Applying Dynamic Difficulty Adjustment to 2D endless runners

CASE STUDY

Nathan Tubb | Experiential Design | 18/11/20

Table of Contents

1 Introduction

Dynamic Difficulty Adjustment (DDA) is a concept relatively unique to the games industry, it allows difficulty to be fine-tuned to the player's skills, even after it has left the developer's hands. Missura and Gärtner, (2009) describe the ideal game as being able to adjust its difficulty according to player performance. Zohaib, (2018) talks about how DDA can make games feel less boring and frustrating when they are too hard or easy. The author goes on to summarize that DDA intends to keep the players engrossed throughout gameplay and to maintain an adequate level of challenge throughout.

This study measures the effectiveness of dynamic difficulty adjustment in 2D endless runners, developing a system that allows all players to take part at a difficulty suited to them. The test involved players skill being acquired by creating an average of both the coins collected and average lifespan in each attempt they took. The results were then used to influence the pacing of the level. Achieving the aims of this study would contribute to users reaching optimal flow (Appendix A) (Csikszentmihalyi, Harper and Row, 1991) which, is shown to produce higher levels of enjoyment within the game thus increasing players intention to play more often and for longer (Wu and Liu, 2007).

Testing took place using online services as the current COVID-19 pandemic prevented face to face testing. Project files were sent to each participant across discord and surveys were answered using google forms. Testing began on the 27/12/20 and ran until the 02/01/21 with a total of 19 participants, 17 of which producing valid results. Upon completion of testing, all results were scrutinized and can be seen in section 5.

2 Hypothesis

The initial hypothesis is that: The use of a dynamic difficulty adjustment system within the endless runner genre greatly increases user enjoyment for players of all skill levels.

A system that dynamically changes the ramping difficulty of oncoming platforms allows players of all skill levels to feel as though they are always sufficiently challenged and that this increases levels of enjoyment and motivation to play.

3 Literature Review

3.1 USEFUL WORK

Hunicke, (2005) explores the basic design requirements for the effective use of DDA offering evaluations. After developing a simple DDA system for an Fps game and gathering data from participants with both active and non-active DDA sessions, his results showed that even simple DDA systems can improve players enjoyment without ruining the sense

of control or self-esteem. This influences the aims of this project as it allows for its scope of the designed system to remain relatively manageable without jeopardizing the results. Hunicke also found little correlation between players self-rating and their actual performance, suggesting that players did not know what difficulty they should be playing at. This will require the project under investigation to use a practical based diagnostic such as test level to group participants rather than a questionnaire or interview as these could be inaccurate.

Developing on the effects of DDA on player experience (PX)Ang, (2017) discuss' uses of traditional player-oriented DDA (pDDA) in which players adjust the difficulty themselves and Ramping DDA (rDDA) where the difficulty naturally increases over time i.e. increase of running speed or falling objects. Firstly, Ang found that whilst both forms of DDA have positive impacts on challenge and skill, rDDA, which is traditional in endless runner games, lacks any form of self-control or support for players that may find the ramping difficulty to slow and boring or fast and difficult. Furthermore, using pDDA can allow for players to solve this issue by setting the difficulty themselves to a more appropriate level. However, Ang found that pDDA also causes a reduction of identification or "loss of self-consciousness". This suggests that the use of this project under investigations automated DDA system would have positive impacts in avoiding the negative impacts of pDDA whilst also alleviating the lack of support experienced when using rDDA systems. Secondly, in certain instances Ang found that some players took offence to the DDA system reduced the difficult, suggesting that players being aware of DDA can have negative effects on PX. As a result of this, participants of the project under investigation will not be made aware of the DDA system developed.

Xue *et al.*, (2019) re-iterate that DDA should be invisible for players so that they do not feel offended by the game. However, they also suggest that many research projects and games produce DDA systems targeted at short term goals such as getting the player out of undesirable situations like boredom or frustration. Xue *et al.*, argue that this type of design makes it impossible to optimize long term goals such as player engagement and retention. These authors argue that long term designs that can increase challenge to meet exponential skill increase should be used instead as they will have a much greater lifespan than systems that only increase to a certain difficulty level. To try and accommodate this within the project under investigation, the DDA system will use a modifier system of changing different values by set amounts to increase or decrease difficulty rather than creating a set range of difficulty levels with preset values that clamp the difficulty range which a player can reach. This allows for an exponential amount of difficulty levels for the players to end up in.

3.2 THE APPLICATION OF DDA IN SIMILAR PROJECTS:

Kaimbault, (2016) studies progression in popular endless runners and suggests a chart (Appendix B) depicting the progressive loop that a player should experience every time they start a new attempt of a level. The study suggests that each time a player begins the

level that they need time to adjust to the speed and controls that they have available to them, each segment on the chart should take a consistent amount of time each run to create a safe area for players to prepare themselves. To improve player enjoyment and interest Kaimbault also suggests that new challenge should be incremented the further the player gets in the attempt with the addition of new mechanics, powerups or abilities. The project under investigation will include this progression loop increasing challenge through player speed and using the introduction period to give players time to prepare themselves each run.

Relating to this study Calder, (2018) explored the effects of DDA in 3d endless runners similar to games like subway surfers. Calder's system attempted to adjust the difficulty and spawn rate of upcoming obstacles in real-time. To do this he took variables such as coins and powerups missed as data to calculate if the difficulty needed to be increased or decreased. This was done as a penalty system which would be calculated every 10 seconds the player lived; an overall performance would then be created being added to a list. The average grade of 3 performance calculations (30 seconds) would then change the difficulty level mid-game. This would increase or decrease the level of sections being spawned. The implementation of this system allowed for successful use of DDA in a similar game, it will provide an excellent base for the design for the system used in the project under investigation.

Pedersen, (2014) also explored DDA in endless runners, using reinforcement learning to balance procedurally generated levels in a lane-based 2d runner. The algorithm used generated 1 of 1024 possible levels and would then ask the player for a rating when they finished it. If the player disliked the level it will be chosen less often, the opposite happens if they liked it. The authors found that the implementation of DDA into this type of game also improved the learning curve of players as they achieved better results at a higher difficulty faster than previously, players also found the game more enjoyable. This shows that the application of DDA into this genre has positive effects on PX and helps demonstrate the benefits of its use.

This review has shown that the used of DDA in games can greatly increase elements flow, enjoyment and skill felt by players. Arguments presented regarding its visibility and participants judgments of skill have made the need for a practical diagnostic and undisclosed system clear to reduce the risk of outlier results from testing. Finally, with current research providing similar experiments in the field, this project has reliable a method to draw on for the creation and implementation of such a system.

4 Methodology

4.1 DYNAMIC DIFFICULTY SYSTEM

The DDA system for this study is closely related to Calder's, (2018), in using the collection rate of coins and powerups to determine whether the difficulty of upcoming platforms should be increased or decreased every 30 seconds.

Differing from Calder, this study uses a combination of coins missed and time alive to determine average performance over three level attempts, at which point the difficulty is changed. Furthermore, only the characters acceleration will be changed to modify difficulty.

Due to the literature review suggesting that DDA systems should be designed with long term goals like player skill increase. This projects approach is intended to allow for exponential player growth and will apply a gradual assessment of skill as appose to every thirty seconds. Secondly, as the scope of this project is smaller than Calders, survival time is expected to be less, therefore, a different system is needed to evaluate performance.

4.2 IMPLEMENTATION

The base performance for each attempt is set to 50, this is then increased or decreased to a value from 0 to 100 by the time (Appendix C) and coin modifiers (Appendix D) which are generated from the level attempt. Each modifier has a weight which determines how strongly it impacts performance (note, the maximum combined value of both modifiers must equal + or - 50).

The average performance value of three runs is then used to determine whether the DDA system increases or decreases the difficulty of the game the boundaries for this value are shown below.

Difficulty Values:

0-22 = - 2 difficulty levels

23-45 = - 1 difficulty level

46-55 = no change

56-77 = + 1 difficulty level

78-100 = + 2 difficulty levels

The difficulty of the game is changed by multiplying the milestones at which the players speed increases, increasing difficulty makes these milestones shorter, whereas decreasing makes them larger. This means that players need to spend either more time at slower or

faster speeds depending on what level they are at (note it is considered harder to stay alive at faster speeds).

4.3 EXPERIMENT

The methods for this main experiment are similar to Pedersens, (2014) who gathered performance metrics and surveyed participants, to establish the quality of his DDA algorithm.

This experiment consisted of three stages, participants first completed a diagnostic to determine their skill level, secondly, they completed a play session of the game which produced metrics on the DDA systems actions, and finally, filled out a survey to determine how they perceived the difficulty, challenge, and enjoyment of the game. These metrics and surveys were then compared to establish the effectiveness of the produced system.

Firstly, to gather data that proves the DDA system works for all skill levels of players, participants were placed into skill groups. To determine which group participants would be in they played a diagnostic. The diagnostic calculated the players' group by taking an average of their performances in the initial difficulty of the level. This performance was gathered the same way as difficulty system however was only used to calculate which group a participant would be in, as shown below in the group breakdown.

Group Break Down Thresholds:

Novice: Participants who performed between 0 and 40

Intermediate: Participants who performed between 41 and 70

Advanced: Participants who performed between 71 and 100

Second, each participant played a further fifteen attempts of the project's endless runner. This allowed the system to make five difficulty evaluations from start to finish. Once this was completed a text file would be created on the participant's computer containing their: name, diagnostic results, and the results of each difficulty evaluation throughout their playthrough (Appendix E).

Finally, participants were again given a survey (Appendix F) that used a combination of adapted flow questionnaire questions developed by Magyaródi *et al.*, (2013)(Appendix G), answering using the Likert Scale, (Likert, 1933). Followed by questions that allowed them to explain their experience. This was then compared to metrics produced by the DDA system, to determine if its decisions had created positive results.

To determine whether the hypothesis of this project was met, predictions were made for each group of participants. Participants in the advanced group were expected to begin the experiment collecting the most coins and surviving over the expected time, these numbers would then fall as the DDA system increased their difficulty and made it harder. In comparison, the opposite was expected of the novice group, whilst the intermediate group were expected to see the least change in their statistics and difficulty.

Finally, these predictions allowed for the data received from playthroughs to determine whether systems judgments worked as intended.

5 Results

5.1 DIAGNOSTIC

The diagnostic separated the valid 17 participants into the groups defined by the DDA system accurately, based on the group breakdown in section 4.2 as can be seen below.

Assignment of participants to skill groups:

Novice: 9

Intermediate: 8

Advanced: 0

As can be seen, no participants performed at a high enough level to be placed in the advanced group.

5.2 DIFFICULTY ADJUSTMENTS

Data in this section represents the average findings of data compiled from the results files (discussed in 4.2, Appendix E).

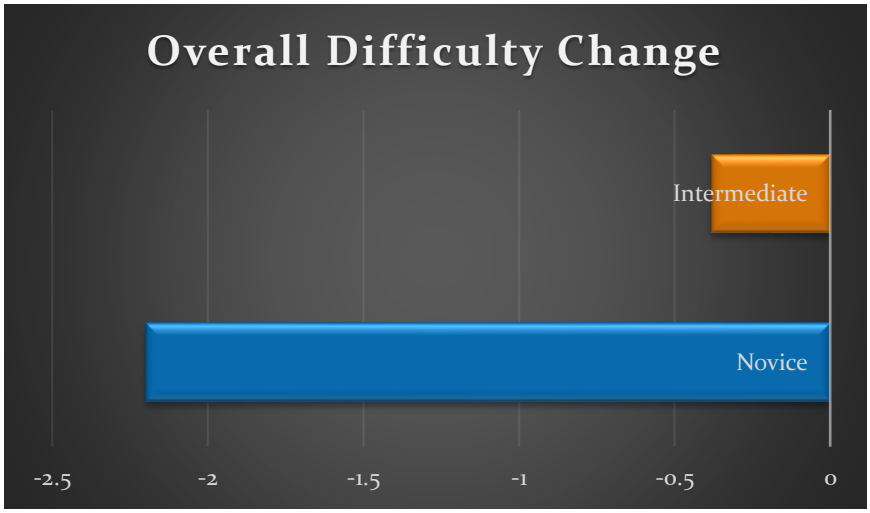


Figure 1 The overall Average difficult change for participants of both groups

As can be seen in figure 1 the novice group had a much greater difficulty decrease in the study, being decreased by 2.2 levels. The intermediates in comparison decreased an average of 0.38 levels.

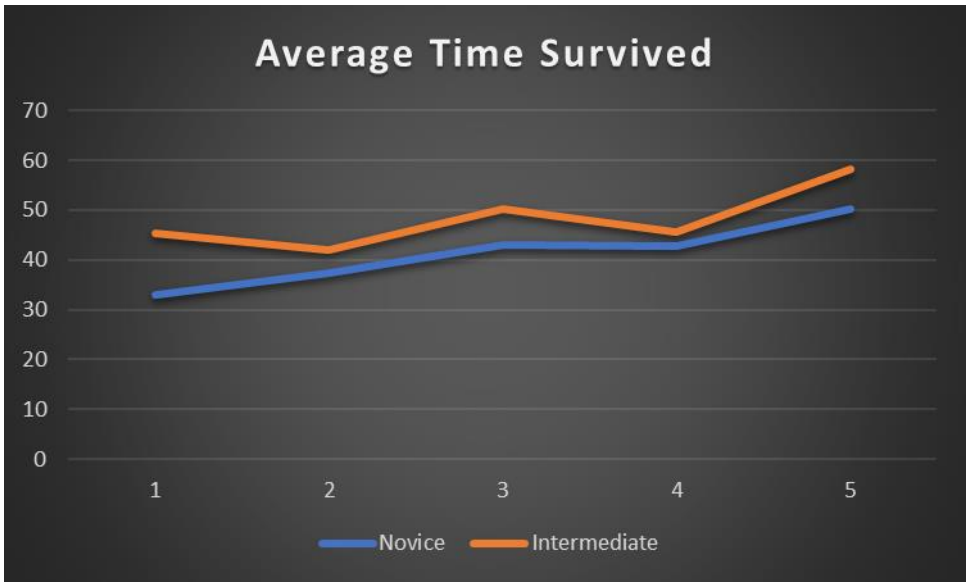


Figure 2 The Average Time that participants in each group survived between each difficulty evaluations

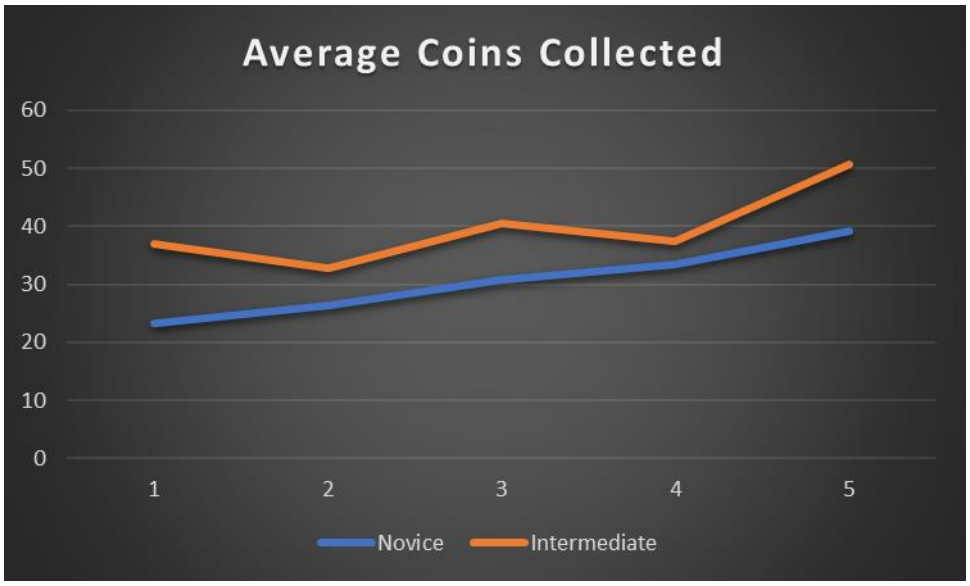


Figure 3 The Average amount of coins collected by participants of groups between each difficulty evaluation

As can be seen in figures 2 and 3 intermediate players performed at a higher level throughout the study, which again supports the hypothesis and suggests that the diagnostic worked as intended. Furthermore, both groups performance is shown to increase throughout the experiment, towards the target goal of 50 seconds survival time.

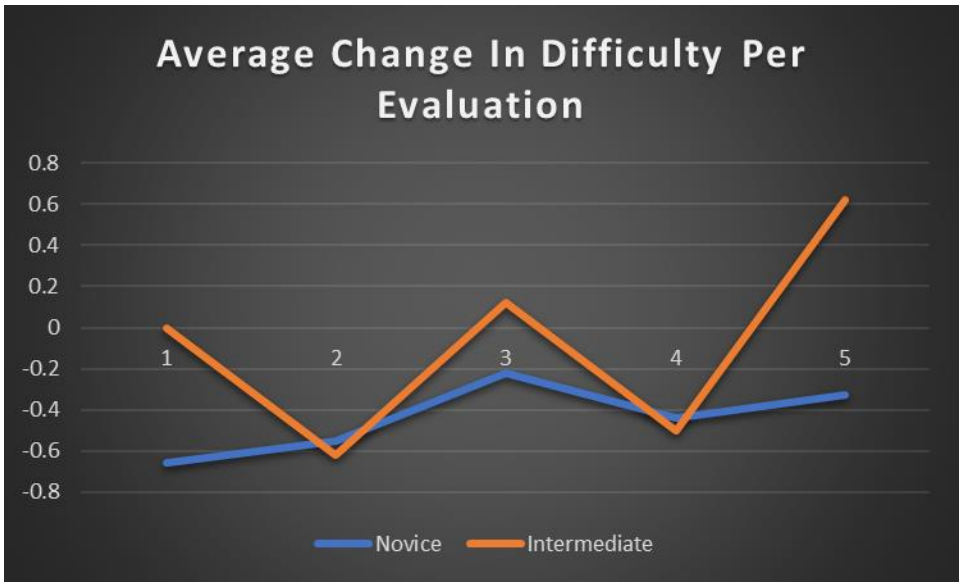


Figure 4 The Average difficulty change for the participants of each group at each of the five evaluations in the study

An outlier seen in figure 4 shows how the intermediate group experienced a great change in their difficulty evaluations as appose to the general increase seen by the novice group. This is proven not to be the system's fault however as both figures 2 and 3 show that participants performed worse during these evaluations.

5.2 SURVEY ANSWERS

5.2.1 Overall Opinions of Gameplay

Firstly, the survey answers using the Likert scale are shown as a bar chart to compare how both groups felt overall about their experience. Secondly, written answers have been generalized into key points that are presented.

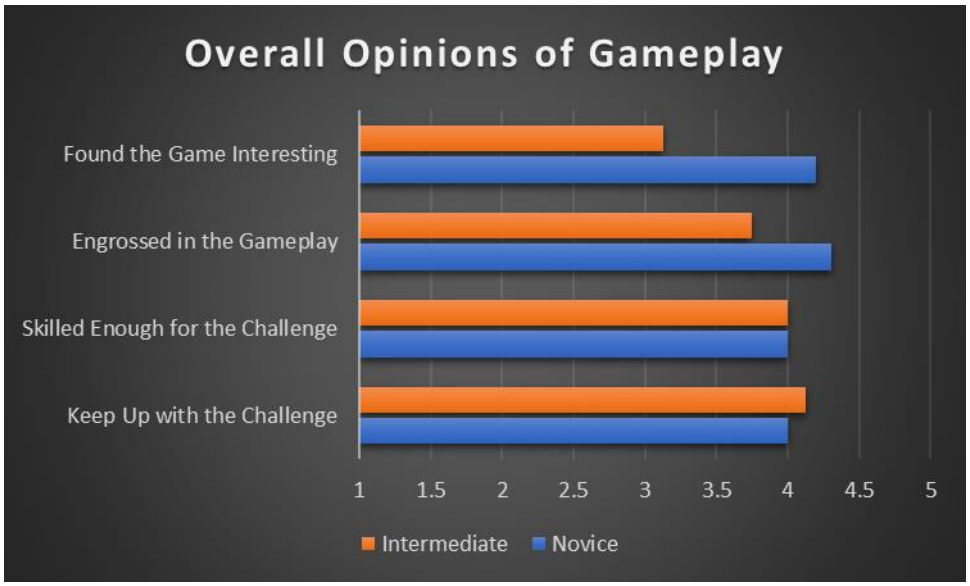


Figure 5 The consensus from players in the two groups on the asked questions

Both groups felt that they were skilled enough and able to keep up with the overall challenge of the game, however, the intermediate group are shown to have felt less interested and less engrossed in the gameplay, suggesting that they enjoyed it less.

5.2.2 Key points raised in Written answers

Novice Group

The novice group suggested that they felt the difficulty was well balanced to their expectations, though mentioned that it became hard very quickly around 40 seconds. They also suggested that coins were difficult to collect and often died when trying to gather more. Furthermore, a number of these participants had some difficulties with the mechanics and control scheme of the game. Finally, some novice participants mentioned that they found the game more enjoyable towards the end of their play session.

Intermediate Group

In comparison to the novice group, intermediates felt that the game was at times a little boring or stale, especially the longer they played. They had mixed opinions on the difficulty and challenge of the game with some mentioning speed was too slow at the start and others commenting on the difficulty of the max speed. However, upon surviving longer, the speed became uncontrollably fast and too difficult. Some participants in this group also had issues with the controls.

6 Discussion

6.1 FINDINGS

Firstly, predictions of group performances (other than advanced) have been proven as the results show a clear comparison between how novice and intermediate participants were handled, and how that affects their performances. As expected, novice players experienced the highest drop in difficulty, which can be seen to correlate with both their improved performance. Intermediates, on the other hand, experienced little change from the starting difficulty and suggesting that they were already close to the level expected for the game. They are then shown by the end of their sessions to have progressed past the initial target and increased their difficulty.

Secondly, Increased performance results could be accredited to players learning the game or becoming used to it. However, the correlation between intermediate participants drop in performance (Section 5.2, Figure 2,3), and difficulty levels (Section 5.2, Figure 4) suggest that difficulty changes were the cause. After each difficulty decrease participants had a much higher score in the next evaluation with their average change being above 0. After each increase participants, performance levels dropped slightly which then decreased the average of the next evaluation. This supports the effects of the DDA system on player performance.

Thirdly, novice players said that they found the game more enjoyable towards its end than its beginning. Furthermore, their answers on the Likert scale show that they felt interested and engrossed in their experience. Correlating this to the DDA systems decisions proving to increase their performance, it can be surmised that the hypothesis for this study is supported.

On the other hand, as several intermediate players stated that the game felt boring and stale after playing for a while, and with their results from the Likert scale questions (section 5.2.1, Figure 5) showing that they felt less engrossed and interested than novices overall. This may be due to the initial speeds of the game as it was mentioned to be too slow at the beginning of attempts, which was found to be boring. This was out of the control of the DDA system as the start speed and max speed of the player character was

constant throughout the experiment. However, it affected players of higher skill. Based on these differing results from the two groups its cannot be concluded that the hypothesis is entirely supported.

Finally, several players in both groups had issues with the game jump controls and platform mechanics. This may have harmed their experience and enjoyment and could have influenced their overall enjoyment.

6.1 RECOMMENDATIONS

As this study failed to determine how the given DDA systems interact with players of an advanced skill level future research should aim to target this group specifically, perhaps through adapting an already created endless runner using players with existing high-level experience.

Secondly, as players in this study found the max and minimum speeds either boring or too difficult, future research should look at the effects of increasing or lowering these with a DDA system, based on the skill level or group of the player.

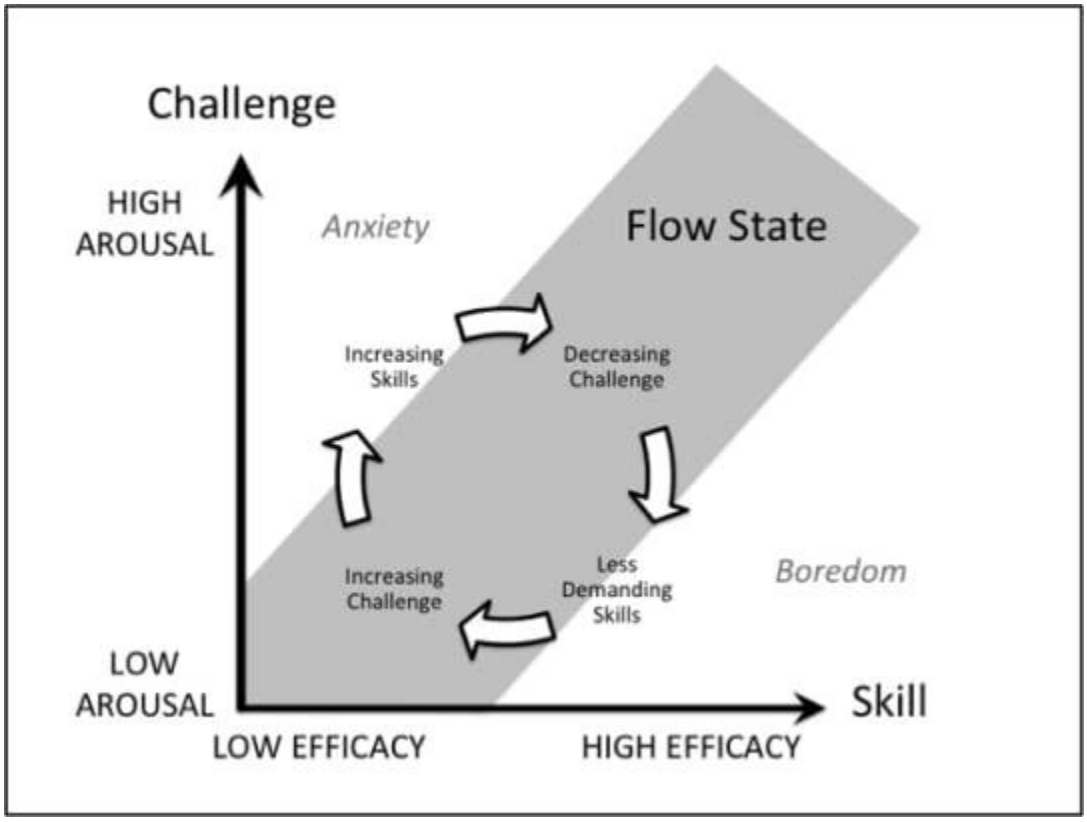
Finally, as several participants encountered issues with the controls and mechanics of the game, future studies of this kind should ensure all mechanics are clearly explained to participants, no matter how simple they might seem.

Conclusions

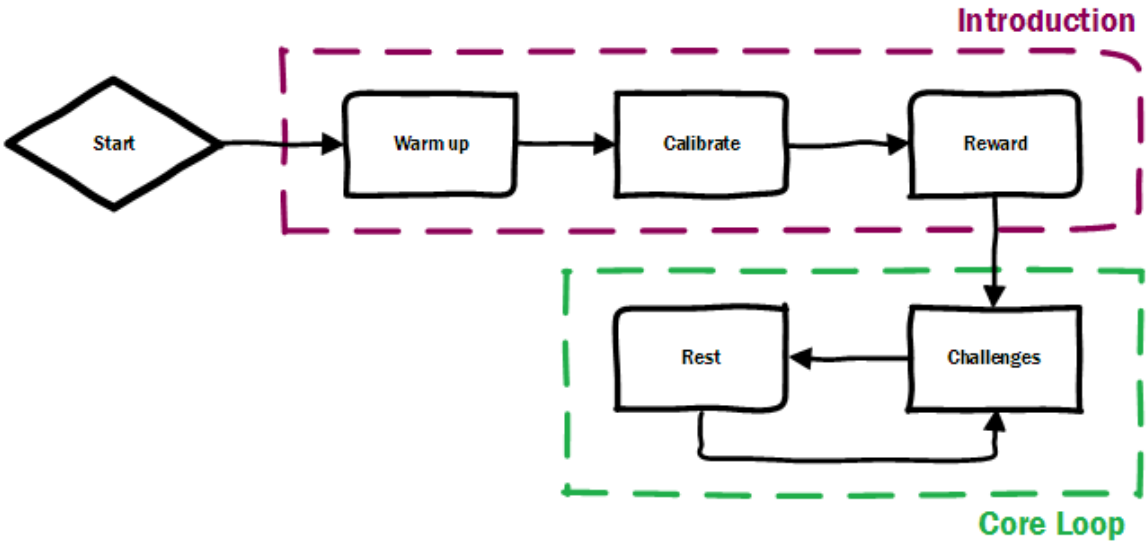
In conclusion, this study proves that the addition of a basic Dynamic difficulty system into 2d endless runners can improve the performance of its players towards the level expected for the game. However, whether it can reduce the performance of players performance above expectation still needs to be studied. Furthermore, whilst certain groups and types of players found more enjoyment from this, others became bored and uninterested, showing the need for greater balancing, and tweaking of mechanics with later iterations of this project or others.

Appendices:

Appendix A: Mihaly Csikszentmihalyi's, theory of flow



Appendix B: Kaimbault's, (2016) chart on the progression of endless runner levels



Appendix C: Time Modifier

The Time Modifier is based on the player reaching or overshooting a survival time specified by the developer of the game. In this project, the survival time will be 50 seconds.

A decimal range from -1.0 to 1.0 is used to determine the player's success in reaching the goal. The value of this range multiplied by the weight factor of the Time Modifier:

Number Range Break Down:

-1 = 0 seconds: Player died in the warmup or when the game starts

0 = survival time target: In this instance 40 seconds

1 = (2 * time alive target): Player dies at or past the 80-second mark

Time Modifier Calculation:

a = Time Modifier

b = Range Value at time of death

c = Modifier Weight

$$a = b * c$$

Appendix D: Coin Modifier

Like the time modifier, this modifier also uses a range system. However, this range starts at -0.5 and ends at 1 rather than -1 to 1 as it is impossible to collect all coins that are presented to the player. This means that the player only needs to collect 25% of the coins to avoid a penalty.

The calculation works out the percentage of overall coins collected based on the number of coins presented, and then finds the value of that percentage in the range.

Percentage coins collected

a = coins collected

b = coins presented

c = percentage collected

$$a / b = c$$

Value of percentage in range Calculation

Mathf.lerp = find a value in a range based on a percentage

Mathf.lerp(-0.5, 1, percentagecollected)

Modifier Calculation

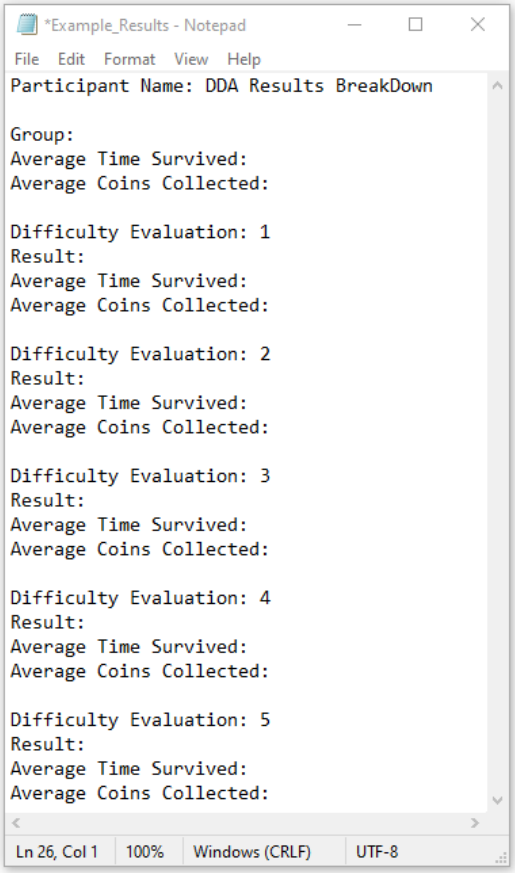
a = value in range

b = modifier weight

c = resulting modifier value

a * b = c

Appendix E: Example of the results text file



Appendix F: Survey used in this experiment

https://forms.gle/8ZhwooyBNYsw7Y92A

1. Please enter the same name you entered when playing the game (written)
2. Please Upload your Results txt file (upload box)
3. Were you able to keep up with the challenge throughout? (likert)
4. Did you find that your skills were balanced with the challenge most of the time? (likert)
5. Did the task totally engross your attention? (likert)
6. Did you find the task interesting? (likert)
7. If you answered 3 or below on the previous question, please explain why (written)
8. Did you find the task too easy or hard at any time, if so, when, and why? (written)
9. Please Share any other further comments or feelings on the difficulty or challenge of the task (written)

Appendix G: Flow Questionnaire produced by Magyaródi *et al.*, (2013)

1. I was able to keep up with the challenges
2. I felt I can meet the requirements of the situation
3. I had a grip on the events
4. I felt I was in control over the situation
5. I knew I was able to solve the task
6. I knew exactly what I had to do, and I acted accordingly
7. This task was not too difficult
8. I felt that what I had to do matched my skills well
9. I could effortlessly perform well
10. My skills were in balance with the challenges of the activity
11. My mind worked in total harmony with my body
12. My attention was not engrossed at all by the activity
13. It was boring for me.
14. The activity totally engrossed my attention
15. I forgot about the progress of time
16. I found the task interesting
17. I fused with the task
18. I forgot about my close environment

Bibliography:

Ang, D. (2017) 'Difficulty in video games: Understanding the effects of dynamic difficulty adjustment in video games on player experience', *C and C 2017 - Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pp. 544–550. doi: 10.1145/3059454.3078706.

Calder, C. (2018) 'Dynamic Difficulty Adjustment in Procedural Content Generation'.

Csikszentmihalyi, M., Harper and Row (1991) 'Flow: The Psychology of Optimal ExperienceFlow: The Psychology of Optimal Experience, by Csikszentmihalyi Michael. New York: Harper & Row, 1990, 303 pp., $19.95, cloth.', *Academy of Management Review*, 16(3), pp. 636–640. Available at: http://journals.aom.org/doi/10.5465/amr.1991.4279513%0Ahttps://www.tandfonline.com/doi/full/10.1080/00222216.1992.11969876%0Ahttps://www.cambridge.org/core/product/identifier/9780511621956/type/book.

Hunicke, R. (2005) 'The case for dynamic difficulty adjustment in games', *ACM International Conference Proceeding Series*, 265, pp. 429–433. doi: 10.1145/1178477.1178573.

Hunicke, R. and Chapman, V. (2004) 'AI for dynamic difficulty adjustment in games', *AAAI Workshop - Technical Report*, WS-04-04, pp. 91–96.

Magyaródi, T. *et al.* (2013) 'Psychometric properties of a newly established flow state questionnaire', *The Journal of Happiness & Well-Being The Journal of Happiness & Well-Being (JHW) The Journal of Happiness & Well-Being*, 1(12), pp. 89–100.

Missura, O. and Gärtner, T. (2009) 'Player modeling for intelligent difficulty adjustment (resubmission)', *LWA 2009 - Workshop-Woche: Lernen-Wissen-Adaptivitat - Learning, Knowledge, and Adaptivity*, (October 2009), pp. 76–83. doi: 10.1007/978-3-642-04747-3.

Pedersen, K. (2014) 'Procedural Level balancing in runner games', *SBC - Proceedings of the SBGames 2014 | ISSN: 2179-2259*, pp. 797–801.

Wu, J. and Liu, D. (2007) 'The Effects of Trust and Enjoyment on Intention to Play Online Games', *Journal of Electronic Commerce Research*, 8(2), p. 128.

Xue, S. *et al.* (2019) 'Dynamic difficulty adjustment for maximized engagement in digital games', *26th International World Wide Web Conference 2017, WWW 2017 Companion*, (April 2017), pp. 465–471. doi: 10.1145/3041021.3054170.

Zohaib, M. (2018) 'Dynamic difficulty adjustment (DDA) in computer games: A review', *Advances in Human-Computer Interaction*, 2018. doi: 10.1155/2018/5681652.

Likert, R. (1932). A technique for measurement of attitudes. Archives of Psychology, 140, 5-55.