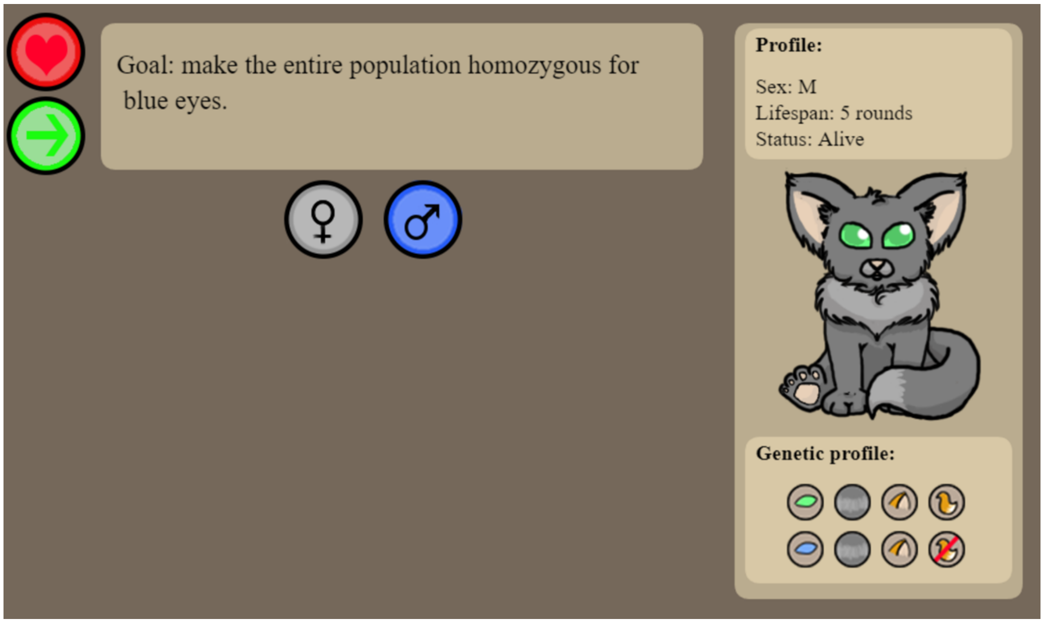

Figure 1: Example of a level.

2.2.2 Controls

Starting up the game by going to the links mentioned previously, the player is greeted by an explanation. This includes the story of the game, as can be seen in figure 2. Closing this screen has the player enter a small tutorial on how to play the game, i.e., what creatures are, how they can learn everything they need to know about them, and how to use the buttons. This is followed-up by a quick refresher course on the concepts of homozygosity, heterozygosity, and dominance and recessiveness. Adding this was done at the request of the beta testers of the game. This, and the other results of beta testing will be discussed in the section on beta testing.

After completing the tutorial, the players are greeted with their first challenge. The goal of the first level is visible at the top of the screen. Underneath this text are two creatures, one male and one female. The players can then choose to either go to the next round, by clicking the green button, or having the two creatures mate. The latter is done by selecting one male and one female and then press the red mate button. As a result of this, the female creature will get pregnant. This is visible on her profile when she is selected again. The other things that are visible on a creature's profile, which

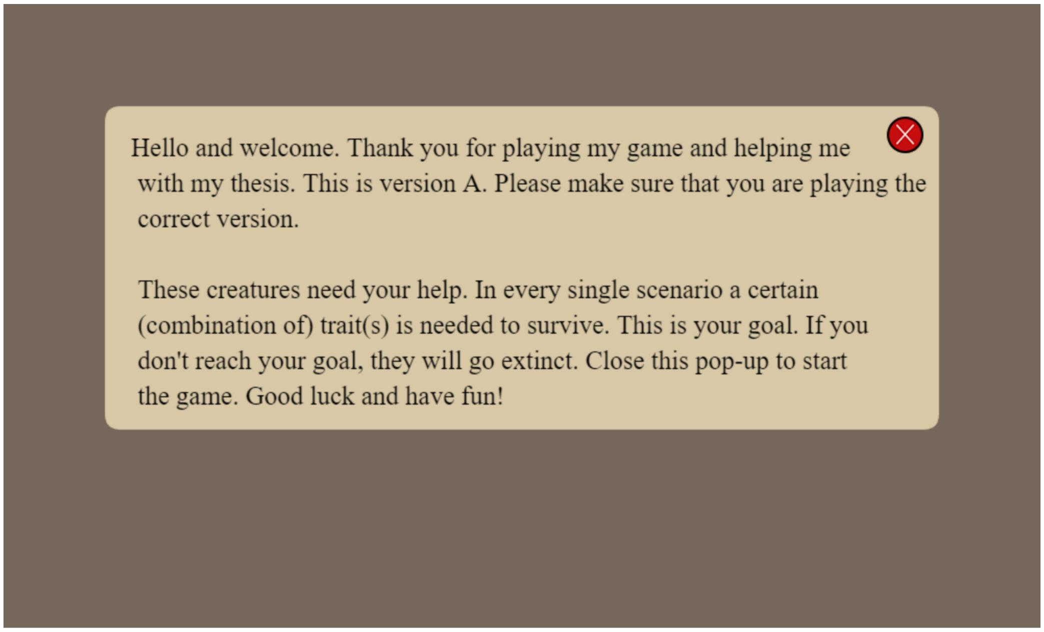

Figure 2: Welcome screen. The player is told which version they are playing. Furthermore, they are clued in on the background story.

can be seen in figure 1, are how long they will be alive for, their sex, whether they are dead or alive, a picture of the creature, and their genetic profile. When the player is done with having creatures mate, they can move on to the next round. By doing this, the creatures will have 1 round subtracted from their lifespan, effectively making them age, and pregnant creatures will give birth. Once a creature has reached a lifespan of 1, they will die the next time the player move on to the next round. This goes on until the player has either reached the goal, or has triggered one of the "lose" conditions.

A player wins when they reach the goal, meaning that the youngest creature (or group of creatures) has the trait or combination of traits specified in the goal. Losing can be done in a couple of ways. The first is that all creatures are dead. This is a losing condition because the creatures have gone extinct, the exact thing you are trying to prevent according to the story. The second losing condition is an all-male or all-female population. When a population consists of one sex only, no mating can occur. Thus, the creatures will go extinct. The last losing condition triggers when one or more of the traits specified in the goal does not occur anymore within the population. In this case, the goal cannot be reached anymore and the player would either have to go on forever of until all creatures are dead.

2.2.3 Progression of Difficulty

The levels of the game increased in difficulty as the level number went up, i.e. level 5 was more difficult than level 3. The exact difficulty of a level was decided by a couple of aspects were taken into account: whether the goal was homozygous or heterozygous, how dominant the desired allele was, how many traits the player had to work with, and whether the starting male or starting female started out with the desired trait and whether they are homozygous or heterozygous for this allele.

First, homozygosity versus heterozygosity. A goal that is homozygous would be easier to accomplish, since the player would only need to focus on one allele. There is no need to work towards an ideal pairing, since that pairing would already fulfill the requirements set by the goal. For heterozygosity, the player would first need to work towards one creature that is homozygous for the allele mentioned in the goal and one that is homozygous for any allele but the one mentioned in the goal in order to get the most effective pairing, i.e., the one that leaves up the least to chance.

Dominance level of the desired allele has more to do with how much the player has to pay attention. Having the desired allele be dominant over all other possible alleles can make it easier to spot which creatures possess this allele, since this is what will show up in the phenotype. This makes it easier to figure out which creatures to pair up for mating.

How many traits the player has to deal with also influences the difficulty. This is simply because it is easier to keep track of just one trait, then keeping track of two or three traits. If they only have to deal with one trait, the player can just focus on that one trait. They do not have to worry about other alleles disappearing from the population, which is what they would have to do if they had to deal with two or three traits.

Lastly, whether the starting male or the starting female starts out with the desired allele and whether they are homozygous or heterozygous. First, male or female. Starting out with a male with the desired allele makes the problem somewhat easier. This is because females get pregnant and being pregnant means that they cannot mate anymore during the round they got pregnant in. So, if a female has the desired allele and no other creature has it, you can only get one child that might have that allele. While, when the

Action Performed	Points
Not mating when the desired combination of alleles does not occur in the population (always)	-30
Not mating when the youngest creature(s) do not all possess the desired combination of alleles (always)	-10
Using a creature that is homozygous for the desired allele for mating (goal: homozygous)	+20 (per creature, per trait)
Using a creature that is heterozygous for the desired allele for mating (goal: homozygous)	+5 (per creature, per trait)
Desired allele does not occur in the match up used for mating (always)	-10 (per creature, per trait)
Using one creature that is homozygous for the desired allele and one that does not contain the desired allele (goal: heterozygous)	+20 (per trait)
Desired allele does not occur in the match up used for mating (always)	-10 (per creature, per trait)
Using two creatures that are both homozygous for the desired allele (goal: heterozygous)	-10 (per trait)
Creature used in the pairing possesses the allele (goal: heterozygous)	+5 (per creature, per trait)

Table 1: Actions a player can perform and how those actions influence the score.

male has the desired allele, you can go on with mating until all females in the population are pregnant. This increases the chance that you get a child that either satisfies the goal, or that at least possesses the desired allele. Whether the starting creature that possesses the allele is homozygous or heterozygous influences that chance that they pass it on to their progeny. A homozygous creature has a higher chance of passing it on than a heterozygous creature, thus making it easier.

2.2.4 Score System

Every level the player starts out with 100 points, to which performing certain actions can add points while performing others can subtract points. Most of these actions have to do with the match the player makes when they try to mate two creatures. An overview of these actions can be found in table 1.

If the goal is homozygous and the player pairs up two creatures for mating that are homozygous for both traits, 20 points get added to the score per creature per trait. If the creature is a carrier of the allele, i.e. it is heterozygous, the player gets five points per creature per trait that this situation applies to. When creatures in the match up do not carry the allele at all, 10 points get subtracted per trait per creature.

The scoring for heterozygosity works somewhat differently, as it is more complex. In this case too, when neither creature has the desired allele, 10 points are subtracted per trait per creature. When the player uses one creature that is homozygous for the desired allele and one that is homozygous for any other allele, they get 20 points added per trait for every trait this is the case for in that specific pairing. Using creature that at least contain the allele nets the player five points for each trait that this is the case for. This includes homozygous creatures. But mating with two creatures that are homozygous for the desired allele is an action that results in 10 points getting subtracted from the score.

Other actions that result in points being subtracted from the score are not mating when the desired combination of alleles does not yet occur within the population and when the youngest creature is not the one with the desired combination of alleles. These result in a subtraction of 30 and 10 points respectively. After either a win or a lose condition has been triggered, the score is calculated by dividing the points by the amount of rounds plus the amount of creatures used. The amount of creatures used is the total amount of creatures minus the two everyone starts out with and this is used to see whether or not people have been using the rounds effectively.

2.2.5 Linear Algorithm

This was version A of the game. When played, the game progressed through the list of levels with steps of two. Two was chosen so that the level between two even levels could be about the same subject, but a little bit easier. The logic behind this decision will start to make more sense during the discussion of the adaptive algorithm. The amount of levels each individual player plays is always the same when playing this version. This means that winning, losing, or the score in general do not affect the progression of the player.

2.2.6 Adaptive Algorithm

The difference between version A and version B is that version B works with the adaptive algorithm. This means that the levels the player encounters is dependent on their performance.

A player's performance is measured by the score they have achieved at the end of each level. This number then decides how many levels the player moves forward. If the player has scored 10 points of less, they only move forward one level since they did not perform that well. This is the same for the losing condition, but in this case a score of 0 is saved, and used, instead of the actual score. Did the player get a score greater than 10 but smaller than 17, they move on three levels. This was done because the player, in this case, has not performed extremely bad, which shows that they had to have some understanding of how the genetic processes work. However, they did not do extremely well either, so they get introduced to a new aspect of Mendelian genetics at a similar level to the one they just played. A score higher than or equal to 17 means that the player did well, so the difficulty can increase while, at the same time, also introducing a new aspect of Mendelian genetics. These numbers were decided upon through the use of beta testing.

Another thing that appears exclusively in the adaptive version are pop-ups. If a player takes too long, they get a hint. This hint provides an example of what genotype they should be working towards. Another pop-up attends the player to the fact that the allele desired by the goal does not occur in their match up if this is the case and asks them if they wish to continue to mate with this specific pair of creatures anyway. The timing for the former pop-up has been decided upon through beta testing, which will be discussed in the section on beta testing.

2.2.7 Asset Design

Due to the nature of the game, namely it being about genetics, I had to assign the assets myself. All of the creatures, the buttons and the symbols for the alleles were designed by me. The buttons were designed with certain associations in mind (Elliot, 2015) . Blue and a male symbol for males, pink and a female symbol for females. The next round button is green, with an arrow pointing to the right. Both are associated with moving on. Lastly, the mate button is red with a heart. This is because both are associated with either love or lust, from which the step towards sex is a small one in the minds of many people.

Figure 3: Examples of creature design.

The creatures were designed to be as cute as possible. Some examples can be seen in figure 3. Making the design as cute as possible was done in order to further motivate the player to try their best to help the population survive. According to Cho (2012) , there are several dimensions of cuteness. The first one is size. Smaller things are considered cuter than bigger things. This is why the creatures were designed to resemble some kind of smaller animal, like a cat or a rabbit. There are also no sharp edges in the designs of the creatures. This is in accordance to the second dimension of cuteness: roundness. Their simplified designs also add to the perceived cuteness, as does their plump body shapes. Lastly, keeping the lines of the drawings thick reminds people of images in children's books, which also has a positive influence on cuteness (Cho, 2012).

2.2.8 Beta Testing

As I mentioned before, I made use of beta testing for the decisions on a couple of gameplay factors. The numbers used in deciding the level progression and the timer for the hint were both based on how well the beta testers did on each level and how long it took them respectively. These beta tests were also used to test one other thing. Because, due to the nature of inheritance, there is a random factor at play. This means that, even though a player makes all the right moves, they can still lose due to bad luck. What I wanted to see is if this occurred often enough for it to be frustrating to the player. If so, preventive measures would have to be taken. However, while it did occur, it did not occur a significant amount of time. This means that the randomness factor was left in. But this conclusion was not reached based on these results alone.

People react differently to different schedules of reinforcement. Research has shown that people show a stronger reaction when the reinforcement is presented variably, so not in fixed ratios or intervals (Skinner Ferster, 1997). I would argue that the algorithm I used to program the inheritance is a form of a variable-ratio schedule, since the desired outcome occurs after an unpredictable number of responses. This is a highly effective schedule, also used in gambling (Skinner Ferste, 1997), which has also proven to be incredibly effective in video games (Nagle et al., 2014). However, this does not mean that tactics do not play a role in completing the level successfully, since there are only a small number of combinations possible in each pair of creatures and choosing the right combination of creatures increases the player's chance of getting the outcome they desire.

3 Methods

3.1 Participants

There were 47 people who joined the experiment by playing one of either versions of the game. Out of these 47 people, two had to be removed. One of them because they did not fill out the questionnaire and the other because they encountered a bug, which they mentioned in their questionnaire. The oldest of the participants was 67 years old, while the youngest was 19. This had to do with the participants having to be 18 or up in age. The average age was 36.6, with a standard deviation of 14.32 and a median of 33. Out of the participants, 23 played version B (adaptive) and 22 played version A (linear). None of them work in the field of Biology, but 3 study Biology in college and 4 took or are taking Biology-related courses in college.

3.2 Instructions

Participants were instructed to play the game in a calm environment where they felt like they could focus. They were also told to play the game in one sitting, as progression could not be saved over multiple playthroughs. Furthermore, they were made aware of the fact that both their score and time would be recorded and saved, but that the time did not have any influence on their final score. The last thing they were instructed to do was to fill out a questionnaire afterwards.This questionnaire is included as Appendix A.

3.3 Measurements

As mentioned before, the time and score was recorded for each participant for each level that they played. The questionnaire included different kinds of questions. Some questions were about personal information, like age, their experience in the field of Biology, and how much they enjoyed both video games and Biology. Then they were asked questions about the game. This included how much they enjoyed it, whether they dropped out, and, for those who played the adaptive version, if the difficulty and adaptive features were appropriate. Then they were given the opportunity to say something about the game if they felt that there was something I really needed to know. Lastly, their openness to the concept of educational video games was tested. Some of these questions were multiple choice, while others used a Likert-scale from 1 to 5.

4 Results

After testing was finished, the collected data was analysed. Participants' experience in the field of Biology was given a nominal value, depending on which answer they gave on the questionnaire. High school experience without the exam was given a 1, since Biology is a mandatory part of the curriculum during the earlier years of high school and Mendelian genetics is one of the aspect of this domain that is taught to every student. Working or having worked in the field of Biology corresponds to the most amount of experience possible, so this was given a 5.

Two one-sided t-tests were done in order to see whether the group that played the adaptive version had a higher average score per level and to see whether this group also happened to be significantly faster. The amount of dropouts per version was also investigated as a means to see whether the linear version was considered either more difficult or frustrating. Furthermore, three linear regressions were performed. One in order to predict the time needed to complete a level based on the age of the player and the version they played, another to predict the average score for the levels from the version and experience of the player, and a third one was performed in order to predict the average score on a level based on the time and version.

4.1 Performance Comparison Between Versions

In order to see whether the participants that played the adaptive version performed better than the ones that played the linear version, a one-sided test was performed on both the mean scores and mean times of the participants. The results of time can be seen in table 2 and of score in table 3. The distribution of time can be seen in figure 4 and the distribution of score in figure 5.

	Linear	Adaptive
Mean	95.05	88.15
Variance	2651.07	2193.77
N	24	23
dF	45	
T-value	0.4961	
P-value	0.31	

Table 2: T-test results from mean time.

By looking at the table it becomes clear that the mean average time per player of the adaptive version is lower than the mean average time per player of the linear version. However, it can also be seen that the one-sided t-test shows that the players of the adaptive version (M = 88.15, std = 46.83) were not significantly faster ($t(45) = 0.496$, $p > 0.05$) than the players of the linear version (M = 95.05, std = 51.49). Time also appears to be distributed normally, as can be seen in figure 4.

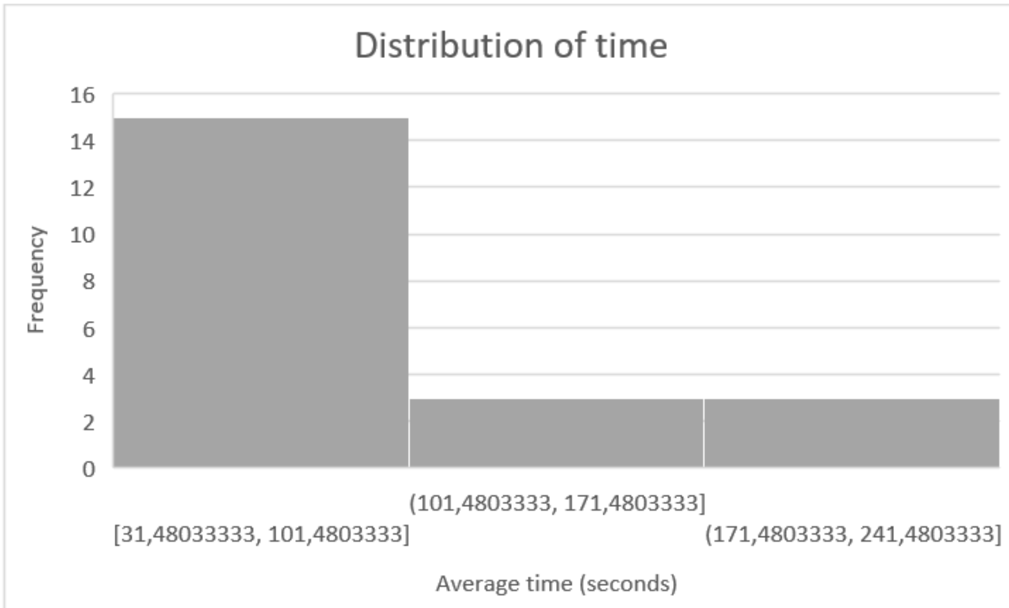

Figure 4: Distribution of average time per player.

Table 3 shows that the mean average score per player of the adaptive version is higher than the mean average score per player of the linear version. Furthermore, it can also be seen that the one-sided t-test shows that the players of the adaptive version (M = 9.38, std = 5) were not significantly faster (t(45) = -1.992, $p < 0.05$) than the players of the linear version (M = 6.67, std = 4.59). Score also appears to be distributed normally, as can be seen in figure 5.

	Linear	Adaptive
Mean	6.67	9.38
Variance	21.10	21.10
N	24	23
dF	45	
T-value	-1.99	
P-value	0.03	

Table 3: T-test results from mean score.

It should also be noted that the average amount of levels played by players

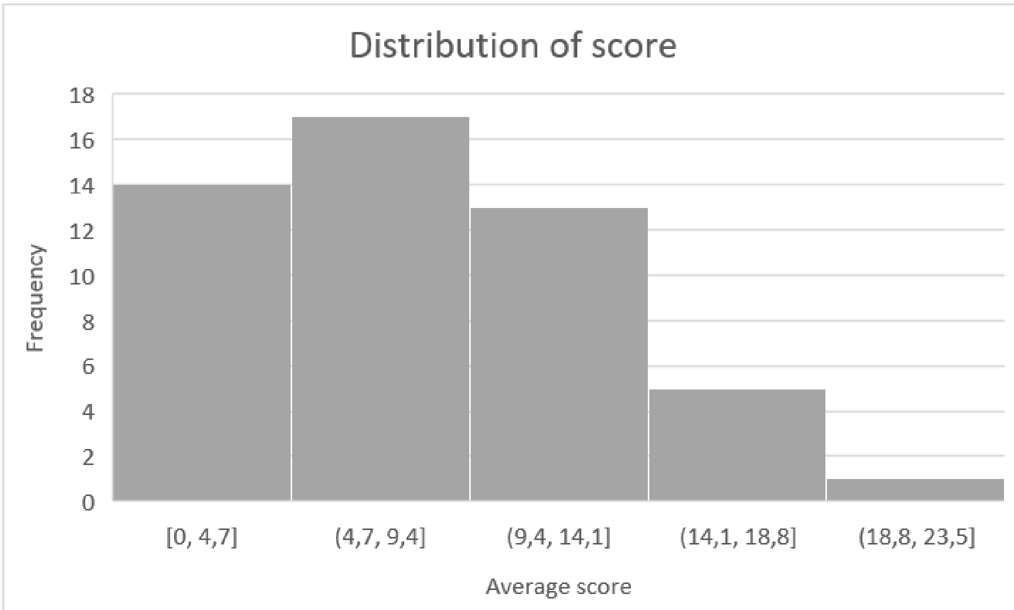

Figure 5: Distribution of mean score.

of the adaptive version was slightly higher, at 8 levels, than the average amount of levels played by those who played the linear version, which was 5.8 levels. In regards to dropouts, 14.29% of the participants who played the linear version admitted to dropping out, while 8.70% of those who played the adaptive version dropped out.

4.2 Adaptation Judgment

The participants were asked in the questionnaire whether the adaptive features were useful and whether the difficulty was appropriate. The distribution of the scores on these two aspects can be seen in figure 6. The scores for both aspects are distributed normally. In order to see if the differences between the versions were significant, I performed two one-sided t-tests.

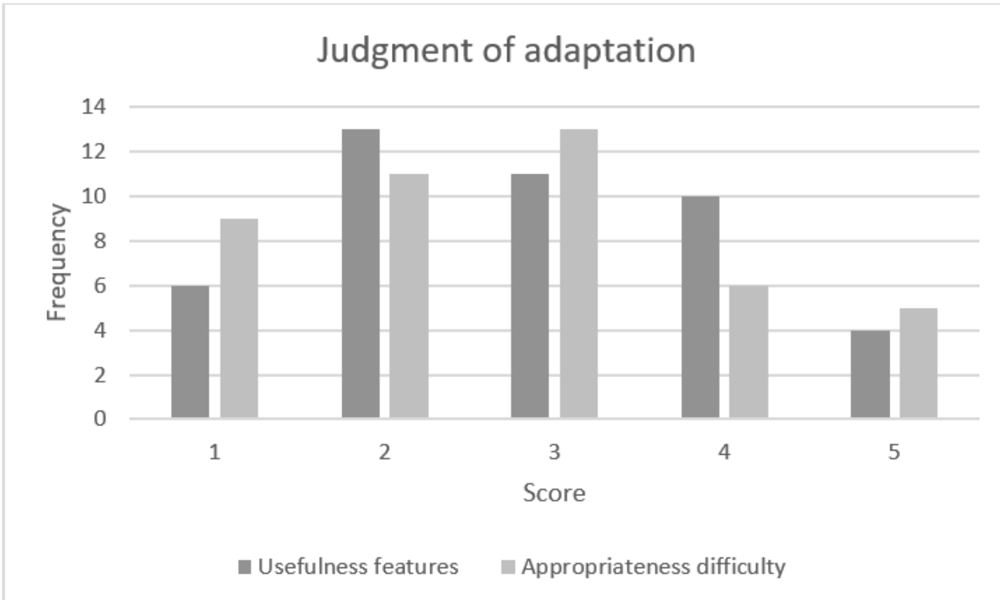

Figure 6: Distribution of feeling of adaptiveness per aspect of adaptiveness.

On average, the participants who played the adaptive version found the features to be more useful (M = 3.35, std = 0.98) than those who played the linear version (M = 2.29, std = 1.19). This difference is significant (t(43) = -3.21, $p < 0.05$). Regarding the appropriateness of the difficulty, those who played the adaptive version felt, on average, that the difficulty was more appropriate (M = 3.26, std = 1.14) than the participants who played the linear version (M = 2.095, std = 1.14). This difference is also significant (t(43) = -3.398, $p < 0.05$).

4.3 Judgment of Learning

In the questionnaire, participants were asked about if they felt like they learned something. One of the questions was about homozygosity and heterozygosity, the explicit learning goal, and the other was about recessiveness and dominance, the implicit learning goal. The resulting answers on these questions can be seen in figure 7.

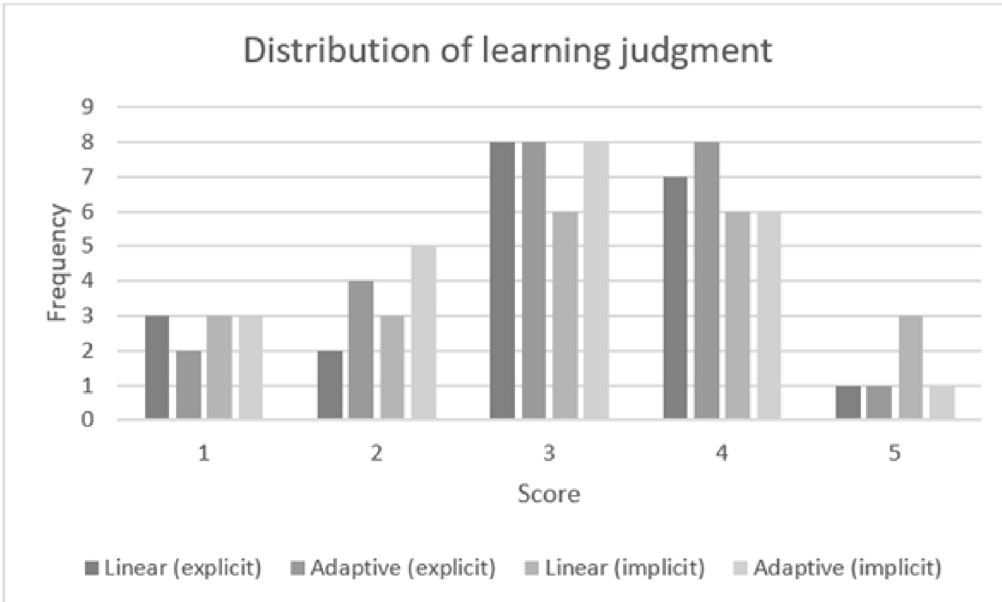

Figure 7: Distribution of learning judgment per version per kind learning.

For all groups, learning judgment is distributed normally. This is the case for both implicit and explicit learning.

4.3.1 Learning Judgment Per Version

Figure 8 shows the distribution of learning judgment for both the adaptive version and the linear version. It can be seen that these scores are distributed normally. A one-sided t-test shows that participants who played the adaptive version ($M = 2.97$, $std = 0.91$) did not score significantly higher ($t(43) = 0.31$, $p > 0.05$) than those who played the linear version ($M = 3.07$, $std = 1.09$).

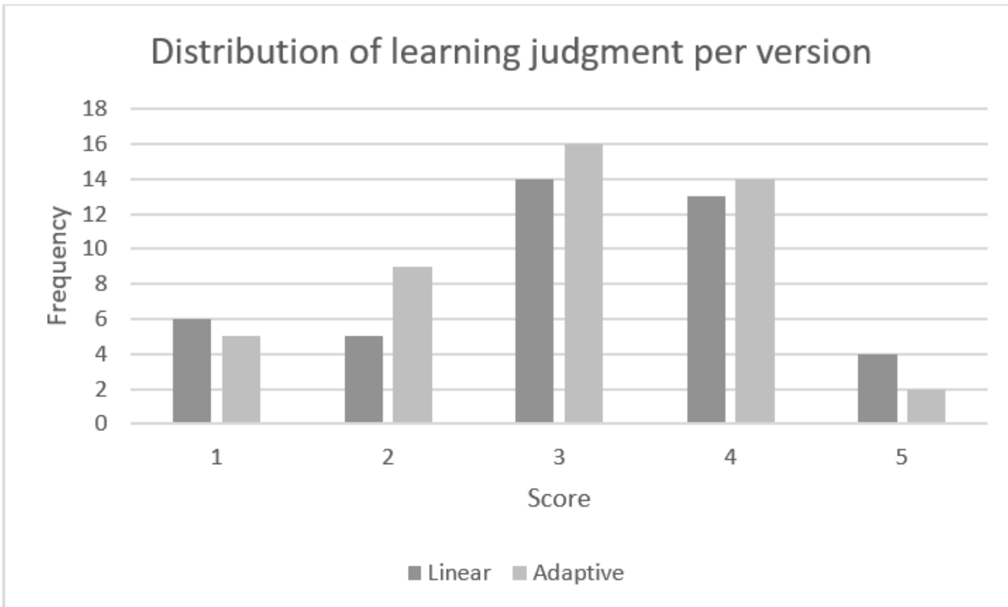

Figure 8: Distribution of learning judgment per version.

4.3.2 Implicit Learning and Explicit Learning Per Version

Figure 7 shows the results for implicit learning and explicit learning per version. A one-sided t-test showed that, on average, participants who played the linear version felt like they learned more about the implicit learning goals, recessiveness and dominance, (M = 3.14, std = 1.28) than about the the explicit learning goals, homozygosity and heterozygosity (M = 3.05, std = 1.12). This difference, however, is not significant ($t(45) = -0.26$, $p > 0.05$). In regards to the participants that played the adaptive version, they felt that they learned more about the explicit learning goals (M = 3.09, std = 1.04) compared to the implicit goals (M = 2.87, std = 1.21). Again, this is not a significant difference ($t(44) = 0.496$, $p > 0.05$). The difference between implicit (M = 3, std = 1.18) learning goals and the explicit (M = 3.07, std = 1.07) learning goals in general was not significant either ($t(88) = 0.28$, $p > 0.05$).

4.4 Enjoyment

Enjoyment was tested by having people answer how much fun they had while playing the video game. The participants who played the linear version rated their enjoyment on average (M = 3, std = 1.73) higher than those who played the adaptive version (M = 2.96, std = 1.72). However, the result of the one-sided t-test performed was not significant ($t(43) = 0.13$, $p > 0.05$). The participants also scored an average score of 3.91 on openness.

4.4.1 Enjoyment and Enjoyment of and Experience With Video Games

I performed linear regression to find out whether enjoyment of video games in general and experience with video games would influence the enjoyment score of my video game. The summary of these results can be seen in table 4.

	Coefficient	p
Intercept	2.56	
Experience video games	-0.09	0.63
Enjoyment video games	0.22	0.35

Table 4: Results from linear regression on enjoyment with enjoyment of video games and experience with video games.

The weights for both experience with video games and enjoyment of video games were not significant ($p > 0.05$). Since the p-values for both variables are not significant, a plot of the data could be misleading and is therefore not shown.

4.4.2 Enjoyment and Enjoyment of and Experience With Biology

In order to find out whether enjoyment of Biology and experience in the field of Biology influenced the enjoyment of my video game, I performed linear regression. The summary of the results of this linear regression can be seen in table 5 and figure 9.

	Coefficient	p
Intercept	2.03	
Experience in Biology	-0.28	0.16
Enjoyment of Biology	0.46	0.003

Table 5: Results from linear regression on enjoyment with enjoyment of Biology and experience in Biology.

The weight for experience in Biology was not significant ($p > 0.05$), but the weight for enjoyment of Biology was significant ($p < 0.05$). Due to this outcome, the line for experience in Biology in figure 9 might be misleading due to its weight being not significant.

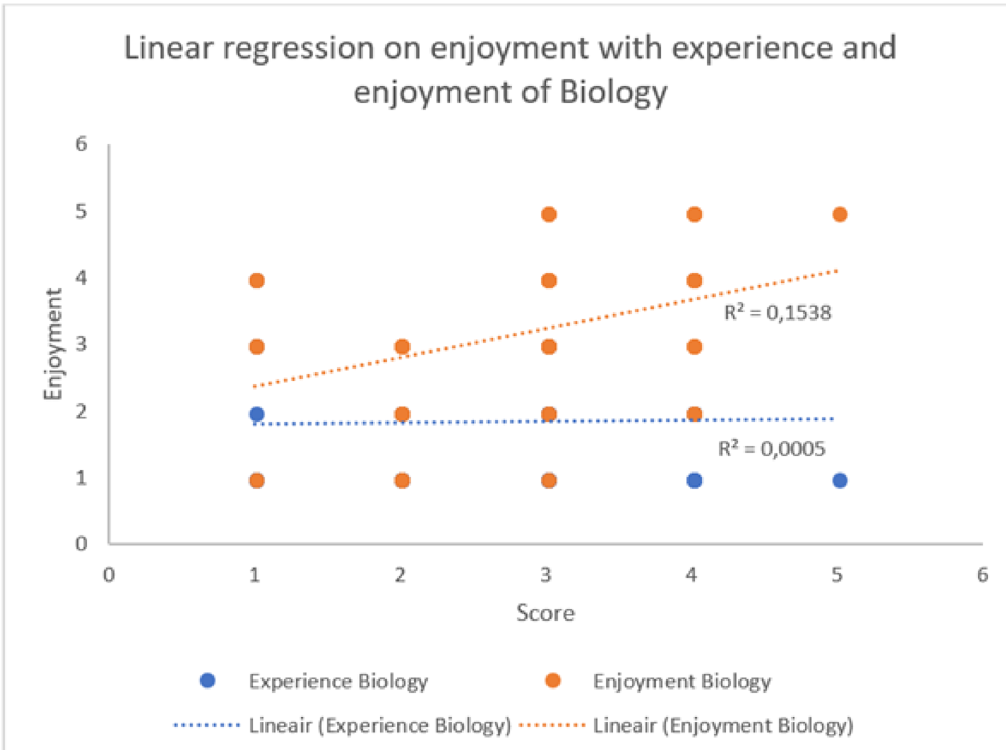

Figure 9: Linear regression on enjoyment with experience in Biology and enjoyment of Biology.

4.4.3 Enjoyment and Age

Lastly, I performed linear regression to see whether age influenced enjoyment of the video game. The summary of these results can be seen in table 6.

	Coefficient	p
Intercept	2.67	
Age	0.008	0.45

Table 6: Results from linear regression on linear regression on enjoyment with age.

Since the weight for age is not significant ($p > 0.05$), the plot of the data was not added due to the fact that it might be misleading.

4.5 Linear Regression On Mean Score With Mean Time

In order to find out whether the mean score is influenced by the mean time, I performed linear regression on these two variables. The summary of these results can be seen in table 7.

	Coefficient	p
Intercept	7.53	
Age	0.005	0.74

Table 7: Results from linear regression on mean average score with mean average time.

As the weight of mean time is not significant ($p > 0.05$), the plot is not shown. This is because a non-significant weight could make the figure misleading.

4.6 Linear Regression On Mean Score With Experience In Biology and Version

Linear regression was performed in order to find out whether experience in the field of Biology and the version could be used to predict the mean score. The results from those who did not fill in the questionnaire were removed for this test, as I did not have any information on their experience within the field op Biology. The summary of the results from this test can be found in table 8 and figure 10.

	Coefficient	p
Intercept	2.44	
Version	5.29	0.0003
Experience in Biology	-1.72	0.04

Table 8: Results from linear regression on mean average score with mean average time.

The weights for both the version and experience in Biology were significant ($p < 0.05$). In the plot you can see that the more experience someone has, the lower their mean score is.

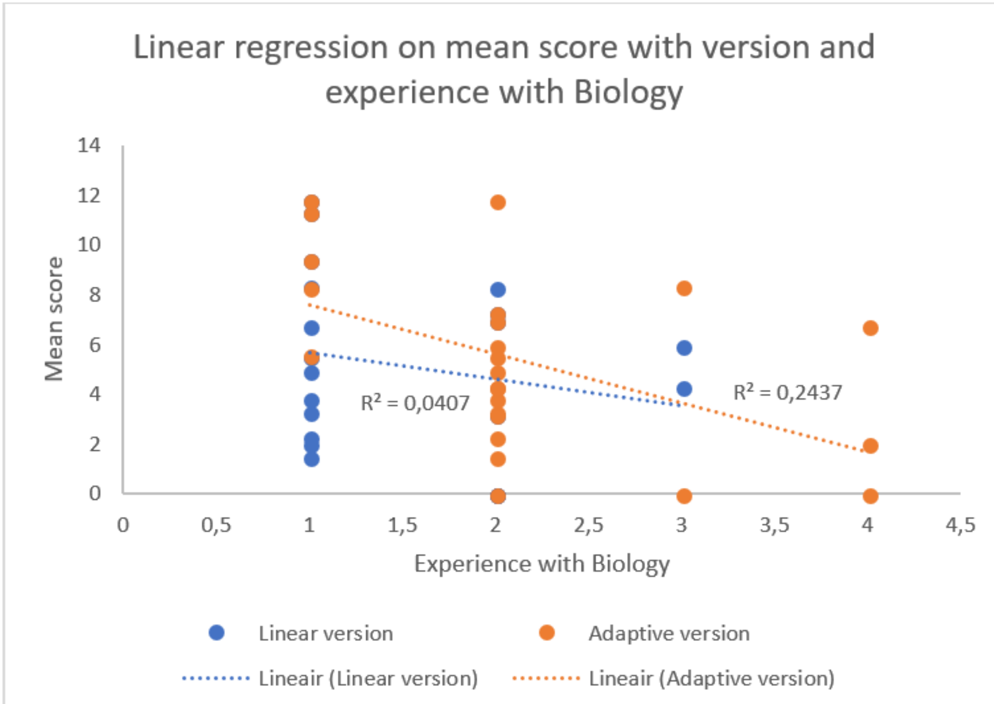

Figure 10: Linear regression on mean score with version and experience in Biology.

4.7 Linear Regression On Mean Time With Age

In order to see whether the mean time of a player can be predicted by using their age and the version they are playing, I applied linear regression. The summary of the results from this test can be seen in table 9.

	Coefficient	p
Intercept	123.35	
Version	-18.81	0.21
Age	0.07	0.89

Table 9: Results from linear regression on mean average score with mean average time.

As can be seen from the table, both the weights for version and age are not significant ($p > 0.05$). This is why the plot is not shown, as non-significant weights could mean that the plot would be misleading.

5 Conclusion

The aim of my thesis was to develop an educational adaptive video game based on evidence, which is also fun to play. In order to answer the research question: "Will a video game teaching the basic concepts of Mendelian genetics with an adaptive learning algorithm improve learning in comparison to the same game with a linear algorithm?" and to check whether or not the game was perceived as adaptive and was considered fun, multiple tests have been done. In this section, the conclusions of the tests will be discussed. At the end of this chapter, the overall conclusion will be given.

5.1 Performance Comparison Between Versions

In order to see whether participants performed better while playing the adaptive version in comparison to the linear version, two one-sided t-tests were performed. One was performed on the mean time per participant, while the other t-test was performed on the participants' mean score. The average mean score from the adaptive version was higher than the average mean score from the linear version (table 2). Furthermore, the results from the one-sided t-test on the score were significant. This means that the version does influence the score, with the adaptive version having the higher score out of the two versions.

In regards to time, those who played the linear version had a higher average time per level than those who played the adaptive version. This difference, however, was not significant, which means that the version played does not influence the average time a player spends on a particular level. It should be noted, however, that, on average, those who played the adaptive version played more levels than those who played the linear version. Since the amount of dropouts was only 5 out of the 45 who made the questionnaire, no t-test between the versions could be done as the results would not be reliable due to the amount of dropouts being too small. However, the amount of dropouts on the linear version was higher than the amount of dropouts on the adaptive version. This does imply that the linear version was more difficult, as more people gave up, but no conclusions can be drawn from these numbers with any kind of certainty.

5.2 Adaptation Judgment

The feeling of adaptation was testing using two questions on the questionnaire. The first was about whether the adaptive features were useful. Those who played the linear version should not have encountered any adaptive features and thus should not have considered them to be useful. The other question asked the participants about whether the difficulty of the levels they played was appropriate. This relates to adaptivity in that the level progression, and thus the difficulty of the levels the players encountered, depended on how well the player did. This means that those who played the adaptive version should find the difficulty of the levels they encountered more appropriate than those who played the linear version.

In regards to the adaptive versions, those who played the adaptive version of the game rated their usefulness on average higher than those who played the linear version. This difference was significant. It can thus be concluded that the adaptive features were considered useful and that they made a difference in how adaptive the game felt. That those who played the linear version did think there were some adaptive features was likely due to that fact that they did not know which version they were playing and that some pop-ups, like those about how to use the mate button and win/lose pop-ups, occurred in both versions.

Furthermore, those who played the adaptive version found, on average, that the difficulty of the levels they encountered was more appropriate than those who played the linear version. The one-sided t-test performed on these results showed that this difference is significant. This means that the version does influence the perceived appropriateness of difficulty as expected.

In conclusion, the results to both questions were deemed significant. The adaptive version score higher on both usefulness of adaptive features and appropriateness of level difficulty. It can thus be concluded that the adaptive version was perceived as being more adaptive than the linear version.

5.3 Judgment of Learning

5.3.1 Learning Judgment Per Version

The participants who played the linear version felt that, on average, they learned more than those who played the adaptive version. This was not a significant result, meaning that the version played does not influence how much the participants felt they had learned.

The direction of the difference might be explained by the fact that those who played the adaptive version played, on average, more levels than those who played the linear version. This can become demoralising to people, due to the fact that losing and doing badly are the things that increase the amount of level a player plays. And losing or doing badly can demotivate people and make them feel like they did not learn anything (Shute Ke, 2012). However, both groups still felt neutral about how much they learned, seeing as how the mean scores are approximately 3, regardless of version.

5.3.2 Implicit Learning and Explicit Learning Per Version

In the introduction the effects of implicit and explicit learning are discussed. Whether complete integration of the learning into the gameplay (implicit) or keeping the learning goals separated from the gameplay (explicit) was more effective was tested by asking people how much they felt they learned about heterozygosity and homozygosity for explicit learning and how much they felt they learned about recessiveness and dominance for implicit learning. The players of the adaptive version felt like they learned more about the explicit learning goals, while the players of the linear version felt that they learned more about the implicit learning goals. However, the results for neither version was significant, so it is likely that this difference is simply a coincidence.

That the difference between implicit learning and explicit learning is not significant in either version seems to be in agreement with the results from Kam et al. (2008), as they stated that educational video games need a mix of implicit and explicit learning.

5.4 Enjoyment

On the questionnaire, the participants were asked about how much fun they had while playing the video game. Between both versions, the linear version score higher on enjoyment than the adaptive version did, but not significantly so. The reason for this difference is similar to the reason for why judgment of learning scored higher on linear as well. On average, participants played more levels when playing the adaptive version. This, however, can be a discouraging experience. Furthermore, from the questionnaire it became clear that people experienced the gameplay to be repetitive after a few levels. This, of course, also affects enjoyment. However, as the result was not significant, the version does not influence the enjoyment.

5.4.1 Enjoyment and Enjoyment of and Experience With Video Games

When people play a lot of video games, and enjoy them, they have certain expectations when they hear the word video game. However, applying linear regression to the results showed that neither experience with video games nor enjoyment of video games affect the enjoyment of the video game I made. This could be because people knew it would be an educational game, for which the expectations regarding enjoyment are not high as they are still not considered to be very much fun, and because people knew this was not an actual video game, but a concept. Both assumptions probably affected the expectations of the participants, which, in turn, affected the score for enjoyment. This is likely the reason why people gave enjoyment a neutral score.

5.4.2 Enjoyment and Enjoyment of and Experience With Biology

I expected that, if someone does not like Biology, they are less likely to like a video game about Biology. Another expectation was that those with a lot of experience in the field of Biology might consider the game boring, due to the concepts taught being incredibly basic.

Applying linear regression to the data revealed that experience in the field of Biology does not influence the enjoyment of the video game, as the results were not significant. Enjoyment of Biology, however, does affect enjoyment of the game. The more someone enjoys Biology, the more they enjoy the video game. This is likely due to the fact that, if a video game is about a subject you enjoy and find interesting, you are more likely enjoy it more than a video game about a subject you do not enjoy or do not find interesting.

5.4.3 Enjoyment and Age

Age was another factor that could influence enjoyment, due to the fact that some people have grown up with video games and others have not. The kind of video games that those people grew up with also differs from generation to generation. This influences the expectations of the player, and thus their enjoyment when actually playing the video game.

From the results of my questionnaire it because clear that those between 35 and 61 years of age seemed to enjoy the game the most. That is why I tried to apply linear regression in order to find out if this perceived relationship was actually there. But the results were not significant, so age does not influence enjoyment.

5.5 Linear Regression On Mean Score With Mean Time

Linear regression was applied to mean score and mean time in order to find out whether the mean score could be used to predict the mean time. This, however, was not the case, as the coefficient found was not significant. This means that the mean time does not have a significant effect on the mean score.

5.6 Linear Regression On Mean Score With Experience In Biology and Version

Another linear regression was used to find out whether the mean score could be predicted by using the player's experience in the field of Biology and the version they played. Both weights were significant, so it can be concluded that both experience in the field of Biology and the version affect the mean score. It should be noted, however, that most people only had done high school level Biology, so the trend lines might not be completely accurate later on due to a lack of data involving more experience than high school level Biology. Nonetheless, it appears that the more experience someone has in the field of Biology, the lower their scores are. This is true over both the adaptive and the linear versions of the game.

5.7 Linear Regression On Mean Time With Age

The last linear regression was performed to see whether the mean could be predicted from the player's age and the version. That the version does not have a significant effect on the mean time is supported by the t-test performed on mean time. The weight of age also was not significant. Therefore, it can be concluded that mean time cannot be predicted from age and version.

5.8 Overall Conclusion

The aim of my thesis was to answer the research question: "Will a video game teaching the basic concepts of Mendelian genetics with an adaptive learning algorithm improve learning in comparison to the same game with a linear algorithm?". The video game made for this purpose was based on evidence based learning techniques and tried to make use of both intrinsically motivating and extrinsically motivating mechanisms in order to make the game fun to play. As suggested by Kam et al. (2008), both implicit learning and explicit learning were implemented in the video game. Furthermore, the game made use of both delayed and direct feedback.

The overall conclusion is that the answer to the research question remains undecided. In short, it is still unclear whether this game improved the learning of player more than its linear counterpart. The adaptive version did, however, improve the scores of the players. At the very least, people appeared to be open to idea of video games as an educational tool. how much they have learned. Further research should be done in order to find better ways to implement an effective adaptive algorithm that improves the learning of players.

6 Discussion

There are some factors that influenced the results. In this section, some of these factors will be discussed. There was also room for improvement regarding the questionnaire, which will also be discussed. Lastly, some further improvements regarding the game and the tests will be talked about.

6.1 Comments on Research

The first factor that influenced the result was that the experiment was not performed in a controlled environment. The participants could play the game at home, therefore, it was impossible to check whether they cheated or took breaks in between levels or if they got distracted. This could have been avoided by doing the experiment under my supervision. However, this approach is extremely time consuming, for both myself and the participants. Furthermore, playing at home would actually be more realistic, as educational video games will likely not be played under the all-seeing eye of a researcher or teacher.

It is also possible that time per level might not be an accurate way to measure performance, at least not between subjects. There are too many factors other than performance that influence the time of the player, including the player getting distracted, struggling with reading, planning out strategies before starting, or having trouble with their internet connection.

Another thing that should be noted is that the game did not really test learning improvement, as there was no indication of their knowledge of the assignments before the game. Furthermore, the game was in English. This means that there was a required level of mastery of the English language, which is not something every single one of the participants seemed to possess.

Lastly, at the start of the experiment there were some issues with the database. This makes the data points provided by the first 22 participants not completely reliable.

6.2 Comments on Questionnaire

On the questionnaire, the participants were given the chance to give some extra comments about the video game. Most participants were rather positive about the video game, understanding that it was just a concept.

One of the main complaints was that it lacked a “winning” feeling. This was because, one a player successfully completes a level, only a pop-up appears. A solution for this problem could be adding some sort of achievement or reward that people can display or use. In short, different kinds of feedback could be added.

Another complaint was that some parts were confusing, even after the tutorial. One part of the problem is that everything was in English, which I already discussed previously. A possible solution could be to translate the video game. Another could be to turn the tutorial in a video tutorial, where the movements are explained and showed on a screen.

Lastly, there were the times when people got a "game over" due to bad luck. A possible solution could be the option to repeat the level when people lost due to bad luck. Cases of bad luck can be divided into two categories: either the allele mentioned in the goal disappeared from the population or the player ended up getting an all-male or all-female population.

6.3 Further Improvements

This video game was based on a couple of cognitive psychological findings. However, not all of these findings could be included into the game. For example, this video game does not promote distributive learning since the player has to finish the game in one sitting. The simplest solution to this problem is the adding the ability for the player to decide when they want to quit and when they want to continue. One example of this could be a save file. This would also make it possible to split the material the game is trying to teach into pieces.

There was also no score or reward, so it was not obvious to the player how well they did. This is a form of extrinsic motivation that could be added to the game in the future. As mentioned before, the game was very basic and thus suffered from repetitive game play. The game could also be made more fun by working out the story to something more than "save these creatures from extinction".

Lastly, the genre I chose to use the "puzzle" genre for my video game. However, not everyone enjoys playing puzzle games. There are many more genres of video games that could be explored. An increase in learning has been found when people are playing a video game from a genre they have an affinity with (Steinkuehler Squire, 2014). That is why this could be interesting, especially because most current educational video games are puzzle games.

References

1. Benabou, R., Tirole, J. (2003). Intrinsic and extrinsic motivation. The review of economic studies, 70(3), 489-520.

2. Cairncross, S., Mannion, M. (2001). Interactive Multimedia and Learning: Realizing the Benefits. Innovations in Education and Teaching International,38(2), 156-164. doi:10.1080/14703290110035428

3. Cho, S. (2012). Aesthetic and Value Judgment of Neotenous Objects: Cuteness as a Design Factor and its Effects on Product Evaluation.

4. Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. Nordic Journal of Digital Literacy, 1(03), 184-214.

5. Elliot, A. J. (2015). Color and psychological functioning: A review of theoretical and empirical work. Frontiers in Psychology,6. doi:10.3389/fpsyg.2015.00368

6. Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., Brosvic, G. M. (2002). Immediate Feedback Assessment Technique Promotes Learning and Corrects Inaccurate first Responses. The Psychological Record,52(2), 187-201. doi:10.1007/bf03395423

7. Ferster, C. B., Skinner, B. F. (1997). Schedules of reinforcement. Acton Mass.: Copley Publ. Group.

8. Global Games Market Revenues 2018 — Per Region Segment. (n.d.). Retrieved June 27, 2018, from https://newzoo.com/insights/articles/global-games-market-reaches-137-9-billion-in-2018-mobile-games-take-half/

9. Habgood, M. P., Ainsworth, S. E. (2011). Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. Journal of the Learning Sciences,20(2), 169-206. doi:10.1080/10508406.2010.508029

10. Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. Psychological Science in the Public Interest, 16(1), 3-34.

11. Kam, M., Agarwal, A., Kumar, A., Lal, S., Mathur, A., Tewari, A., Canny, J. (2008). Designing e-learning games for rural children in India. Proceedings of the 7th ACM Conference on Designing Interactive Systems - DIS 08. doi:10.1145/1394445.1394452

12. Ke, F. (n.d.). A Qualitative Meta-Analysis of Computer Games as Learning Tools. Gaming and Simulations. doi:10.4018/9781609601959.ch701

13. Metcalfe, J., Kornell, N., Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. Memory Cognition,37(8), 1077-1087. doi:10.3758/mc.37.8.1077

14. Michael, J. (2006). Where's the evidence that active learning works?. Advances in physiology education, 30(4), 159-167.

15. Nagle, A., Wolf, P., Riener, R., Novak, D. (2014). The Use of Player-centered Positive Reinforcement to Schedule In-game Rewards Increases Enjoyment and Performance in a Serious Game. International Journal of Serious Games,1(4). doi:10.17083/ijsg.v1i4.47

16. Peirce, N., Conlan, O., Wade, V. (2008). Adaptive Educational Games: Providing Non-invasive Personalised Learning Experiences. 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning. doi:10.1109/digitel.2008.30

17. Randel, J. M., Morris, B. A., Wetzel, C. D., Whitehill, B. V. (1992). The Effectiveness of Games for Educational Purposes: A Review of Recent Research. Simulation Gaming,23(3), 261-276. doi:10.1177/1046878192233001

18. Revans, R. W. (1982). The origins and growth of action learning. Studentlitteratur.

19. Schnotz, W., Böckheler, J., Grzondziel, H. (1999). Individual and cooperative learning with interactive animated pictures. European Journal of Psychology of Education,14(2), 245-265. doi:10.1007/bf03172968

20. Schwan, S., Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. Learning and Instruction,14(3), 293-305. doi:10.1016/j.learninstruc.2004.06.005

21. Schwartz, B. L., Son, L. K., Kornell, N., Finn, B. (2011). Four principles of memory improvement: A guide to improving learning efficiency. IJCPS-International Journal of Creativity and Problem Solving, 21(1), 7.

22. Shute, V. J., Ke, F. (2012). Games, learning, and assessment. Assessment in game-based learning, 43-58. Springer, New York, NY.

23. Steinkuehler, C., Squire, K. (2014). Videogames and learning. Cambridge handbook of the learning sciences, 377-396.

24. Susi, T., Johannesson, M., Backlund, P. (2007). Serious games: An overview.

A Questionnaire

1. Age?

2. Which version did you play?
   O Version A
   O Version B

3. Experience with Biology?
   O High school, but didn't take the exam
   O High school and took the exam
   O I study/studied Biology in college
   O I did/am doing some Biology-related courses in college
   O I work in the field of Biology (or something related)

4. How much do you like Biology?
   1 2 3 4 5

5. How often do you play video games?
   1 2 3 4 5

6. Do you like video games?
   1 2 3 4 5

7. How much fun did you have playing the video game?
   1 2 3 4 5

8. Did you quit before completing the game?
   O Yes
   O No

9. How much did you feel like you've learned about the concepts of homozygosity and heterozygosity?
   1 2 3 4 5

10. How much did you feel like you've learned about the concepts of dominance and recessiveness?
    1 2 3 4 5

11. How useful were the pop-ups?
    1 2 3 4 5

12. Was the difficulty appropriate?
    1 2 3 4 5

13. Anything else you want to say about the video game?
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....
    ....

14. Last question: (adaptive) video games are the future of our educational
    .
    1 2 3 4 5