Real-Time Adaptive Game Difficulty: Optimising Player Flow and Skill Evolution

Master Computer Science

Evan Meltz

Name:	Evan Meltz
Student ID:	s3817911
Date:	[14/08/2025]
Specialisation:	Artificial Intelligence
1st supervisor:	Mike Preuss
2nd supervisor:	Giulio Barbero

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Contents

1 Introduction

As video games become increasingly popular, a wider variety of players start to emerge, with different skill levels and playstyles. This increases the need for more customised experiences that can adapt to different players. One way this is currently done is through Dynamic Difficulty Adjustment (DDA), which is implemented to foster flow. However, it often disregards specific player characteristics and focuses primarily on skill level alone. In this thesis we bring together the theory of flow, DDA and player modelling to investigate how they impact player experience and how we can better utilise these methods to provide a more personalised experience to individual players.

1.1 Problem Definition

While the theory of flow has been widely used to analyse players' experience, it is often too simplistic when it comes to addressing the detailed complexities of real gameplay, making it insufficient to guide adaptive systems for individual players.

Player modelling offers a way to represent individual player behaviours, but on its own, it is primarily used retrospectively — more as a hindsight tool than a dynamic one.

Dynamic Difficulty Adjustment (DDA), however, uses player modelling dynamically, as it aims to adapt the game's challenge in real time based on player skill. Yet existing DDA systems lack concrete methods to balance challenge, skill, and player behaviour in a holistic and dynamic way. Specifically, most approaches focus solely on skill, failing to address playstyle and its nuances.

The gap, therefore, is the need for real-time systems that tailor the play experience to both a player's skill level as well as their playstyle. This thesis identifies and addresses this gap by proposing a system that embeds real-time player modelling directly into the game loop — much like DDA, but continuously monitoring key performance indicators to classify not just player skill, but also playstyle. The system then deploys dynamic countermeasures, personalising the game to each specific player in a broader context.

The ultimate goal is not simply to make the game harder or easier, but to enhance engagement and support flow while encouraging players to adapt and explore new playstyles, fostering both skill improvement and playstyle evolution.

1.2 Research Questions

We design a system of in-game adjustments (*countermeasures*) that adapt to specific aspects of individual playstyles. We then analyse the system created in terms of play experience, in particular flow (as function of skills and challenges), affective responses, and playstyle adaptation. Therefore, we define the following research questions:

RQ1: Real-Time Challenge Adaptation
What is the impact of real-time adaptive countermeasures on player experience and the core components of flow?

- **Sub-question 1 (Psychological Flow):** How does real-time challenge adaptation influence key psychological components of flow?
- **Sub-question 2 (Affective Response):** How do these components evolve across the play experience, and do adaptive countermeasures impact them differently for low- and high-skill players?

RQ2: Refined Modelling
Which combination of performance metrics and self-reported surveys yields the most accurate real-time model of the player's skill state and receptiveness to new playstyles and strategies?

RQ3: Playstyle Experimentation
Which playstyle's countermeasure had the most significant effect on encouraging players to refine, evolve, or change from their preferred playstyle?

1.3 Related Work

1.3.1 Flow Theory in Games

The concept of *flow* describes a psychological state of complete immersion and optimal experience, arising when a perfect balance is struck between the challenge posed by a game and the skill of the player [Csikszentmihalyi, 1990, Cowley et al., 2008]. When this balance occurs, players enter flow - they lose track of time and become deeply absorbed in the gameplay.

If the challenge exceeds the player's skill level, the player will experience frustration or anxiety; whereas if the challenge is too low, the result is boredom. The *core components* that characterise this experience include frustration and boredom at either end of the spectrum, while the flow state itself involves feelings of satisfaction, control, and excitement [Sweetser and Wyeth, 2005]. These components provide a practical framework for analysing player experience and have influenced both game design as well as some adaptive AI approaches.

1.3.2 Player Modelling

Player modelling refers to the process of creating a computational representation which reflects a human player's in-game behaviour.

It is widely used to study gameplay, train game AI, and analyse players' experience. It can also be used to guide the behaviour of non-player characters (NPCs) as collaborators, opponents, or neutral agents [Machado et al., 2011].

Since its establishment, player modelling has evolved, resulting in various methodologies and definitions, but still typically requires specifying a model type as well as what input data should be used. The first model type is the top-down approach, where a theoretical framework guides the formulation of a model that is then later validated through empirical testing. The second one is the bottom-up approach, which builds a model by directly mapping raw player inputs to state representations, relying on observed data and annotated player states without initial assumptions. Hybrid approaches that combine elements from both strategies are also employed [Yannakakis et al., 2013].

Recent surveys highlight key taxonomies, challenges, and trends, while also detailing the range of AI techniques applied to player modelling — including neural networks, finite state machines, and Monte Carlo tree search [Farooq and Kim, 2024, Machado et al., 2011]. Recent work has even explored using player models to infer player personality traits from in-game behaviour and affective expression [Habibi et al., 2023]. Across the literature, frameworks for player modelling have ranged from early dynamic models [Charles and Black, 2004] to holistic, player-centric approaches that seek to adapt to performance, affect, and playstyle [Bontchev, 2016, Farooq and Kim, 2024].

In practice, data mining and machine learning techniques have enabled predictive modelling of player groups, but their utility for predicting and tailoring to individual behaviour remains limited [Yannakakis et al., 2013].

A key challenge in the field is the problem of generalisability: models may adapt only to a limited set of game states and fail to generalise to new situations, thereby reducing their effectiveness as soon as the game genre or player behaviour shifts. Moreover, researchers face a lack of publicly available and robust datasets, which are needed to build and benchmark comprehensive player models [Hooshyar et al., 2018].

1.3.3 Dynamic Difficulty Adjustment (DDA)

Dynamic Difficulty Adjustment (DDA) encompasses a broad range of techniques which aim to dynamically adapt a game's challenge or environment in real time, based on the player's current performance and behaviour [Zheng, 2024, Zohaib, 2018]. The primary goal of DDA is to maintain player engagement as well as psychological flow.

DDA approaches are found across many genres, such as action/adventure games, puzzle/strategy games and racing/sports games. Implementations can be rule-based — like using simple thresholds on player statistics to increase or decrease difficulty — or more sophisticated, like employing machine learning and reinforcement learning to predict player state and make more calculated adjustments [Shao et al., 2019, Hossan et al., 2024, Xiong, 2019]. Common DDA mechanisms include adjusting enemy strength, spawn rates, resource availability, and even altering game rules dynamically. Many modern systems employ the earlier discussed practice of player modelling, integrating real-time data on performance, actions, and sometimes even affective responses to guide adaptation.

However, despite these advances, most DDA systems primarily aim to balance overall challenge relative to skill, often neglecting the broader spectrum of player behaviour and playstyle. This means that many opportunities to support/challenge playstyles, encourage experimentation, or foster skill development are missed. Recent research highlights the need for more holistic DDA frameworks that personalise the play experience by taking both skill and playstyle into account [Hossan et al., 2024, Xiong, 2019].

1.3.4 Final Remarks

Overall, while the focus of most of this work has been to tailor gameplay to better match or support a player's current skill — to help them sustain a state of flow — there has been little focus on attempting to model more nuanced playstyles or actively countering established playstyles to promote adaptation.

2 Methodology

2.1 Experimental Game

The system for this experiment was based on a modified version of the *3D Wave Shooter* Unity asset [Buckley, 2025]. The game is a wave-based, top-down shooter in which the player's objective is to survive each wave by defeating all spawned enemies while managing health, resources, and navigating the environment. Between waves, players also have access to a shop where they can purchase weapons and upgrades using in-game money earned from defeating enemies.

To facilitate player modelling and adaptive difficulty, the base game was extensively customised. Key modifications include the integration of detailed performance tracking (e.g., damage dealt, movement, weapon switching, time moving/idle), real-time logging, and the embedding of surveys to assess player experience at specific points during gameplay. Additional systems were implemented to allow dynamic adjustment of game variables related to specific aspects of the user's playstyle (countermeasures), including enemy spawn rates and behaviour, resource availability, and environmental factors, all driven by the outputs of the player modelling and DDA algorithms.

The core gameplay loop is structured around progressive waves that increase in difficulty. After certain waves, the system will also collect player feedback through surveys. These adaptations ensure that the experiment captures detailed gameplay and self-reported data and provides a flexible test bed for evaluating the effectiveness of this more nuance-inclusive dynamic difficulty adjustment and playstyle countermeasures.

This game was chosen for its balance between simple, accessible gameplay as well as its potential to discover and capture potentially interesting, nuanced playstyles and allow for the implementation of unique countermeasures.

2.2 Pilot Phase(Pre-Experiment)

In our methodology, we isolate the main metrics defining different playstyles in a pilot phase. We then define appropriate thresholds and countermeasures for each of these dimensions and test them in our experimental setup.

To establish a foundation for player modelling and adaptive difficulty, the pilot phase of the project focused on developing robust in-game metric gathering and logging. The purpose of this phase was to implement code that could continuously track key performance metrics during gameplay as well as gain insights into users' initial reactions to the game. This study's methodological approach to player modelling, adopts a hybrid of top-down and bottom-up techniques. The top-down component involved defining initial categories and thresholds informed by prior literature and design intuition, while the bottom-up component focused on empirical data collection to capture the actual diversity of player behaviour observed in practice to refine and update the initial categories and thresholds [Yannakakis et al., 2013].

The following metrics were logged for each participant and gameplay session (consisting of two waves):

- **Shooting Accuracy:** The ratio of shots landed to shots fired, reflecting aiming skill and control.
- **Wave Completion Time:** The total and average time required to clear each wave, used as a proxy for skill, risk-taking, or efficiency.
- **Damage Taken:** The amount of health lost per wave, indicating defensive ability or risk aversion.
- **Movement Metrics:** Time spent moving vs. idle, total distance travelled, and average movement speed, to distinguish between active and passive playstyles.
- **Weapon Usage Patterns:** Frequency and diversity of weapon switching, revealing exploratory versus habitual behaviour.

After the implementation of the metric gathering system, this pilot test was conducted with eight participants who would each play two consecutive waves of the game, with the goal being to collect a dataset representative of diverse playstyles and skill levels, and to empirically inform threshold definitions and clustering hypotheses for subsequent player modelling. These early play sessions provided feedback on the reliability of the logging system, the variability of player behaviour, and informed adjustments to metric selection and categorisation.

This phase set the groundwork for the hybrid player modelling and countermeasure design in later phases, ensuring that subsequent algorithms would be data-driven and sensitive to the nuanced patterns of both skill and playstyle observed in real players.

2.2.1 Insights and Metric Selection

Analysis of pilot gameplay revealed several actionable trends:

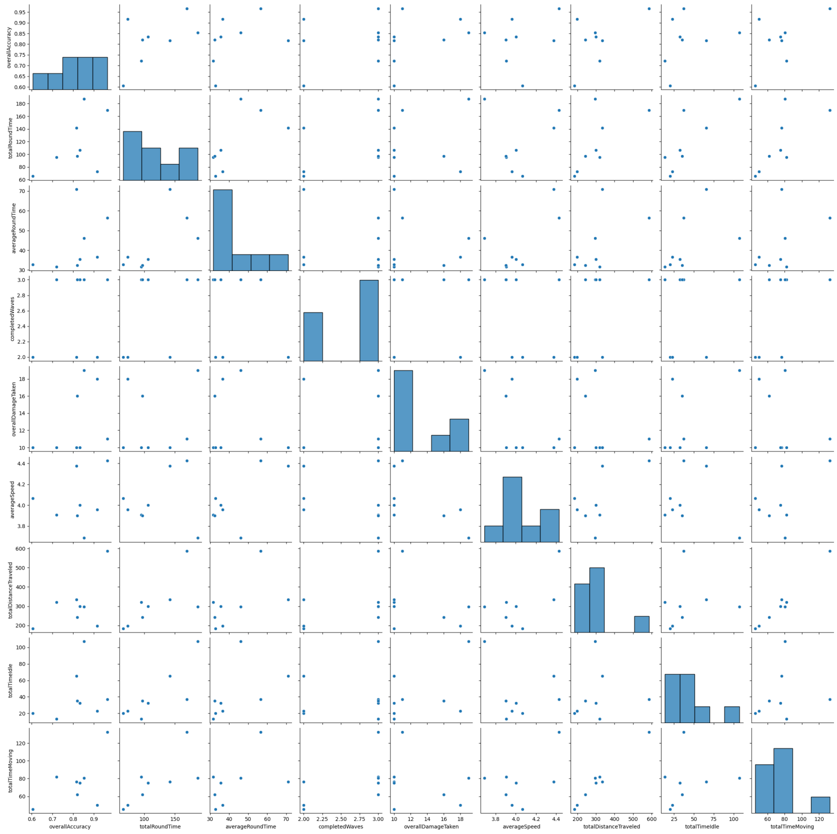

Figure 1: Pilot pairplots: Notable trends include idle time vs. damage taken and movement vs. accuracy. These informed quadrant and metric selection for the main experiment.

- **Movement and Survival:** Players who spent more time moving tended to survive longer rounds and had higher overall accuracy. This justified the inclusion of the later introduced Movement-vs-Idle (MvI) metric as a playstyle dimension.
- **Idle Time and Vulnerability:** There was a noticeable trend between increasing idle time and higher damage taken (see Figure 1), further supporting the decision to treat idle-vs-moving as a core playstyle.
- **Accuracy and Round Duration:** Higher accuracy was linked to longer survival, establishing its later weight in the PI calculation.
- **Thresholds:** Median pilot values were used to set initial classification boundaries. Metrics with insufficient variation were adjusted using domain knowledge.

2.3 Player Modelling and Countermeasure System Design

Building on from the foundation established during the pilot phase, a real-time player modelling pipeline and a dynamic countermeasure system was developed to personalise the gameplay and promote adaptation.

Composite Performance Index (PI): The Composite Performance Index (PI) was developed as the closest representation possible of a player's overall skill using a weighted sum of four core, standardised performance metrics: overall accuracy, total round time, average round time, and total distance travelled. This is based on similar approaches that use composite scores — specifically, the Expertise Performance Index (EPI) and the Maximum Similarity Index (MSI) — which aggregate standardised gameplay metrics using both PCA and domain-informed weighting to differentiate novice and expert performance in serious game environments [Loh and Sheng, 2015, 2014].

Initial Approach: PCA-Derived Weights: Using the results from the pilot study, we first used Principal Component Analysis (PCA) to assign weights to each metric for the composite score and establish a skill threshold of $PI_{\text{cut}} = -0.1727$. However, PCA-derived weights, while mathematically optimal to explain variance, did not fully align with our domain priorities.

Final Weights: Domain-Informed Approach: We ultimately chose to assign weights based on practical and pedagogical considerations:

- **Accuracy (30%):** Proportion of successful hits out of total shots fired, identified as the most relevant indicator of player skill.
- **Total Round Time (25%):** The total time taken to complete all waves.
- **Average Round Time (25%):** Mean time taken per wave, to account for pacing and consistency.
- **Total Distance Travelled (20%):** Measures player mobility, contributing to their tactical flexibility.

Each metric is first standardised (converted to a z-score) using the pilot sample mean and standard deviation:

$$Z_{M,i} = \frac{X_{M,i} - \mu_M}{\sigma_M}$$

The PI formula is then:

$$\text{PI}_i = 0.30\, Z_{\text{OverallAccuracy},i} + 0.25\, Z_{\text{TotalRoundTime},i} + 0.25\, Z_{\text{AverageRoundTime},i} + 0.20\, Z_{\text{TotalDistanceTraveled},i}$$

Because each $Z_{M,i}$ is zero-mean in the pilot sample, the composite PI is also centred at zero. Thus, we define the skill threshold as:

$$\boxed{\text{PI}_{\text{cut}} = 0}$$

Players with $\text{PI}_i \geq 0$ are classified as *high-skill*, and those with $\text{PI}_i < 0$ as *low-skill*.

Sensitivity Analysis: To ensure robustness, we also performed a sensitivity analysis by adjusting weights (e.g., $\pm 10\%$) and recalculating PI. The results showed that the ranking and separation between high- and low-performing players remained consistent, even with moderate weight changes. This demonstrates that our chosen composite PI is *robust* and does not depend heavily on the exact weights chosen. This procedure therefore builds confidence in both the interpretability and stability of the PI as a real-time skill metric.

Playstyle Metrics and Quadrant Classification: To capture playstyle, three key composite metrics were chosen, each with empirically or domain determined thresholds:

1. **Movement Playstyle:** Assessed using the Movement versus Idle ratio (MvI). The threshold was set at the median value (average of the 4th and 5th values) from the eight pilot participants' scores, resulting in:
$$\mathrm{MvI}_{\mathrm{cut}} = \frac{2.1747 + 2.2415}{2} \approx 2.2081$$
So, players with MvI ≥ 2.2081 are classified as moving, and those below as idle.
2. **Damage Playstyle:** Based on the Damage Taken per Minute (DPM). Again, the threshold was set at the median score of the eight pilot participants:
$$DPM_{\mathrm{cut}} = \frac{6.0310 + 6.2198}{2} \approx 6.1254$$
So, players with DPM ≥ 6.1254 were classified as high damage, and those below as low damage.
3. **Weapon Usage Playstyle:** Defined by Weapon Switch Rate per Minute (WSR). Since only two waves were played in the pilot, most participants had insufficient time to purchase or switch weapons, making the pilot data for this metric unreliable. Therefore, the threshold for WSR was determined based on domain knowledge, with WSR ≥ 2 classified as high switching and values below as low switching.

- **Implementation Notes:** Each intervention modifies only a *single* gameplay parameter (mobility, defense, or loadout) to prevent conflicting and confounding effects. Countermeasures are applied immediately at the start of the intervention wave (3rd of 5 waves) and persist for its duration. This independent-quadrant approach ensures clear, non-overlapping "gameplay lanes" for analysis.

Table 1: Thresholds for playstyle metrics derived from pilot data and domain knowledge.

Metric	Threshold
Composite PI	0
Damage per Minute (DPM)	6.1254
Move–Idle Ratio (MvI)	2.2081
Weapon Switch Rate (WSR)	2.00

Each playstyle metric was paired with the skill classification (high/low PI) to assign the player to one of four quadrants for each playstyle, e.g.

- High Skill / Moving
- High Skill / Idle
- Low Skill / Moving
- Low Skill / Idle

and similarly for damage and weapon switching metrics, see Figures 2 to 4.

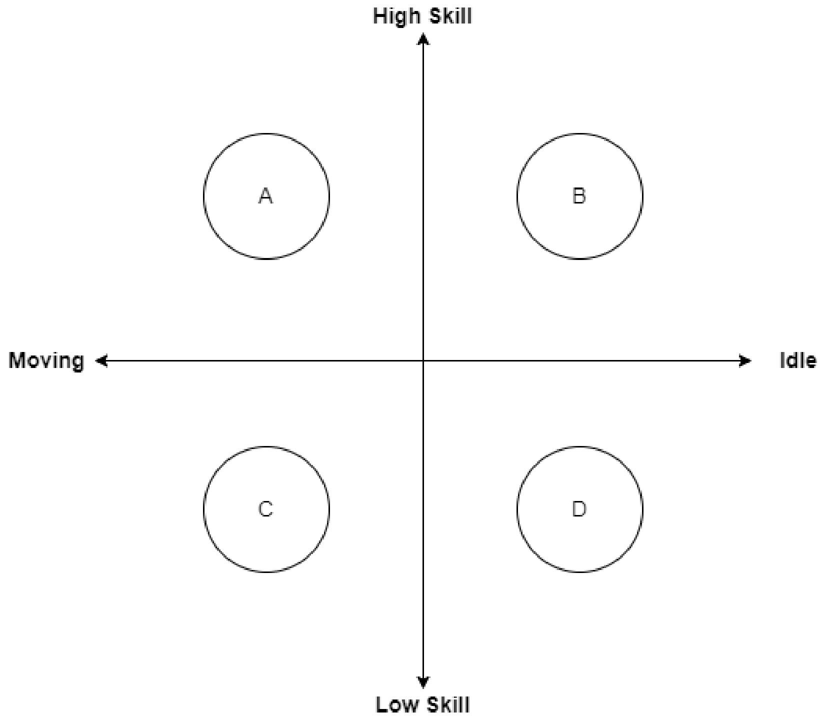

Figure 2: Movement playstyle quadrant classification (threshold: 2.2081 MvI).

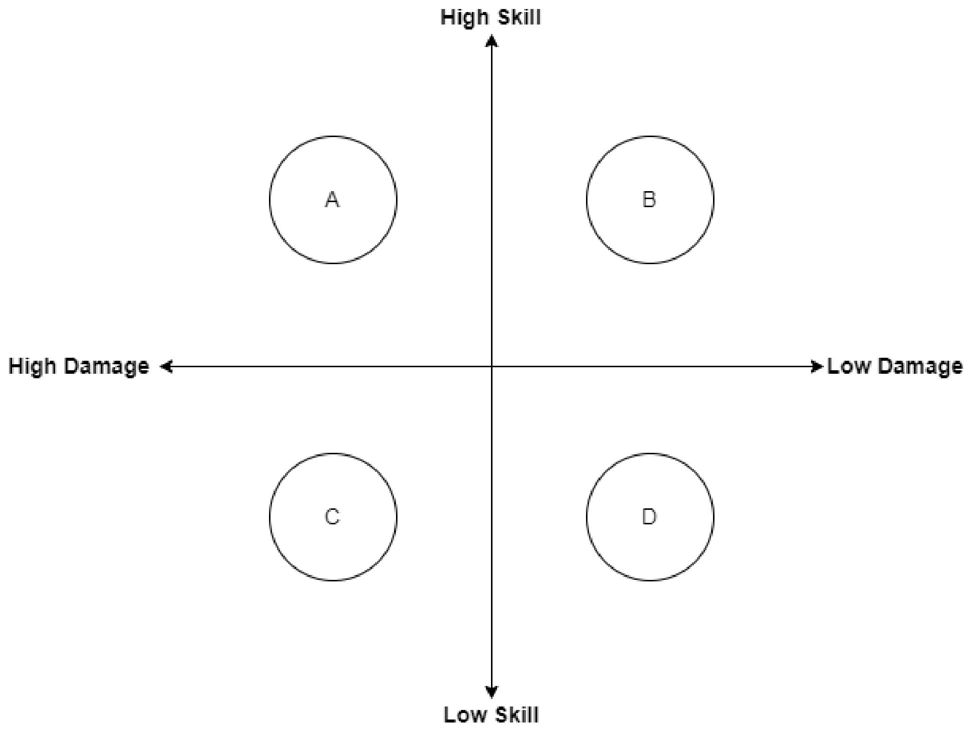

Figure 3: Damage playstyle quadrant classification (threshold: 6.1254 DPM).

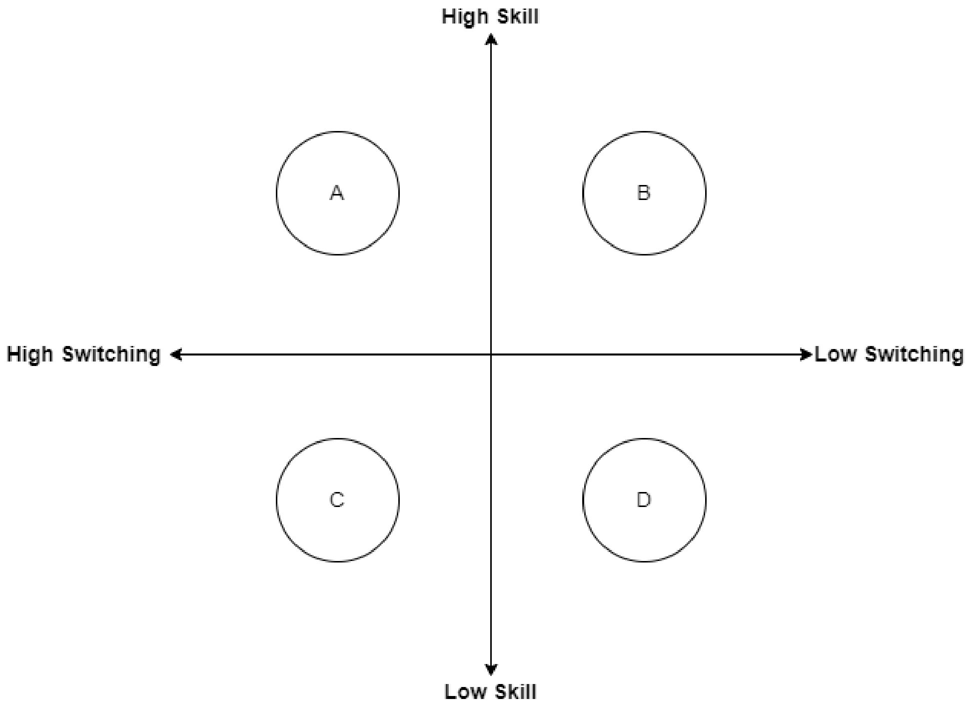

Figure 4: Weapon usage playstyle quadrant classification (threshold: 2 WSR).

Countermeasure Design and Deployment: For each of these three playstyle metrics, a specific countermeasure was designed for each quadrant A to D, resulting in a total of twelve distinct countermeasures. These countermeasures were intended to nudge players towards the threshold — encouraging less habitual behaviour, increased adaptation, and ultimately supporting skill development and exploration.

For example, a player classified as "low skill, moving" would periodically see additional rock cover spawned around the map to encourage more careful movement, while a "high skill, idle" player would encounter fast-moving "dart" enemies to force greater mobility. Weapon and resource drops, enemy types, and temporary buffs/debuffs were all dynamically adjusted based on the player's quadrant assignment.

Tables 2 to 4 summarise all countermeasures mapped to skill–playstyle quadrants.

	Low Skill	**High Skill**
Idle	Movement boost pickups will spawn to encourage players to move.	Adds to enemy spawns, fast "dart" enemies (1.5× move speed), forcing player to stay mobile.
Moving	Spawn rock cover, periodically around map, encouraging player to move less and more carefully.	Enemies will spread out from each other to prevent players from herding them.

Table 2: Movement playstyle countermeasures by skill quadrant.

	Low Skill	High Skill
Low Damage	Spawn damage boost pickups.	Add a single heavy "mini-tank" enemy (3× HP).
High Damage	Reduce enemy spawn rate (20% spawn-interval reduction).	Taking damage lowers player speed and reduces visibility (blood on screen) for a few seconds.

Table 3: Damage playstyle countermeasures by skill quadrant.

	Low Skill	High Skill
Low Switching	Spawn weapon pickups (guns purchasable from the store).	Most used weapon gets a damage debuff and the ammo magazine capacity is halved.
High Switching	Swap bonus: when player equips a previously unused weapon, a temporary damage buff is applied to it.	Lock weapon swap for a short window to force players to improvise with a single weapon.

Table 4: Weapon usage playstyle countermeasures by skill quadrant.

Each player received one active countermeasure for each playstyle axis, resulting in a total of three possible interventions. Note, any countermeasure in the low skill column was designed as an assistive playstyle to help players improve while trying to encourage them to move towards alternate styles, whereas any countermeasure in the high skill column was designed as a disruptive countermeasure intended to make continuing playing in that playstyle a hindrance in order to encourage them to experiment with the alternate. The logic was designed not only to sustain optimal engagement and flow, but also to gently nudge players toward experimenting with new strategies or adapting their existing ones.

All model outputs, surveys, and game state changes were logged in real time for subsequent analysis, enabling a detailed evaluation of both the player modelling accuracy and the efficacy of countermeasures.

The model reclassified players after giving them time to react to the countermeasures, allowing for adaptive responses throughout the session.

2.4 Experiment Setup

The main experiment had a total of 30 participants, divided evenly between two phases (15 per phase). Each participant completed a pre-game survey, played five consecutive waves of *3D Wave Shooter* with post-wave (mid-game) surveys and a post-game survey, and experienced the real-time adaptation logic as follows:

Wave Sequence and Adaptation:

1. **Pre-Game Survey:** Participants completed a consent form and provided their gaming experience.
2. **Waves 1 & 2:** No countermeasures were applied. The game was played at base difficulty, with slight increases per wave as in a typical wave shooter of this kind. These waves were used solely to gather baseline metrics for skill and playstyle classification.
3. **Wave 3 & 4:** Countermeasures were applied based on the player's skill level and playstyle quadrants determined from their metrics in waves 1 and 2.
4. **Wave 5:** Final set of countermeasures, now based on updated classifications from waves 3 and 4, to allow players the opportunity to adapt and experience new interventions.
5. **Mid-Game Surveys:** After each wave, players rated the earlier identified core flow components (challenge, excitement, satisfaction, frustration, boredom, control) on a Likert scale of 1–7.
6. **Post-Game Survey:** After wave 5, a final survey assessed the overall experience and allowed for any additional comments to be given.

Wave and Shop Calibration: To ensure accurate, consistent, and interpretable results, we carefully calibrated game waves and shop mechanics:

- **Round Design:** The experiment used five rounds: two for baseline measurement, three for post-intervention adaptation.
- **Health and Money:** Player health was removed to ensure players would play all 5 waves for comparability; but to maintain stakes of receiving damage, monetary penalties for taking damage were applied.
- **Shop Simplification:** Only weapon purchases and ammo were available, removing purchasable stat buffs to avoid confounds and ensure accurate measurability of playstyle metrics.
- **Difficulty Scaling:** Enemy compositions, coin rewards and deductions, as well as shop prices were tuned such that the average player could pursue multiple viable strategies (dual weapon unlock, ammo focus, long-term savings) and encounter steadily rising but consistent challenge. Wave durations were pilot-tested to take between 45–60 seconds to complete.

In-depth details of enemy balancing, coin budgeting, weapon DPS calibration, and all parameter settings are provided in Appendix A.

Thresholds for Skill and Playstyle Metrics:

Phase 1: - As mentioned, the threshold for PI was centred at 0. For the playstyle metrics, thresholds were initially determined from the medians of pilot data, except for weapon switch rate which was set by domain knowledge due to insufficient pilot switching:

- **Composite Performance Index (PI):** 0
- **Damage per Minute (DPM):** 6.1254
- **Movement versus Idle Ratio (MvI):** 2.2081
- **Weapon Switch Rate (WSR):** 2 (domain knowledge)

Phase 2: Using the full dataset from Phase 1 participants, thresholds were recalculated in order to better align with the actual median performance across all players, thereby improving the accuracy and fairness of player classification. Only those that differed substantially from pilot medians were changed:

- **PI:** Updated to -0.987. Figure 5 visualises the sorted $\mathrm{PI}_{\mathrm{initial}}$ scores for all Phase 1 players, with the colouring indicating k-means skill clusters. Several candidate thresholds are displayed:
  - **Median (–0.987):** Provides the most balanced split (8 "low skill" vs. 7 "high skill"), and is easy to justify statistically.
  - **Cluster midpoint (–0.854):** Maximised the silhouette score (0.45), indicating well-separated clusters, but produced an unbalanced split.
  - **Other percentiles (40th: –1.064, 60th: –0.950):** Evaluated for robustness.

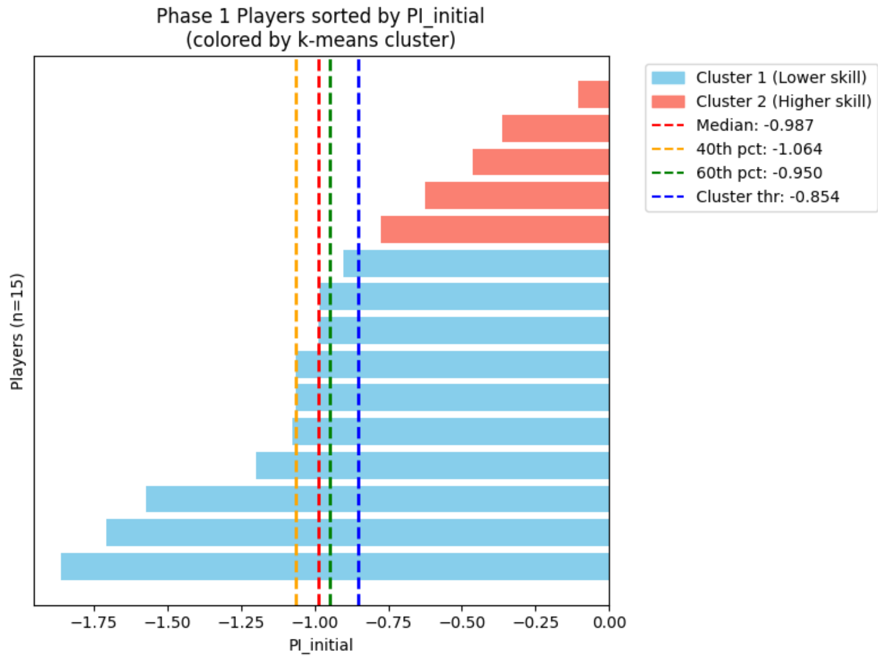

Figure 5: Phase 1 players sorted by initial PI, coloured by k-means cluster. Vertical lines indicate candidate thresholds: median (red), 40th percentile (orange), 60th percentile (green), and cluster midpoint (blue).

Although the cluster midpoint (-0.854) yielded the highest silhouette score (0.45), the median split (-0.987) provided the most balanced division (8 vs. 7) and was straightforward to justify. Thus, the threshold of the PI was updated from 0 to the phase 1 median of -0,987 to ensure a fairer classification of high and low skilled players.

- **DPM:** Remained at 6.12 (median remained consistent with pilot)
- **MvI:** Remained at 2.21 (median similar consistent with pilot)
- **WSR:** Updated to 9.92 (as this median from Phase 1 was much higher and the initial value was set from domain knowledge)

This two-phase structure allowed for empirical refinement as well as validation of the adaptation system.

All gameplay data, classifications, countermeasure activations, and survey responses were logged for analysis.

2.5 Survey Instruments

Three sets of surveys were administered to participants during the experiment: a pre-game survey, a set of mid-game surveys, and a post-game survey. These were designed respectively to assess player background, capture real-time experiential data, and collect summative feedback.

Pre-Game Survey:

- Informed consent statement.
- Demographic information (age, gender, etc.).
- Gaming experience (years of gaming, frequency, typical genres played).
- Self-assessed skill level (Likert scale).

Mid-Game (Post-Wave) Survey: After each wave, participants rated the following components on a 1–7 Likert scale (1 = strongly disagree, 7 = strongly agree):

- I felt challenged during the last wave. [*Challenge*]
- I felt excited while playing. [*Excitement*]
- I felt satisfied with my performance. [*Satisfaction*]
- I felt in control of the game. [*Control*]
- I felt frustrated during the last wave. [*Frustration*]
- I felt bored during the last wave. [*Boredom*]

(See Appendix B for full survey text.)

Post-Game Survey: After completing all five waves, participants completed a post-game survey, including:

- Overall enjoyment.
- Perceived fairness of the game.
- Perceived adaptivity/personalisation of the game.
- Open-ended feedback on experience, adaptation, and suggestions for improvement.

All survey responses were timestamped and linked to the corresponding game data for each participant.

3 Results

This section presents the findings from the main experiment. Descriptive analyses are followed by visualisations of how player metrics and skill levels changed in response to the adaptive countermeasures.

3.1 Metrics Across Waves

Figure 6 plots the average values of the key gameplay metrics across all five waves for all 30 players. The vertical red dotted line at wave 2 marks the point after which the adaptive countermeasures are introduced.

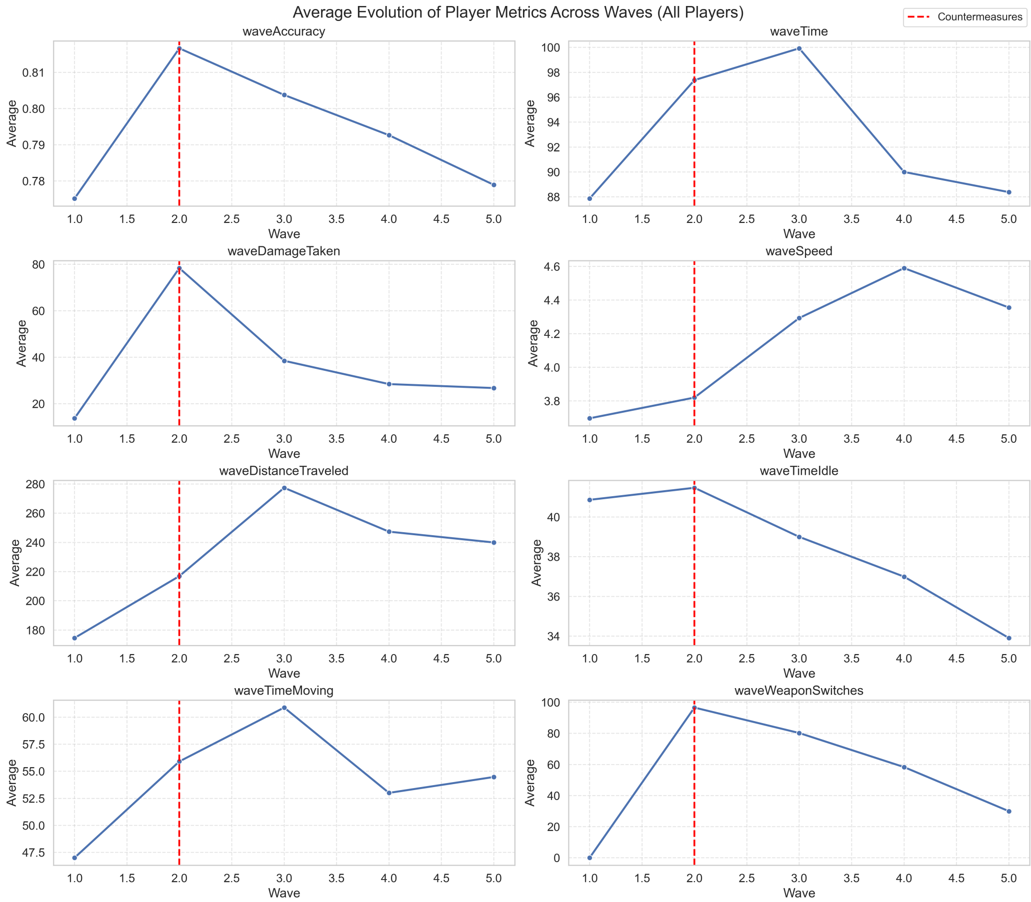

Figure 6: Mean gameplay metrics by wave. (Red vertical line indicates when countermeasures were first applied.)

Prior to the introduction of countermeasures (waves 1 and 2), accuracy, damage taken, and weapon switching steadily increased and by the end of wave 2 had peaked. Immediately following the introduction of countermeasures (waves 3–5), all three showed a notable decline. In the case of damage taken and weapon switching, these are both part of composite playstyle metrics, meaning the interventions are catered towards countering playstyles identified from these metrics, which likely suggests the countermeasures did in fact hinder player's favoured playstyle. However, the reduction in accuracy observed in the later waves, potentially reflects the disruptive effect of the countermeasures on high-skill players, who then faced greater challenges thereby affecting their accuracy.

3.2 Skill Level Changes by Initial Skill Group

Figure 7 neatly shows how player skill levels evolved from before to after the countermeasures, grouping players by their initial skill classification (high or low skill). The bars show the number of players who improved, worsened, or crossed the skill threshold after countermeasures.

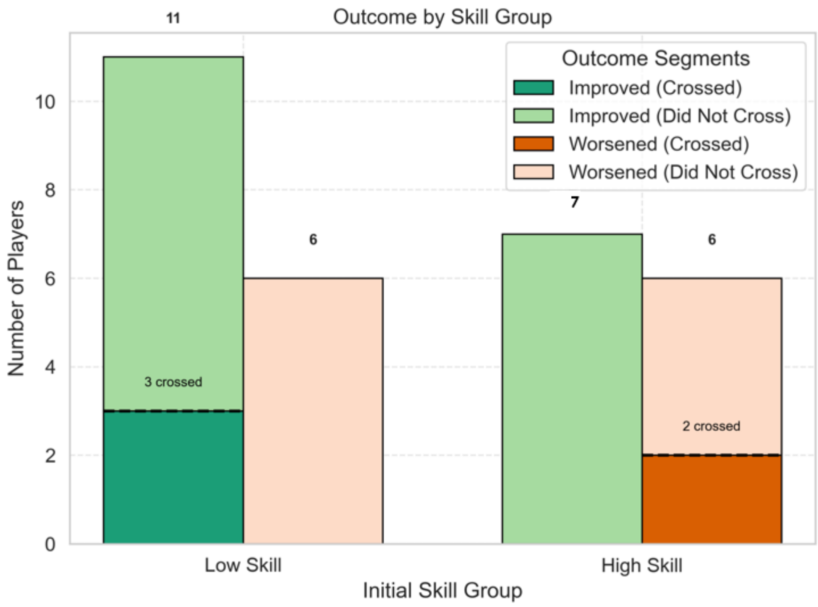

Figure 7: Number of players whose skill level improved, worsened, or crossed threshold after countermeasures, grouped by initial skill classification.

- For initially low-skill players, the countermeasures appeared to have a positive effect as the majority improved their skill over the session, with 11 showing improvement and 3 crossing over into the high skill group. Only 6 worsened.
- For initially high-skill players, 7 improved and 6 worsened, with 2 dropping below the high skill threshold. This more balanced distribution could suggest that disruptive countermeasures succeeded in challenging high-skill players without causing an excessive difficulty increase.

3.3 Distribution of Skill Change

Figure 8 presents box and whisker plots of the change in the Composite Performance Index (PI) for both initial low-skill and high-skill players.

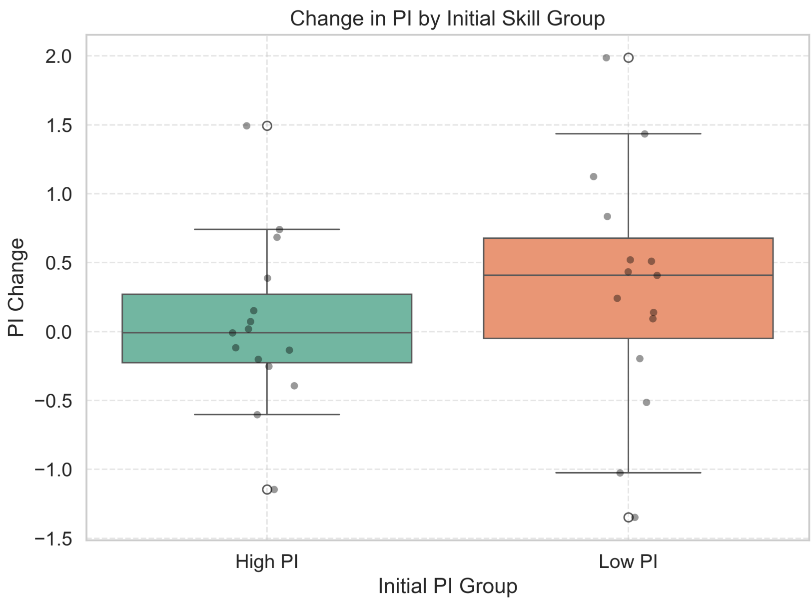

Figure 8: Distribution of change in Composite Performance Index (PI) for low-skill and high-skill players.

- Low-skill players exhibited a larger variation in PI (and also, as we saw in the previous figure, most players exhibit a positive change value), suggesting that assistive countermeasures effectively promoted player skill improvement.
- High-skill players showed more stable PI change distribution, with changes split more evenly between positive and negative, likely indicating on average the disruptive countermeasures were challenging but not disproportionately disadvantageous.

4 Discussion

4.1 RQ1: (Real-Time Challenge Adaptation): What is the impact of real-time adaptive countermeasures on player experience and the core components of flow?

Sub-question 1 (Psychological Flow): How does real-time challenge adaptation influence key psychological components of flow?

To answer this, we analysed the mid-game survey data where players rated their experience on a 1–7 Likert scale across the core metrics: challenge; excitement; satisfaction; control; boredom; and frustration after each wave.

Figure 9 presents the average of these ratings across all five waves.

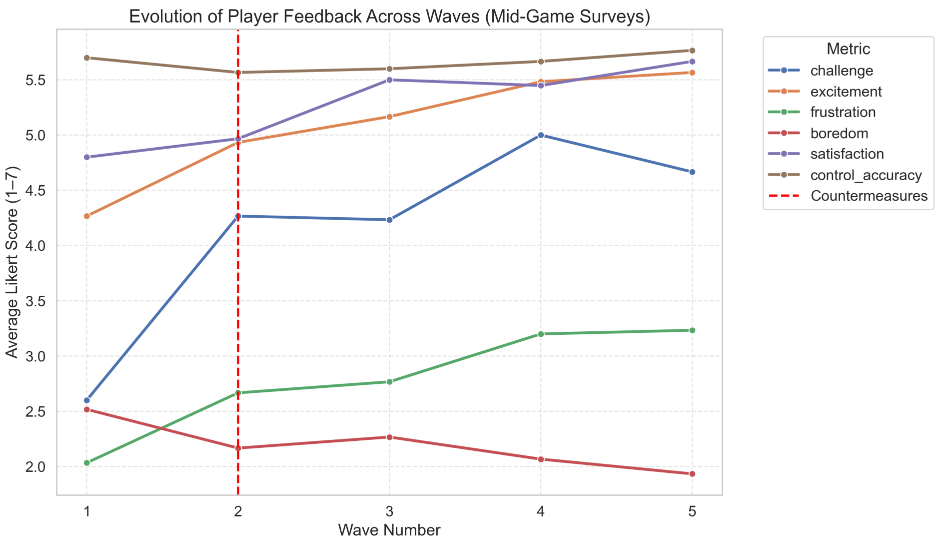

Figure 9: Mean self-reported challenge, excitement, satisfaction, control, boredom, and frustration across the five waves. Red dotted line marks the start of adaptive countermeasures.

- **Challenge:** Ratings consistently rose up to 2, reflecting the increasing enemy spawn rate and general difficulty increase as gameplay progressed. After the introduction of countermeasures, challenge ratings plateaued and starting stabilising more in the median between 1 and 7, potentially indicating the adaptive system stabilised perceived challenge overall for both high and low skill players.
- **Excitement, Satisfaction, and Control:** These metrics did not show any substantial trends with challenge, except for a slight one with excitement, but not enough to conclude anything.
- **Boredom and Frustration:** With regards to these metrics and their relation to the challenge, boredom appeared to decrease as challenge increased, and both

plateaued in tandem after countermeasures took effect. This could imply that players responded well to the increase in challenge and found it stimulating enough to subdue boredom. Frustration, by contrast, rose in nearly exact inverse proportion to boredom. This relationship is perfectly in line with the flow theory, which, as mentioned, suggests that as games become more challenging, players become less bored; but if challenge exceeds skill, frustration rises.

To further connect these results to flow theory, we roughly modelled the flow experienced by plotting the average Composite Performance Index (PI, as a proxy for skill) for waves 1 and 2 (pre-countermeasures) and waves 3 and 4 (with countermeasures) against the average challenge rating corresponding to those waves, separately for high- and low-skill players (Figure 10).

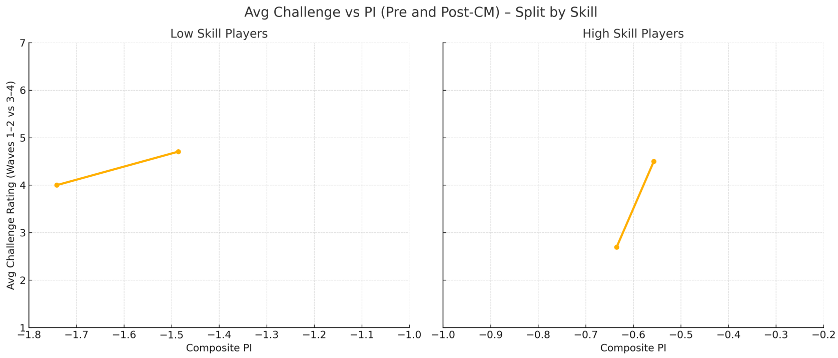

Figure 10: Mean skill (PI) vs. mean challenge, before and after countermeasures, for high- and low-skill players.

Due to the incomparable values of the axes, the results of this graph are limited, however, from what can be seen: for low-skill players, skill scores increased at a higher rate than challenge after countermeasures, suggesting a tendency toward the boredom side of flow. In contrast, for high-skill players, challenge ratings increased more than skill, pushing them toward the frustration side. This can be interpreted as a positive outcome as it indicates the countermeasures were over-effective, where the assistive countermeasures assisted to the point of boredom and the disruptive ones disrupted to the point of frustration. So if the countermeasures do indeed have the desired effect, but to an excess, this is modifiable, and future tuning of the countermeasures could bring both groups closer to optimal flow.

Overall, these findings demonstrate that the real-time adaptive system did, to an extent, stabilise challenge and also produced patterns in line with the flow theory regarding boredom and frustration, thereby validating the methodology as well as the underlying modelling of player skill and playstyle.

Sub-question 2 (Affective Response): How do these components evolve across the play experience, and do adaptive countermeasures impact them differently for low- and high-skill players?

To investigate how the psychological components of flow evolved throughout the play session and whether adaptive countermeasures had different effects on low- and high-skill players, mid-game survey responses were analysed by skill stability group (referring to those who remained high skill throughout, those who remained low skill, and those who switched skill group either from low to high or vice versa).

Figure 11 shows the evolution of the core self-reported metrics (challenge, excitement, satisfaction, control, boredom, frustration) across the five waves, plotted separately for each skill stability group. Also note for the y-axes on the figure below, they are scaled differently for each metric, as we are interpreting the change in the values across the waves rather than the actual values themselves.

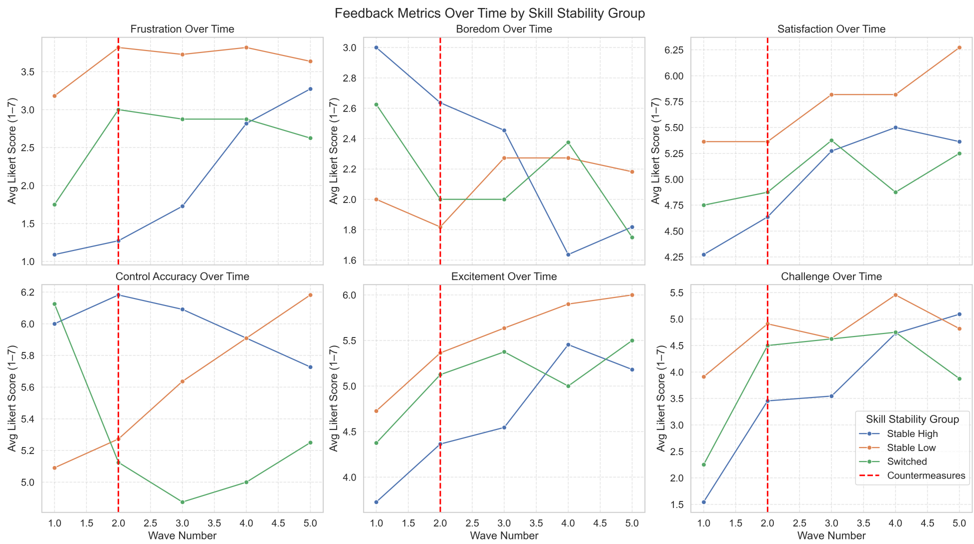

Figure 11: Evolution of self-reported flow metrics by skill stability group across five waves.

- **Frustration:** Among high-skill players, frustration rose sharply after the introduction of countermeasures, further confirming the earlier assumption that disruptive countermeasures were effective in challenging this group and pushing them toward the frustration side of the flow.
- **Boredom:** For low-skill players, we again see they experienced somewhat of an increase in boredom post-countermeasures, indicating that assistive interventions may have reduced challenge too much, nudging these players towards the boredom side.
- **Excitement and Satisfaction:** Across all groups, excitement and satisfaction generally increased after the countermeasures took effect, even for high-skill players, which may indicate that, even if frustrated, these players still found the added challenge somewhat engaging.

- **Control:** Notably, low-skill players reported increasing feelings of control after countermeasures, while high-skill players felt less control. This nicely indicates the distinction between the assistive and disruptive countermeasures.

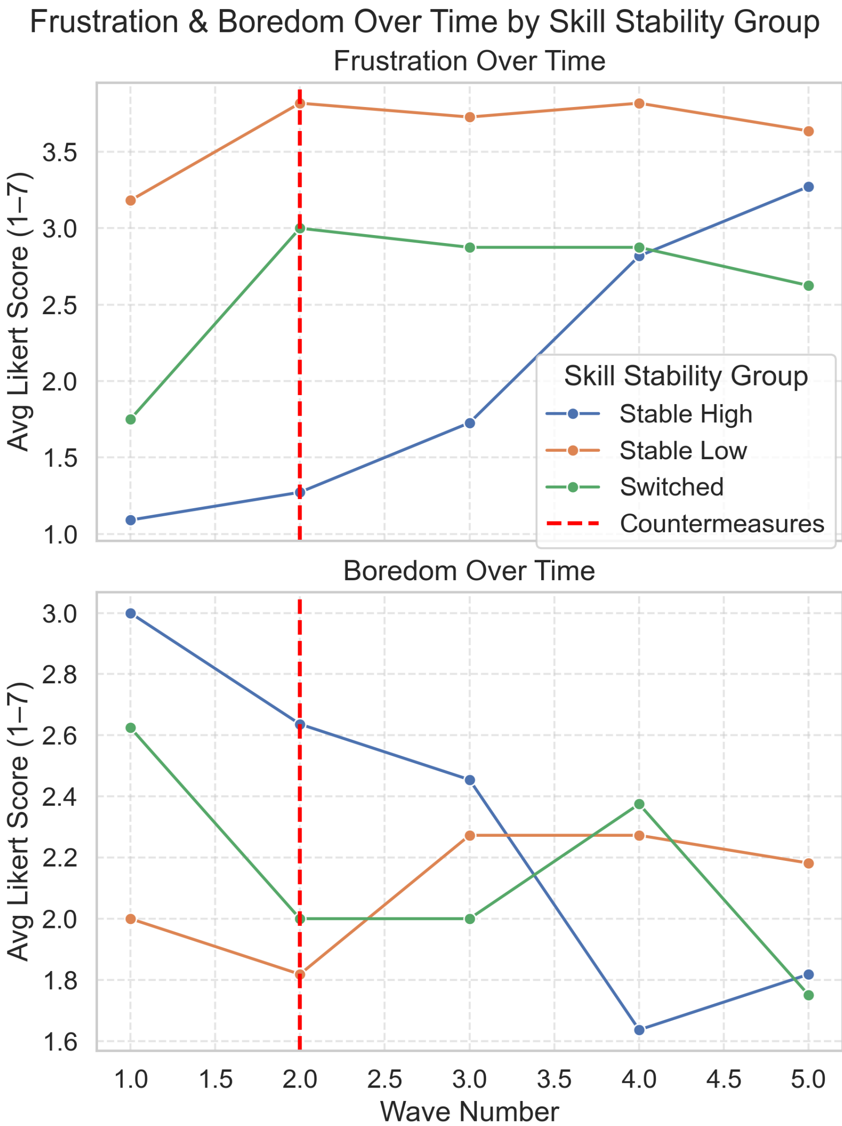

Figure 12: Evolution of self-reported boredom and frustration by skill stability group across fives waves.

Focusing specifically on boredom and frustration, seen in figure 12 above, we see that for high-skill players, these metrics tracked in almost perfect opposition: as boredom decreased, frustration rose, perfectly validating the presence of flow theory in this study.

Interestingly, however, while frustration increased more for high-skill player than for low-skill players, high-skill players did not reach the level of frustration which low-skill players did. This could be explained by the psychological effect of frustration: once players become frustrated early on (likely as a result of too high of a challenge for low-skill players before countermeasures), it is difficult to return to a state of flow or lower frustration especially within the same play session.

In summary, adaptive countermeasures had differentiated effects on player groups: disruptive interventions successfully challenged and engaged high-skill players (at the cost of increased frustration), while assistive interventions helped low-skill players feel more in control, albeit sometimes at the risk of boredom. These findings highlight the need for further tuning to ensure both groups are kept within the optimal flow channel throughout play.

4.2 RQ2: (Refined Modelling): Which combination of performance metrics and self-reported surveys yields the most accurate real-time model of the player's skill state and receptiveness to new playstyles and strategies?

To answer this research question, we implemented an end-to-end modelling pipeline using a Random Forest classifier with 100–500 trees, tuned via grid search and stratified 5-fold cross-validation plus permutation testing. Random Forest was chosen for its ability to handle non-linear interactions between features and its robustness to overfitting.

The modelling pipeline was as follows:

- **Input features:** Pre-countermeasure (Waves 1–2) performance metrics (*PI*, *MvI*, *DPM*, *WSR*) and mid-game survey responses.
- **Target labels:** Post-countermeasure (Waves 3–4) binary skill and receptiveness states. For *skill*, the already established PI score and threshold were used. For *receptiveness*, a similar score was constructed by taking the equally weighted combination of the z-scored changes (Δ) in the three composite metrics (*MvI*, *DPM*, *WSR*):

  $$\text{receptivenessScore}_i = \frac{z(\Delta\text{MvI}_i) + z(\Delta\text{DPM}_i) + z(\Delta\text{WSR}_i)}{3}.$$

  The threshold for receptiveness classification was set as the median of this score across all players.
- **Training:** Models trained on Phase 1 data; best feature sets identified by ablation/blocking.
- **Validation:** Frozen models tested on Phase 2 (hold-out) data.

To avoid information leakage, only pre-CM data (waves 1–2) was used as input for predicting post-CM states (waves 3–4).

Ablation Results:

We conducted systematic ablation tests to identify the most predictive combination of performance metrics to ensure only the minimal and most useful input data was used. The tables below show Area Under Curve (AUC) scores for each performance feature combination.

Table 5: AUC for different performance feature sets in skill prediction

Features	AUC
PI Only	**0.95**
DPM Only	0.90
All Composites	0.90
All Raw Metrics	0.80
Composites + Raw Metrics	0.90

Table 6: AUC for different performance feature sets in receptiveness prediction

Features	AUC
All Composites	**0.90**
MvI Only	0.85
WSR Only	0.60
All Raw Metrics	0.40
Composites + Raw Metrics	0.70

We then compared models using: the best identified combination of performance features by themselves; the survey responses by themselves, and both combined:

Table 7: Comparison of AUC: performance only, survey only, and combined models

Model	Perf Only	Survey Only	Combined
Skill	**0.950**	0.900	0.800
Receptiveness	**0.900**	0.400	0.900

Hold-out Validation (Phase 2):

When the best-performing models were evaluated on new, unseen Phase 2 data, the results were as follows:

Table 8: Phase 2 holdout evaluation metrics for skill and receptiveness models

Target	Accuracy	Precision	Recall	F1	ROC-AUC
Skill	0.80	0.67	0.80	0.73	0.88
Receptiveness	0.73	0.73	0.89	0.80	0.86

Interpretation and Key Findings:

- **Skill:** The PI composite alone achieved the highest AUC (0.95) for skill prediction, with survey data offering little additional benefit. This is logical as is implies early-game PI is a robust indicator of post-countermeasure skill state.
- **Receptiveness:** The set of composite playstyle (*MvI, DPM, WSR*) together best predicted receptiveness, with survey-only models performing substantially worse and both combined offering no additional benefit to just the performance features alone while coming at a higher cost.
- **Survey Data:** Survey responses added little to no predictive power over performance data alone, particularly for receptiveness.
- **Generalisation:** Both models retained strong predictive performance on new participants in Phase 2, indicating good external validity with room for improvements.
- **Practicality:** With just four lightweight features per wave, the system could be deployed in real-time to continuously predict player skill and receptiveness during live gameplay.

4.3 RQ3: (Playstyle Experimentation): Which playstyle's countermeasure had the most significant effect on encouraging players to refine, evolve, or change from their preferred playstyle?

To address this final research question, we analysed how many players moved towards, crossed, or moved away from each playstyle threshold as a result of the deployed countermeasures.

Analysis and Findings:

For each core playstyle metric (Movement, Damage Taken, Weapon Switching), players were classified according to their position relative to the metric's threshold both before and after countermeasures. The groups were:

- **Moved closer to threshold:** Changed behaviour towards the desired (more adaptive) range, but did not cross.
- **Crossed threshold:** Switched from one side of the playstyle threshold to the other (e.g., from idle to moving).
- **Moved away from threshold:** Became more entrenched in their original playstyle.

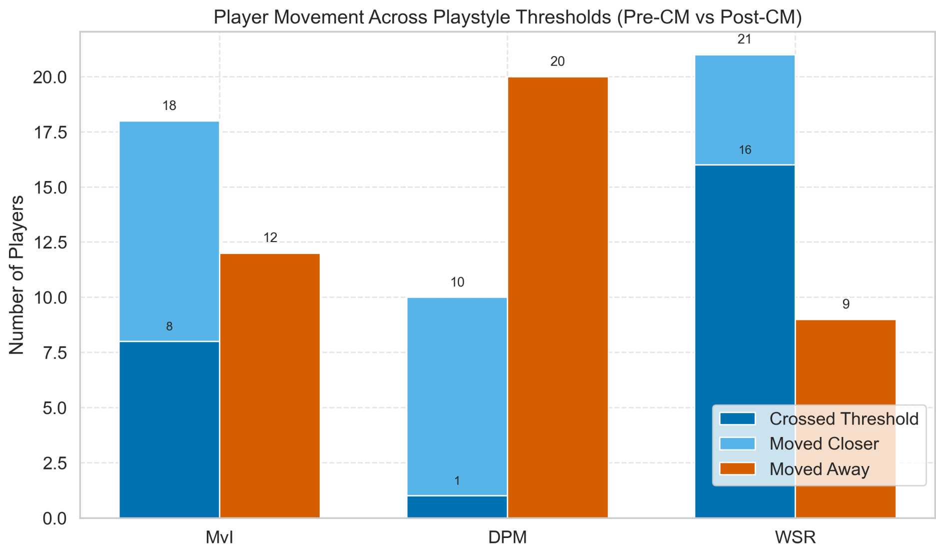

Figure 13: Number of players who moved closer to, crossed, or moved away from each playstyle threshold after countermeasures.

Key Results:

- **Weapon Switching Countermeasure:** Most effective — **21** players moved closer to the switching threshold, with **16** of those crossing it, and only **9** moved away.
- **Movement (MvI) Countermeasure:** Also effective — **18** moved closer, with **8** of those crossing, and **12** moving away.
- **Damage Taken (DPM) Countermeasure:** Largely counterproductive as majority of players moved away (**20**), and only a third (**10**) moving closer with just **1** crossing.

These results suggest the weapon switching countermeasures had the most pronounced and consistent effect in encouraging players to experiment or adapt, while the damage-based measures often had the unintended consequence of reinforcing entrenched playstyles.

Interpretation:

- **Weapon Switching:** Countermeasures designed to push players to experiment with more weapons proved most effective, with the majority adapting or crossing into the opposite playstyle. These results suggest such adaptive responses are especially effective in dynamic, multi-weapon shooter games.
- **Movement:** The movement based interventions were also fairly effective, encouraging nearly two-thirds of participants to adapt or cross the movement threshold.
- **Damage Taken:** Measures targeting adaptation of damage taken appear to require redesigning or more nuanced tuning, as they were by far the least effective at breaking entrenched defensive or aggressive habits.

Overall, these results show that adaptive countermeasures, especially those targeting weapon usage and movement, can successfully drive playstyle experimentation and player adaptation in real time in games of this genre.

Summary of Main Findings

In summary, this study, within its bounds, demonstrated that real-time adaptive countermeasures can effectively encourage player skill and playstyle adaption in a live game environment, with assistive and disruptive interventions impacting engagement and flow as predicted by theory. Furthermore, the lightweight modelling pipeline showed strong generalisation to new players with plenty of room for more potential improvements. And finally, targeted countermeasures — particularly those related to weapon switching and movement — were most successful at encouraging adaptation and experimentation. However, some measures, such as those targeting damage taken, require further refinement.

5 Conclusion

This study demonstrates that real-time adaptive countermeasures can effectively influence both player skill as well as playstyle in a dynamic game environment. The system's integration of composite performance metrics, continuous feedback, and targeted interventions not only supported classic patterns predicted by flow theory, but also facilitated meaningful adaptation among players. The modelling results also show that a small set of gameplay features are sufficient to robustly predict both skill and receptiveness to change, providing a promising foundation for real-time player modelling in games with immense potential for extensions and improvements.

5.1 Limitations

To fully contextualise these findings, several limitations must be acknowledged:

- **Sample Size:** The study involved a relatively small pool of 30 participants (plus 8 pilot testers). This limited the overall statistical power of the results and prevented the use of parametric hypothesis testing, such as t-tests, due to non-normal data distributions.
- **Playstyle Categorisation:** The approach taken in this study to model and classify playstyle, while interpretable and logical, may still be overly simplistic to be able to fully capture the full spectrum or nuance of player behaviours. More granular or data-driven models could provide more accurate and richer insights into the most prevelant playstyles.
- **Controlled Game Design:** The experiment was conducted within a single, controlled wave-based shooter. As a result, the generalisability of the findings to other game genres or more open environments is uncertain without extensions to the study.
- **Long-Term Effects:** Since all data was gathered within a single play session, there was no assessment or insight into the long-term learning or retention of adapted behaviours.

5.2 Future Work

Several promising aspects of this study appear ripe for future research:

- **Physiological and Affective Data:** Incorporating physiological signals (e.g., heart rate, facial tracking, or EEG) could enable more reliable, real-time detection of player affect and engagement, beyond self-reported surveys.

- **Richer Playstyle Modelling:** Applying advanced machine learning methods, such as reinforcement learning or unsupervised clustering, may allow for finer-grained and more dynamic identification of player skill and playstyles.
- **Generalisability:** Testing the adaptive framework in different game genres — such as stealth, racing, or puzzle games — will be critical to assessing its broader applicability.
- **Long-Term Adaptation:** Extending a similar experiment across multiple play sessions would enable a clear assessment of whether players actually retain new skills or behaviours learned from adaptive countermeasures.

This work provides a step towards real-time, personalised game experiences that adapt not only to player skill but also to playstyle, supporting both engagement and player development.

References

Boyan Bontchev. Holistic player modeling for controlling adaptation in video games. In *Proceedings of the International Conference on e-Society*, 2016.

Daniel Buckley. 3d wave shooter. https://assetstore.unity.com/packages/templates/packs/3d-wave-shooter-133749, 2025. Unity Asset Store, Version 2.1, Released March 31, 2025.

Darryl Charles and Michaela Black. Dynamic player modeling: A framework for player-centered digital games. In *Proceedings of the International Conference on Computer Games: AI, Design and Education*, 2004.

Ben Cowley, Darryl Charles, Michael Black, and Ray Hickey. Toward an understanding of flow in video games. *Computers in Entertainment*, 6(2):1–27, 2008.

Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper and Row, 1990.

Sehar Shahzad Farooq and Kyung-Joong Kim. *Game Player Modeling*, pages 774–778. Springer International Publishing, Cham, 2024. doi: 10.1007/978-3-031-23161-2_14. URL https://doi.org/10.1007/978-3-031-23161-2_14.

Reza Habibi, Johannes Pfau, and Magy El-Nasr. Modeling player personality factors from in-game behavior and affective expression. *arXiv preprint*, 2023. doi: 10.48550/arXiv.2308.14224.

Danial Hooshyar, Moslem Yousefi, and Heuiseok Lim. Data-driven approaches to game player modeling: A systematic literature review. *ACM Computing Surveys*, 50(6):90, 2018. doi: 10.1145/3145814.

Md Mosharaf Hossan, Mostafa M. Fouda, and Farjana Z. Eishita. Adaptive game design using machine learning techniques: A survey. In *2024 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, pages 144–150, 2024. doi: 10.1109/IoTaIS64014.2024.10799330.

Christian Sebastian Loh and Yanyan Sheng. Maximum similarity index (msi): A metric to differentiate the performance of novices vs. multiple-experts in serious games. *Computers in Human Behavior*, 39:344–356, 2014. doi: 10.1016/j.chb.2014.07.022. URL https://doi.org/10.1016/j.chb.2014.07.022.

Christian Sebastian Loh and Yanyan Sheng. Measuring expert performance for serious games analytics: From data to insights. In Christian S. Loh, Yanyan Sheng, and Dirk Ifenthaler, editors, *Serious Games Analytics*, pages 67–86. Springer International Publishing, Cham, 2015. doi: 10.1007/978-3-319-05834-4_5. URL https://doi.org/10.1007/978-3-319-05834-4_5.

Marlos C. Machado, Eduardo P. C. Fantini, and Luiz Chaimowicz. Player modeling: Towards a common taxonomy. In *2011 16th International Conference on Computer Games (CGAMES)*, pages 50–57, 2011. doi: 10.1109/CGAMES.2011.6000359.

Kun Shao, Zhentao Tang, Yuanheng Zhu, Nannan Li, and Dongbin Zhao. A survey of deep reinforcement learning in video games. *CoRR*, abs/1912.10944, 2019. URL http://arxiv.org/abs/1912.10944.

Penelope Sweetser and Peta Wyeth. Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, 3(3):3–3, 2005.

Hongyou Xiong. Dynamic difficulty adjustment: Developing an adaptive game ai with machine learning. Master's thesis, California State University, Northridge, 2019.

Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. Player modeling. In *Artificial and Computational Intelligence in Games*, 2013. URL https://api.semanticscholar.org/CorpusID:590230.

Tian Zheng. Dynamic difficulty adjustment using deep reinforcement learning: A review. *Applied and Computational Engineering*, 71(1):157–162, 2024. doi: 10.54254/2755-2721/71/20241633. License: CC BY 4.0.

Mohammad Zohaib. Dynamic difficulty adjustment (dda) in computer games: A review. *Advances in Human-Computer Interaction*, 2018:5681652, 2018. doi: 10.1155/2018/5681652. URL https://doi.org/10.1155/2018/5681652.

A Game Calibration and Balancing

A.1 Test Structure and Rationale

To ensure robust and interpretable results as well as a fair comparison between participants, the game and experiment underwent extensive calibration prior to the main study to better replicate a game of this style. The following adjustments were made:

- **Rounds:** The experiment consisted of 5 rounds. Rounds 1–2 served as baselines for player metrics, round 3 was the intervention point (countermeasures applied), and rounds 4–5 assessed post-intervention adaptation.
- **Health Removal:** Health was removed to prevent early elimination. Instead, each hit caused a monetary penalty, maintaining in-game stakes.
- **Monetary System:** Per-round coin earnings were adjusted so that all players could afford purchases in the shop to keep up with the increasing difficulty. Taking damage resulted in lost coins which was also accounted for in these adjustments.
- **Shop Simplification:** Only weapon unlocks and ammo purchases were available. Stat-buff upgrades (move speed, reload, DPS multipliers) were disabled to avoid confounding playstyle metrics and decreasing effectiveness of the countermeasures.
- **Consistent Wave Difficulty:** Each wave was designed to last 45–60 seconds for the average player, with careful tuning of enemy numbers and spawn intervals.

A.2 Wave-by-Wave Configuration

Wave 1 (Baseline):

- **Enemies:** 6 Purple (fast, 50 coins), 6 Red (standard, 25 coins), 4 Blue (tank, 50 coins); total = 16 enemies.
- **Total Coin Reward:** $6 \times 50 + 6 \times 25 + 4 \times 50 = 650$ coins.
- **Hit Penalty:** 10 coins lost per hit.
- **Completion Bonus:** 50 coins for finishing the wave, offsetting up to five hits.
- **Net Coins Formula:**

$$\text{Net}_1 = 650 + 50 - 10H = 700 - 10H,$$

  where H is the number of hits taken.
- **Target Budget:** At least 650 coins for all players.
- **Wave Duration:** At 0.5 enemies/s, 16 enemies $\approx$ 32s; observed times: 45–60s.

Shop Structure:

- **Available Purchases:**
  - Shotgun (150 coins)
  - SMG (500 coins)
  - Rocket Launcher/Flamethrower (800 coins)
  - Ammo bundles (100 coins to refill all guns)
- **Trade-Offs:** Under the 650-coin budget, four main strategies were observed:
  1. Dual unlock: Shotgun + SMG (spend all 650)
  2. SMG + ammo
  3. Shotgun + more ammo
  4. Save 650 coins for Rocket/Flamethrower in Wave 2

Weapon DPS Calibration:

From solo playthroughs, the average times to clear Wave 1 with each weapon:

$$\begin{aligned}
\bar{t}_{\text{pistol}} &= 81.2\,\text{s},\\
\bar{t}_{\text{shotgun}} &= 65.8\,\text{s},\\
\bar{t}_{\text{SMG}} &= 69.1\,\text{s},\\
\bar{t}_{\text{rocket}} &= 63.8\,\text{s},\\
\bar{t}_{\text{flame}} &= 72.6\,\text{s}.
\end{aligned}$$

Corresponding kill rates and relative DPS boosts:

$$\begin{aligned} r_{\text{pistol}} &= 0.1971\,\text{enemies/s}, \\ r_{\text{shotgun}} &= 0.2432, \\ r_{\text{SMG}} &= 0.2317, \\ r_{\text{rocket}} &= 0.2510, \\ r_{\text{flame}} &= 0.2204. \end{aligned}$$

Relative boost over pistol:

$$\begin{aligned} p_{\text{shotgun}} &\approx 0.2339, \\ p_{\text{SMG}} &\approx 0.1754, \\ p_{\text{rocket}} &\approx 0.2731, \\ p_{\text{flame}} &\approx 0.1180. \end{aligned}$$

ROI for each weapon:

$$\text{ROI}_j = \frac{p_j}{C_j},$$

$$\text{ROI}_{\text{shotgun}} \approx 0.00156, \quad \text{ROI}_{\text{SMG}} \approx 0.00035, \quad \text{ROI}_{\text{rocket}} \approx 0.00034, \quad \text{ROI}_{\text{flame}} \approx 0.00015.$$

Thus, Shotgun and SMG are best initial investments.

Budget Simulation: With 650 coins, the "greedy" purchase is Shotgun + SMG, for a cumulative DPS boost:

$$P_{\text{total}} = (1 + p_{\text{shotgun}})(1 + p_{\text{SMG}}) - 1 \approx 0.4503.$$

A.3 Difficulty Scaling for Waves 2–5

Wave 2:

- **Target end-of-wave budget:** 750 coins.
- **Enemy composition:** Add 2 Red (+50), 1 Blue (+50), 1 Purple (+50) = 150 coins added to Wave 1. New total: 7 Purple, 8 Red, 5 Blue (20 enemies, 800 coins gross).
- **Spawn interval:**

$$\Delta_2 = \frac{3\,\text{s}}{1 + P_{\text{total}}} \approx 2.07\,\text{s}$$

- **Wave duration:** Validated via five playthroughs; 60% of runs within ±5% of Wave 1 times.
- **Hit penalty:** 10 coins/hit; no completion bonus.

Waves 3–5:

- **For each wave:** Cumulative DPS boost is recalculated as players acquire new weapons. Spawn intervals decrease as player power increases.
- **Enemy counts:** Increased by +1 Purple, +2 Red, +1 Blue per wave.
- **Budget targets:** Ensured players could afford next-tier weapon or meaningful ammo after each wave.
- **Wave 3:** 7P, 8R, 5B (20 enemies), $\Delta_3 \approx 2.07\,\mathrm{s}$, target 800 coins.
- **Wave 4:** 8P, 10R, 6B (24 enemies), $\Delta_4 \approx 1.625\,\mathrm{s}$, target 950 coins.
- **Wave 5:** 9P, 12R, 7B (28 enemies), $\Delta_5 \approx 1.453\,\mathrm{s}$. No budget target since the game ends after Wave 5.

B Survey Instruments

Evan Meltz Thesis - Pre Game Survey

Start of Block: Default Question Block

Consent You are invited to participate in my research study examining player behavior in a custom-built 3D-Wave shooter game. Participation is voluntary, and you may withdraw at any time by closing the browser. Anonymity & Data Use: Your identity will remain completely anonymous. We will collect in-game metrics and link them to your responses only via a random session ID—no personally identifying information will be recorded except for age and gender. The aggregated, anonymized data will be used solely for academic research in my master's thesis. By selecting "I Agree", you confirm that you have read this information, and that you consent to participate under the terms above.

- ○ I Agree (I consent to participate under the conditions described.) (1)
- ○ I Do Not Agree (I do not consent) (2)

Q1 What is your age?

____

Q2 What is your gender?

- ○ Male (1)
- ○ Female (2)
- ○ Other (3)
- ○ Prefer not to say (4)

Q3 Approximately how many hours per week do you play video games?

- ◯ I don't play video games (1)
- ◯ Less than 5 hours (2)
- ◯ Between 5 - 10 hours (3)
- ◯ More than 10 hours (4)

Q4 Have you ever played a wave-based shooter before today?

- ◯ Yes (1)
- ◯ No (2)
- ◯ Im not sure (3)

Q5 How many years have you been playing action shooters (top-down shooter / third-person shooter / first-person shooter)?

- ◯ I have never played these games (1)
- ◯ Less than 2 years (2)
- ◯ Between 2 - 4 years (3)
- ◯ Over 4 years (4)

Q6 Which is your dominant hand?

◯ Right (1)

◯ Left (2)

◯ Both (3)

Q7 Please rate the accuracy of this statement: "I consider myself a highly skilled shooter-game player who reacts quickly to on-screen threats and can dodge or avoid enemies under pressure." (1 = Not accurate; 7 = Extremely accurate)

◯ 1 (1)

◯ 2 (2)

◯ 3 (3)

◯ 4 (4)

◯ 5 (5)

◯ 6 (6)

◯ 7 (7)

Q8 Please choose one of the following:

◯ I prefer a fast-paced, aggressive approach (playing on the offensive). (1)

◯ I prefer a cautious, defensive approach (playing from cover). (2)

◯ Im not sure. (3)

Q9 Please choose one of the following:

- ○ I like to experiment with multiple weapons. (1)
- ○ I stick to one weapon until mastery. (2)
- ○ Im not sure. (3)

Q10 Please rate the accuracy of this statement: "When I play games I often lose track of time, feel fully absorbed when I’m playing at my best, and generally feel in control of what happens." (1 = Not accurate; 7 = Extremely accurate)

- ○ 1 (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 (7)

End of Block: Default Question Block

Evan Meltz Thesis - Mid Game Survey

Start of Block: Default Question Block

Q1 How challenging did you find this Wave? (With 1 being the least challenging and 7 the most challenging)

- ◯ 1 (1)
- ◯ 2 (2)
- ◯ 3 (3)
- ◯ 4 (4)
- ◯ 5 (5)
- ◯ 6 (6)
- ◯ 7 (7)

Q2 Please rate your excitement / thrill level for this wave (with 1 being the lowest and 7 the highest)

- ◯ 1 (1)
- ◯ 2 (2)
- ◯ 3 (3)
- ◯ 4 (4)
- ◯ 5 (5)
- ◯ 6 (6)
- ◯ 7 (7)

Q3 Please rate your frustration level for this wave (with 1 being the lowest and 7 the highest)

- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)

Q4 Please rate your level of boredom for this wave (with 1 being the lowest and 7 the highest)

- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)

Q5 Please rate your level of satisfaction for this wave (with 1 being the lowest and 7 the highest)

○ 1 (1)

○ 2 (2)

○ 3 (3)

○ 4 (4)

○ 5 (5)

○ 6 (6)

○ 7 (7)

Q6 Please rate the accuracy of this statement for this wave: "'I felt I had good control over my character’s movement and aiming and the game responded fairly to my actions." (with 1 being not at all accurate and 7 being very accurate)

○ 1 (1)

○ 2 (2)

○ 3 (3)

○ 4 (4)

○ 5 (5)

○ 6 (6)

○ 7 (7)

End of Block: Default Question Block

Evan Meltz Thesis - Post Game Survey

Start of Block: Default Question Block

Q1 Please rate the accuracy of this statement: "During the entire session, I felt fully absorbed in the gameplay, I felt in control of what was happening during the game and I lost track of time while playing." (1 = Not Accurate; 7 = Very Accurate).

- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)

Q2 I noticed the game’s difficulty changed at some point without me prompting it.

- Yes (1)
- No (2)
- Im not sure (3)

Q3 If yes, did you feel the changes were: [1 = Too Easy → 4 = About Right → 7 = Too Hard].

- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)
- I didnt notice (8)

Q4 Please rate the accuracy of this statement: "Compared to the beginning, I felt my skills improved by the end of wave 5 and I learned new tactics I hadn’t used before." (1 = Not Accurate; 7 = Very Accurate).

- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)

Q5 Please rate the accuracy of this statement: " I felt motivated to try new strategies I would normally avoid in my usual playstyle." (1 = Not Accurate; 7 = Very Accurate).

- ○ 1 (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 (7)

Q6 Please rate the accuracy of this statement: "Overall, I felt the game’s difficulty adjustments as the waves progressed were fair." (1 = Not Accurate; 7 = Very Accurate).

- ○ 1 (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 (7)

Q7 Please rate the accuracy of this statement: "At times, I felt frustrated by how the difficulty changed." (1 = Not Accurate; 7 = Very Accurate).

- ○ 1 (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 (7)

Q8 Please rate the accuracy of this statement: "Overall, I enjoyed playing this game session and would like to play this game again under similar conditions." (1 = Not Accurate; 7 = Very Accurate).

- ○ 1 (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 (7)

Q9 Please share any additional comments about the game: • What did you like most? • What did you like least? • Do you have any suggestions for how the game adjusts its difficulty? • Did you encounter any bugs or technical issues?

____

____

____

____

____

End of Block: Default Question Block