Game Balance by Scaling Damage:
Scaling Game Difficulty by Changing Players Damage Output

Björn Jeffsell

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden
This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Bachelor of Science in Digital Game Development. The thesis is equivalent to 10 weeks of full time studies.

Contact Information:
Author:
Björn Jeffsell

University advisor:
Olle Hilborn
School of Computing

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden

Internet : www.bth.se/com
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57
ABSTRACT

There is a lot of different kind of games which creates many different ways to balance the difficulty of games. This study will look at if damage output from a player is a good variable to scale in order to create a better balance and make the game feel more rewarding overall, based on that a game would be enjoyable if a player feels that it is rewarding to play the game. By letting both inexperienced and avid players test a part of a game with different settings for the damage output to see if the players finds the game to be more rewarding if the difficulty is set to a higher setting (lower damage output). The conclusion is that it is that damage output cannot directly affect how overall rewarding a player finds the game, but can affect other variables that in turn make the game feel more rewarding.

Keywords: Videogames, Game-balance, Difficulty scaling.
CONTENTS

Contents
Abstract........................................................................................................................................1
Contents......................................................................................................................................... ii
1 Introduction .................................................................................................................................1
2 Method.......................................................................................................................................4
  2.1 Implementation ......................................................................................................................4
  2.2 Testing ...................................................................................................................................6
3 Results .......................................................................................................................................7
  3.1 Discussion ..............................................................................................................................7
  3.2 Future work ...........................................................................................................................9
4 References .................................................................................................................................10
5 Appendix ...................................................................................................................................11
  5.1 Appendix A ............................................................................................................................11
  5.2 Appendix B ............................................................................................................................11
  5.3 Appendix C ............................................................................................................................12
  5.4 Appendix D ............................................................................................................................14
1 Introduction

A video game can be defined as a form of entertainment. People play games to be entertained in different ways. While a player is playing a game there is a lot of different aspects that might affect the players’ opinion of the game, some of these things can be: graphical interface, the background story, the input device, and, in particular game balancing as stated by Andrade et al. (2006). Balance can for example be how hard or easy the game is for the player to complete or play.

Balancing a game has proven to be important for the player to find the game satisfying, for example as illustrated by Andrade et al. (2006) when they validated their hypothesis that there is a mutual influence between game balance and user satisfaction. They were able to do this by tests where players faced different AI (artificial intelligence, in other words computer controlled) enemies in a fighting game with results showing that the AI that performed closest to the users abilities were the one that provided the highest user satisfaction.

So how much the player is challenged in the game affects the players satisfaction of the game, as shown by Juul (2009) in a study where people tested a game and then had to rate the game, where he states:

“There were indications that the players ratings were closely tied to their performance in the game, such that a player performing badly would dislike the game, a player performing fairly well would like the game, but a player performing very well would also dislike the game.”

He also states that “the most positive players were the ones that failed some and then completed the game”.

This was however a study on people who “were generally avid game players”, which somewhat fails the concept of seeing if the challenge of the game affects the players thoughts of the game, since it was only conducted on one target group. The study Andrade et al. (2006) was conducted on 50/50 between avid players and new players, and still shows that balance is important to players. But it does not in the same way show how much the players would like to be challenged as it is shown in Juul (2009) “the most positive players were the ones that failed some and then completed the game” which shows more on how much a player wants to be challenged, instead of just showing that a player wants to be challenged overall.
Another thing that separates the Andrade et al. (2006) and Juul (2009) studies is the adaptation of the learning curve, in the Andrade et al. (2006) study, players were given time to learn the game until all reached a standard where they could play the game on a minimum level, whilst in the Juul (2009) test players started the real test directly, to be able to get “insight to the players initial experience of learning a new game”.

Lankveld et al. (2010) showed that “frustration increases with difficulty”, this can be seen as a sign that players like to be challenged, but only to a certain length after which it simply becomes frustrating to play the game.

That study was however designed so that while playing on the hard difficulty setting the player should not be able to complete the game at all, which might prevent the player from getting any feeling of it being rewarding to complete, since the game was constructed to never allow the player to complete it.

There is however a big difference in how to balance difficulty in games due to the big differences of game types, for example casual games such as many social-media games, first person shooter (FPS) games such as Counter-Strike and platform games such as Super Mario.

Juul (2009) describes a few different punishments that will be a setback for players in different types of games, such as:

“Energy punishment: Loss of energy, bringing the player closer to life punishment

Life punishment: Loss of a life (or “retry”), bringing the player closer to game termination

Game Termination: Game Over

Setback punishment: Having to replay part of the game; losing abilities”

One example could be that in casual games there might not be a difficulty setting and there might not be a way to “lose” the game either, only small setbacks if you fail a certain task, for example a farming game where your tomatoes will go bad if you do not harvest them. This creates another balance, the player can only advance in the game, although it might take a longer time if you fail tasks but you will not lose anything by doing so.

In other games such as platform games there is often a bigger setback if you fail a task, such as if your character falls out of the map you will probably reset at the last checkpoint or at your last save point. This causes you to lose some of the progress you made (from the save point to the part where you fell) and could thereby be a bigger setback than to just be able to continue playing from where you were, like you would in the previous example. FPS games is often much alike in the punishment for failing, often with resets to the last save point or a previous point in the game making you replay part of the game.
So depending on what sort of game a player chooses to play, there would be different way of balancing it since some might not have a distinct way of “losing”.

These days most games (often except the ones that does not have setback), have an option for difficulty setting where the player can choose how hard the game will be. This creates a few problems since there is often only a few options such as easy, normal, and hard, with no guarantee that the player will fit into any of those categories. The problem also follows throughout the game as the player might get better and better at the game as he/she plays further, which could make the original setting wrong (provided it was right when starting to play the game).

There is also the problem of what that setting changes, it could be a lot of different things, like how well the AI behaves, how many enemies the player will face, and how much damage the player can output.

Since we cannot be sure of what affects a player in either a positive or a negative way when changing different variables, depending on the difficulty setting, this study will try to answer if damage output of the player could be one of the variables that would give a positive response from players (how overall rewarding they find the game) if changed with the difficulty setting.

This study will test both experienced and inexperienced players in a FPS game with different difficulty settings. The different settings will only change the players’ damage output towards enemies in an attempt to answer what the players’ finds to be most rewarding, based on that if a player would find a game rewarding they would also be satisfied with and/or want to play the game. The other studies that have been cited were performed on a range of different sort of games Juul (2009) used what can be seen as a platform/casual game, Andrade et al. (2006) that was conducted with a fighting/platform game, and Lankveld et al. (2010) with was is a sort of Role Playing Game (RPG). Since this study will focus on a FPS game this will provide a new perspective from than the older studies and hopefully some new conclusions.

The hypotheses are as follows:

$H_{10}$: The change in damage output depending on what difficulty the player is playing on changes nothing in regards to how overall rewarding the game is perceived.

$H_{11}$: The change in damage output (between hard and easy difficulty setting) will affect the players into feeling that the game is more rewarding.

$H_{12}$: The told difficulty level will not affect the player even if he/she is playing on a different difficulty than he/she has been told.

$H_{13}$: A player playing on an easier difficulty level than the one they have been told will find the game less rewarding.

$H_{14}$: If a player is playing on a harder difficulty setting than believed by the player (playing on a hard difficulty setting whilst believing it is set to easy) the game will seem more rewarding.
2 **Method**

The test players were given instructions on how to play a game as well as information where it was clearly marked what difficulty level they would be playing the game on.

After they had completed the game (map) they were given a form with a few different questions as well as a few statements that they filled in how much they agreed with.

2.1 **Implementation**

The map for the study was created for a modified version of the game “Half-Life 2, Episode 2” made by Valve\(^1\) which contains much of what could be considered standard when it comes to FPS games in terms of controls, different selection of enemies and so forth.

The game also contains a free level editor as well as tools to create modifications of the game, which made it a good candidate for creating the experiment, one of the big advantages is also its ability to output the logs of a played game making it easier to collect the data generated by the test players.

The game made it very easy to change the players´ damage output as it uses a file with plain text format to do so. The players were constricted to using a standard pistol for the game, so that all would start with the same weapon and being unable to pick up any other weapons during the course of the test, to make sure all the players were playing on the same conditions.

However since the study was only supposed to change the players damage output towards enemies it was easier to reverse it, changing the enemies health rather than the pistols damage.

The main reason for reversing it was to avoid changing the players’ damage towards physics object that can be present throughout the map such as crates and barrels, some of which can be destroyed by shooting at them, others that simply get affected by physics when shooting at them (indestructible).

Since there were a few different sorts of enemies the player would be faced with, the health was changed by a set percent of 300% to make all enemies equally much harder to defeat between the change of the easy difficulty level and the hard.
The percentage was decided based on tests playing through the game, both by experienced players and inexperienced players (who would otherwise have been participating in the “real” tests). These results were not included in the study since the game was played several times by the same persons and changes were sometimes made between them. The first tests were to see if an inexperienced player could go through the map without too much difficulty (on easy settings). When it was established that a player could reach the end, the tests for a good percent to use for raising the difficulty to hard started, with 300% looking like a good candidate after a few tests. It was established that an experienced player would have some difficulty completing the game but should be able to without too much trouble.

Since there would be a lot of different sort of players playing the game (both novice players and experienced) the map was designed to give a smaller challenge in the beginning with one smaller room before the main area (room) of the map to give the players a sort of informal introduction to the game, this was not part of the study other than a choice of level design. Even though the smaller room had only a few enemies, the settings on some of them were supposed to challenge the player by rushing towards them, making the player aware that there were enemies and what most of them would look like (humanoid).

The data-log of the test consisted of a text document with info on:

- How many shots that was fired.
- How many that missed/hit.
- How many enemies that were killed.
- How many times the player died.

Since the gathered data was logged to a text file of events with one line of text for each event, for instance “Player fired a shot that missed”, a small program that compiled the log and converted it to numbers was created.

The time of the test were also included, with the play-test being limited to a maximum of 10 minutes.
2.2 Testing

The study was performed on 20 persons, both experienced video game players as well as inexperienced by letting them play a short FPS map. The test persons had one difficulty level that they were told they would be playing on, and then one difficulty level that was the real one that the game was set to, meaning that not all players would be playing on the same setting that they were told by the instructions in the beginning of the test. Since there was two different difficulty settings for this test this created four different groups, two groups that were playing on the same settings they were told, and two groups that would be playing on the opposite difficulty setting from what they were told.

The players would after the test answer a few questions regarding what they thought of the experience, such as how rewarding they found it, by filling in how much they agreed with different statements using Likert scales from 1-5 (one being “I do not agree at all” to five being “I agree completely”) not forcing the test person to choose sides but instead allowing the “middle” option.

All statements had one positive version and one negative, both which were presented and answered in the form.

The test started by the test person being given a paper with information of the test (Appendix A & B (translated)) as well as instructions of the game and controls. It was also clearly marked what difficulty level the game would be set to.

After they had played through the game (map) once and were then asked to fill in the form with questions which can be found in Appendix C.

The questions were designed to answer if the test persons thought the game was rewarding to complete, if they thought it was balanced for the difficulty level that they played on, if they thought it felt challenging, if the enemies were too easy or too hard to defeat and if the effort it took to complete the game was too much or little.
3 RESULTS

The first tests were done matching all four groups against each other in ANOVA-tests, showing that the only differences between groups were in the categories of how challenging the player found the game and how much effort a player found that it took to complete the game, shown in Figures 1 and 2 (please see Appendix D for all result figures).

When looking at the results from the matching between the groups there was no connection between the groups and how overall rewarding they found the game, shown in Figure 3.

Looking at the results from the different difficulty levels (splitting into 2 groups no matter what difficulty level they were told they would be playing on, just the real difficulty they were playing on) there is a clear difference between how much effort the two different groups think they experienced, this is shown in Figure 4. As well as how challenging they found the game, shown in Figure 5.

When checking all 20 test persons independent of what group they belonged to with ANOVA tests, there was a clear connection between how much effort a player thought it took to complete the game and how overall rewarding they found the game, with a p-value of less than 0.05. This is shown in Figure 6.

There is also a connection between how much effort a player thought it took to complete the game and how long time it took for the player to complete the game, shown in Figure 7.

3.1 Discussion

Looking at results from the four different groups showed that the players from group 2 (that thought they were playing easy but was playing on the hard setting) did have a bit higher average score of how overall rewarding they found the game (Figure 3), however the change was not significant and according to the ANOVA paired samples t-tests where all four groups were matched against each other it showed a P value of almost 0.5, ten times higher than needed. Other than that small difference, all groups had almost the same average score on the questions about how overall rewarding they found the game (around 3, compared to 3.9 for group 2). These results speaks against the $H_{11}$ hypothesis, however some of the other ANOVA results where the groups were matched against each other show that the difficulty level did affect how challenging the players found the game as well as how much effort they experienced that it took to complete the game, Figures 4 and 5.
This changes a lot since the effort it took to complete the game was shown to affect how overall rewarding a player found the game when testing against all 20 test persons using correlation tests, it is then clear that the null hypothesis was rejected since there was some change, just not the expected one (overall rewarding). Both of the correlation methods (Spearman and Pearson) found a pattern between how overall rewarding a player found the game and how much effort it took to complete the game (Figure 6). This leads to believe that the connection for effort and overall rewarding is much more individual and not so much affected by what group the player was in since there was no group based increase in overall rewarding but there was group based increase in challenge and effort.

Meaning that if a player felt that it took effort to complete the game, it felt more rewarding, the data cannot show that it was the damage output that made the game more rewarding, nor can it rule it out.

So the conclusion is that the null hypothesis (H₁₀) was rejected since there was some differences (challenge and effort, Figures 4 and 5) between the different groups, even if there is no real way to tell if this was only caused by the damage output there is no way to exclude that it did effect since there were some differences between the groups, even if it was not directly how overall rewarding the players found the game.

Since the H₂₁ hypothesis could only be measured from one group (since there was only one group playing on an easier difficulty level than they were told) it was easy to see the results, even when looking at the earlier results for the H₁₁ hypothesis (Figure 3), since all the groups had a close average of how overall rewarding they found the game. Since the hypothesis is related it can be seen as the results for the H₂₁ is the same as for the H₁₁, being that the perception of how overall rewarding the game was did not change, but areas that affect how overall rewarding the player found the game such as how much effort it took, were affected. Due to this, the H₂₀ hypothesis was rejected.

The H₂₂ hypothesis gets the same results as the H₂₁ since all the groups had almost the same level of overall rewarding, although group 2 was the only one with a higher difference it is not significant (Figure 3).

Some of the results that were not part of any of the hypotheses show some interesting results, such as the effort being affected by or affecting the time it took to complete the game, when checking all 20 test persons in the correlation tests, both tests showed that there was a link between the two, as shown in Figure 7.
3.2 Future work

One thing that could have been done better is the aspect of the learning curve, since some players had experience with FPS games before it could have affected them to learn this game faster, or already having some experience in very similar games. However this could be solved by a short introduction map that all players would have to play through before being able to begin the actual test which would make sure all players have a minimum skill with the game before starting the test-map.

Another test that might have been useful for the study would be if there would have been one more difficulty setting aimed to be even harder than the setting that was hard now, to see if it would affect a player negative to be challenged too much.

The study could also have benefited from having the log collect the shots fired for each separate life of the player, as well as how it was done now with the entire shots fired across all lives.
4 REFERENCES

[1] www.valvesoftware.com


5 Appendix

5.1 Appendix A
The test is anonymous and the results will only be saved with a number assigned later. After the test you will be asked to fill out a form with some questions.
The entire test will be about 10-15 minutes, the game-test is limited to 10 minutes.

You will play through a short First Person Shooter game on Easy-difficulty level. The easy level is suppose to make the game relatively easy for the player to complete, often players can go through it without major difficulties.

The controls are as follows:
Mouse - Aim
Left click - Shoot
W - Walk forward
S - Walk backwards
A - Go left
D - Go right
Space - Jump
Shift - Run
Ctrl - crouch

5.2 Appendix B
The test is anonymous and the results will only be saved with a number assigned later. After the test you will be asked to fill out a form with some questions.
The entire test will be about 10-15 minutes, the game-test is limited to 10 minutes.

You will play through a short First Person Shooter game on Hard-difficulty level. The hard level is suppose to make it difficult to complete the game, having to restart the game one or a couple of times is fully normal.

The controls are as follows:
Mouse - Aim
Left click - Shoot
W - Walk forward
S - Walk backwards
A - Go left
D - Go right
Space - Jump
Shift - Run
Ctrl - crouch
5.3 Appendix C

Age

Gender

How long have you been playing video games? (Years)

How long have you been playing FPS (First Person Shooter) games? (Years)

How many hours a week do you spend playing video games?

How many hours a week do you usually spend playing FPS-games?

What difficulty level did you play on during the test?

What games do you usually play? (choose all that applies)

- None
- FPS (First Person Shoter)
- RPG (Role Playing Game)
- MMO (Massively Multiplayer Online game)
- Co-Op
- MOBA (Such as Dota/Lol/Hon)
- The difficulty level of the game was well adapted
- Casual (Such as facebook games)
- Other[ ]

Scale questions

Choose how much you agree with the statement, 1 You don't agree at all, to 5 You fully agree.

The enemies were too difficult to defeat

The difficulty of the game made it feel rewarding to complete.

The effort it took to complete the game was too small

The difficulty level of the game wasn’t well adapted

It took too much effort to complete the game
It was easy to figure out how to play the game to win
I didn’t find the game challenging enough
The game was too easy to make it worth the effort of completing it

The game felt rewarding to complete
The game was too challenging
The difficulty level I played on felt balanced
The enemies were too easy to defeat
The difficulty of the game didn’t make it rewarding to complete
The difficulty level I played on felt unbalanced
5.4 Appendix D

Figure 1: The different groups average answers about how much effort it took to complete the game.

(First is the told difficulty level, second is the what they actually played on).

Figure 2: The different groups average answers about how much challenge it took to complete the game.

(First is the told difficulty level, second is the what they actually played on).
Figure 3: The different groups average answers about how overall rewarding they found the game.

(First is the told difficulty level, second is the what they actually played on).

Figure 4: The two different difficulty levels average answer about how much effort it took to complete the game.
Figure 5: The two different difficulty levels average answer about how challenging it was to complete the game.

Figure 6: The average effort compared to each step of overall rewarding from all the test players.
**Figure 7**: The average effort compared the time it took for the player to complete the game.