Journal on Interactive Systems, 2025, 16:1, doi: 10.5753/jis.2025.5630

Evaluating a Procedural Content Orchestrator Gameplay Data and Classifying User Profiles

Leonardo Tórtoro Pereira [**São Paulo State University (UNESP)** | ***leonardo.t.pereira@unesp.br***]
T. Yuji Teoi [**University of São Paulo** | ***yujiteoi@gmail.com***]
Claudio Fabiano Motta Toledo [**University of São Paulo** | ***claudio@icmc.usp.br***]

Institute of Geosciences and Exact Sciences, São Paulo State University, Av. 24 A, 1515, Bela Vista, Rio Claro, SP, 13506-700, Brazil.

Received: 03 March 2025 • **Accepted:** 01 July 2025 • **Published:** 04 July 2025

Abstract: *Background:* Content Orchestration is a novel research field focused on coordinating distinct types of algorithmically generated game content. *Purpose:* Thus, the lack of research in this area hinders the analysis of gameplay data and player profiling in games with orchestrated content. *Methods:* This paper is an extension of a work that collected and analyzed gameplay logs of 15 players who played 119 game sections of 12 different dungeons of a top-down action game. The game's Levels, Rules, and Narrative content were orchestrated and adapted to player profiles defined from a pre-test questionnaire. PCA and clustering techniques were used to highlight relevant gameplay metrics for distinguishing play styles. In this extension, we used the gameplay data alone to train classifiers with and without data augmentation to predict a user's profile, measuring the accuracy, precision, recall and f1-score with a train-test split and a 5-fold cross-validation for a more robust accuracy. We also implemented data augmentation on our gameplay metrics sample. *Results:* We identified, through the previous work, two components of PCA explaining a total of 65% of data variability, containing data such as *Lock Usage Rate*, *Enemy Kill Rate*, *Map Completion*, and *Completed Immersion Quests*. We also found *game difficulty* as an important level component for impact clustering. Through data augmentation, we achieved novel results, such as a mean accuracy of almost 95%, measured with a 5-fold cross-validation, for the Histogram-based Gradient Boosting classifier when predicting a player's profile based on their gameplay data, even with our small sample size. *Conclusion:* Our work guides developers and researchers to choose relevant gameplay metrics to determine players' play styles. Our extended results suggest that we can predict player's profiles through gameplay metrics and data augmentation, even for small samples. More studies are needed to validate our findings, with a larger and more diverse player-base.

Keywords: Procedural Content Generation, Player Profiling, Content Orchestration, Gameplay Metrics Evaluation, Machine Learning

1 Introduction

The field of procedural content generation (PCG), grounded in game content creation by algorithmic means [Togelius *et al*., 2016], relates to innovative games that can virtually create infinite diversified content, reducing production costs and increasing development efficiency. Still, most of the procedurally generated content in industry and academia focuses on a single content type (e.g., only game levels or procedural biome background). In this context, the concept of *content orchestration*, which is the coordination of two or more distinct types of procedural content, was proposed as a novel approach to multifaceted content. However, it still lacks research [Liapis *et al*., 2019].

Another field challenge is related to adaptive PCG, where the procedural content is dynamically adjusted, promoting a challenging, but not too punishing, player experience [Cowley *et al*., 2008]. It is relevant to notice that players' engagement reasons are diverse and have different meanings [Yee, 2006], so it is possible to classify types of players into profiles and address each profile with adapted content. Recent studies help us investigate different taxonomies and typologies regarding motivations, behavior, and personality characteristics, and also the most commonly used ones, their contradictions, and convergences Carneiro *et al*. [2022]; Kirchner-Krath *et al*. [2024].

Therefore, player modeling (related to dynamic profiling) and player profiling methods (related to static and once-defined profiling) compromise the definition of player profiles through implicit and explicit data. Still, as supported by Yannakakis and Togelius [2018], a player modeling method can improve its performance using an accurate player profiling method. However, neither player modeling nor profiling has enough research in the context of content orchestration.

This work extends the study presented by Pereira *et al*. [2024b]. In the mentioned study, we collected implicit (collected through questionnaires) and explicit data (gameplay logs) from 15 players and 119 play sections from a procedurally orchestrated and player profile-adapted top-down action game. Furthermore, we performed PCA and clustering with the explicit data to compare with the player profiles defined from the implicit data. In the study, we commented on the limited sample size. Here, we mitigate the drawbacks of the small sample size using data augmentation, in Section 4.3. Furthermore, we provide an enriched introduction about player profiling, add a background section to clarify major concepts (Section 2), add more details about the tested game and the study's setup in Section 3, and discuss in more depth our findings and limitations (sections 5 and 6, respectively).

Our work outlines the relevant gameplay metrics that need to be collected for content adaptation and indicates those less relevant metrics, using a Principal Component Analysis. Moreover, we identify that the characteristics of the played content are as relevant as the player's profile to evaluate their metrics, according to clustering algorithms. And, finally, we discovered that statistical data augmentation techniques can be used in small datasets to enhance the quality of a classifier algorithm that can predict a player's profile using only their gameplay metrics with a mean accuracy of almost 95%, measured with a 5-fold cross validation.

Therefore, this work may help developers and researchers as a guide on what gameplay characteristics are more relevant to be collected to differentiate between play-styles. It is also an initial work towards classifying play-styles using only gameplay metrics, enabling developers and researchers to reduce the reliance on explicit data collection, minimizing the use of intrusive self-reporting, creating a more seamless flow to collect user data and avoid self-reporting bias. Finally, possibly our major contribution, we validate with supervised learning metrics that, with data augmentation, we can train a model that can correctly predict a player's current indicated profile with their gameplay metrics from the last dungeon, allowing applications that can adapt content in real-time, such as for real-time content generators and orchestrators.

1.1 Ethical issues

This work is related to the research project approved by the Brazilian Research Ethics Committee under CAAE number 83177524.7.1001.5411. We do not collect personal data, to ensure accordance with Brazil's General Law on the Protection of Personal Data.

As the research involves the collection of gameplay data and questionnaire answers from players, they firstly were informed about the research goals and data collection through a game screen, as shown in Figure 2, following resolutions 466/2012 and 510/2016 from Brazil's *Conselho Nacional de Saúde*). The players were invited to continue the game only if they consented to the aforementioned issues and could stop playing at any time they wanted.

Nonetheless, we sought players for the study through social network posts (with the link to the game's download, and a brief explanation of the experiment and how their data would be collected and used) and mailing lists. As no personal data was collected, we did not have any information about who accessed the link and downloaded the game.

2 Background

This section presents the essential concepts for understanding our work. We also present studies using similar methods.

2.1 Procedural Content Generation

In 2014, *Minecraft*[1], one of the most popular games of all time, was acquired by Microsoft for 2.5 billion dollars [Microsoft, 2020]. Its simplicity, balance and unique aesthetic explain the game's success. However, it is worth it to account for the relevance of the procedural generation.

Procedural content generation is the creation of game content via computer algorithms [Togelius *et al.*, 2011]. As it seems, this definition encompasses various content, including textures, levels, item parameters, and dynamic enemy spawn.

PCG offers advantages such as automated content creation, reduced production costs, and improved efficiency [Togelius *et al.*, 2016]. Another use of PCG is to address the element of uncertainty in gameplay, which is essential for player's engagement and fun [Salen and Zimmerman, 2003; Koster, 2005]. Unlike simple randomization, PCG systems generate uncertainty through AI, advanced algorithms, or human-PCG collaboration.

Based on the procedural generation of levels and constant replayability, some game genres are only feasible using PCG, such as in *Roguelikes*. In the Minecraft case, virtually infinite worlds are created procedurally through a random seed and gradient noise. Each new world can provide a new game experience.

2.2 Creative Facets of Games

The modern game's complexity encompasses a vast number of development fronts, each producing a specific game piece. In this context, some authors find it useful to classify game content into six *creative facets* [Liapis *et al.*, 2014, 2019]:

- **Audio**: any sound asset for games, such as music, sound effects, and voice;
- **Levels**: any asset the player "passes through", such as dungeons, rooms, objects, or character placement;
- **Narrative**: any type of game narrative, such as character's story, world lore, and quests;
- **Rules**: any type of game rule, such as character status, weapon status, and parameters for win or loss condition;
- **Visuals**: any visual asset for games, such as textures, sprites, 3D models, skins, and visual effects;
- **Gameplay**: any test or simulation of gameplay, such as bots, AI testing, and user testing.

Despite some game content could stay ambiguous, such as in the game *Baba is You*[2] where the rules and levels are intrinsically mixed, this classification into creative facets provides a useful nomenclature for developing focused procedural generators.

2.3 Content Orchestration

In the context of PCG, content orchestration is the coordination of the procedural generation of two or more distinct creative facets [Liapis *et al.*, 2019].

Liapis *et al.* [2019] proposed this concept to address the fact that, in industry and academia, the PCG is mostly performed in only one creative facet. They defined a PCG system called *content orchestrator*, which is classified as:

- **Top-down**: The orchestrator acts as a maestro in a symphony. The musicians (i.e., the procedural content generators) follow the exact instructions of the maestro;

[1]`https://www.minecraft.net/` (Accessed July 3, 2025)

[2]`https://hempuli.com/baba/` (Accessed July 3, 2025)

- **Bottom-up**: The musicians "improvise" based on what the other musicians are playing, like in jazz music. Each musician has a "sheet music" with simple instructions;
- **In-between**: A mix of the two approaches.

Notice that this sub-field of PCG is novel and lacks research.

2.4 The Overlord Content Orchestrator

Overlord is an *In-between* content orchestrator responsible for starting and coordinating the generation of three PCG systems, each responsible for a creative facet: levels, rules, and narrative [Pereira *et al.*, 2022].

The orchestrated content is summarized below:

- **Levels Generation**: the generator produces dungeons containing rooms with enemies, NPCs, and items (e.g., collectibles, lore items, equipment, and the winning item);
- **Rules Generation**: the generator selects enemy parameters, such as health, damage, projectile speed, attack speed, move speed, active time, rest time, and type of move;
- **Narrative Generation**: the generator selects a pre-made quest template according to the player profile.

The orchestrator considers the player profile to post-process the content generated by the distinct PCG systems, maintaining an adaptive PCG. The orchestrator's capacity to adapt the content was tested using prototypes of top-down and platformer genre games and obtained positive results for subjective fun and profile adaptation [Pereira *et al.*, 2022; Teoi *et al.*, 2024].

For this work, we use implicit data collected from a top-down prototype game, orchestrated by Overlord. We note that the game's user experience was tested and described in Pereira *et al.* [2022]; Teoi *et al.* [2024]. The former showed that users who played content suggested for their profiles had, through a visual analysis in Figure 14 from Pereira *et al.* [2022], perceived the game as more fun, more challenging, more difficult to face enemies and find the exit, with users liking more the exploration, more balanced and with users liking more the key-lock puzzles when compared to people who played content not recommended for their profile.

Nonetheless, a statistical analysis to compare both groups was not present, but was shown in Pereira [2022], analyzing the same dataset. There, they provided evidence that people who played content generated towards their profile had more fun, liked more the exploration, had more fun completing the quests, liked better the rewards, and also though the quests were more human-made than those that played content generated towards a non-matching profile. They presented a Mann-Whitney U test to compare the mean from both population, with an alpha of 0.05. The stated statistical power was of 0.947 for a Cohen's D of 0.8. Therefore, we have evidence that the orchestrator can make the content be perceived as overall better for users that it creates content towards their profile.

For the latter, Teoi *et al.* [2024] shows a visual and statistical analysis from a smaller sample size in a different game (a platformer), using the same orchestrator. The players from their study liked more the challenge and exploration, the challenges from the quests, and also had more fun with the quests when playing content matching their profile.

2.5 Player modeling and Player Profiling

Adaptive PCG is a dimension of PCG in which a player's in-game interactions influence the procedurally generated content. A notable example is the game *Left 4 Dead*[3], where a PCG system named *AI Director* adjusts game difficulty, controls enemy spawn position, and places items and weapons according to the player's performance in the game. This prospect of content adaptation creates anticipation on players, that have been reported to be more immersed in a difficulty-adapting game over one that does not adapt. However, just by stating that the game adapts its difficulty was enough to make them more immersed, even if no adaptation was done Denisova and Cairns [2019].

Isbister [2016a] highlights the emotional design importance for player experience, highlighting player's meaningful choices and *flow*. Nunes and Darin [2024a] shows that audio impacts player experience through "immersion, arousal, affect (positive and negative), and challenge" and requires consistency across distinct player preferences. Thus, a successful content adaptation should enable players to perceive their influence on the game through their actions, while providing not-too-easy and not-too-difficult gameplay.

Still, for an accurate content adaptation, correctly identifying the player's skills and motivations for play is essential. Two approaches could address this challenge in this context: *player modeling* and *player profiling*.

Player modeling is a dynamic way of identifying player's characteristics through algorithms and gameplay metrics. An accurate model updates according to the player progression and skill improvement [Machado *et al.*, 2011].

The static method for acquiring information from players is called player profiling. The profiling occurs only once, usually by a questionnaire presented before or after a play section [Yannakakis *et al.*, 2013].

Despite the difference between these two terms, using a player profiling method can improve the accuracy of a player modeling method [Yannakakis and Togelius, 2018]. Although both methods are heavily used in industry and well-researched in academia, little research exists on content orchestration.

2.6 Implicit and Explicit Data Collection

In the context of *Adaptive Interfaces and Agents*, *implicit input* includes all actions a user performs with the system with no explicit purpose of the user's information reviling [Jacko, 2012, p. 318]. Thus, *explicit input* relates to purposely presenting a questionnaire or form for the user to fill.

From the perspective of digital games, implicit data collection relates to gathering player information through gameplay metrics without directly asking players for something.

[3] `https://pt.wikipedia.org/wiki/Left_4_Dead` (Accessed July 3, 2025)

In contrast, explicit data collection directly requires players to answer questions, usually using a questionnaire or a profile selection panel [Pereira *et al.*, 2024a].

Player modeling and profiling methods can use implicit and/or explicit data collection approaches [Yannakakis and Togelius, 2018].

2.7 Related Work

This work relates to analyzing and comparing data collected implicitly and explicitly through the lens of player profiling. As our explicit player profiling method was already tested and presented positive results regarding content adaptation [Pereira *et al.*, 2022; Teoi *et al.*, 2024], our PCG system still lacks of an accurate method for dynamic adaptation through implicit data collection.

Thus, our focus is measuring the effectiveness of our collected gameplay metrics. In this context, we summarize related works encompassing the collection and use of explicit data thought gameplay collection.

Bernhaupt [2015] described a set of generic and gender-focused metrics used in previous studies (not in the context of PCG or player profiling). Some metrics included game duration, determination of important events, and player's interactions with game objects.

Heijne and Bakkes [2017] collected multiple game metrics across gameplay blocks of the village tutorial, enemy combat, puzzle solving, and exploration. They used these metrics to correlate players' profiles and questionnaire responses of faced difficulties and activity preferences. Although they obtained no conclusive result, they pointed out a set of gameplay metrics to accomplish desired goals.

Melhart *et al.* [2019] collect explicit data using the *Ubisoft Perceived Experience Questionnaire* (UPEQ) and implicit data of gameplay metrics to feed machine-learning models. They collected 26 metrics, including player level, days played, duration of gaming sessions, among others. From players' activities within the game, they identified several play styles after using player profiling techniques. Their classification method and collected data differ from our work, given that they employed *support vector machine* models, with data from a specific third-person shooter game named *Tom Clancy's The Division*.

Loria and Marconi [2018] compared profiles obtained from a *Hexad* questionnaire and abstract player behaviors from metrics of a gamified application. They found a low relation between the metrics and Hexad types, but positive results regarding mapping abstract actions and crafting believable characters.

de Lima *et al.* [2021] described a game narrative adaptation from the player's in-game actions, including enemy combat, item use, and interaction with NPCs. The story arc added tension to the narrative, changing narrative dynamics and initiating game events. They highlighted the adaptive capacity for creating a more compelling narrative trajectory by comparing their narrative-adaptive system with a default narrative system.

3 Method

3.1 Technologies

We developed the game used for this study using the game engine *Unity 3D*[4], and *C#* as a programming language.

The project is free and open-source on *GitHub*[5].

3.2 Top-down action game

The game developed for our research is a top-down genre game, with the player and enemies having mechanics of moving (up, down, right, left) and shooting (with distinct weapons and projectile speed). The rooms are filled with procedurally generated blocks and can contain enemies, NPCs, lore items, collectibles, and equipment. Figure 1 illustrates a game level. More gameplay screenshots and *GIFs* can be seen at this project's repository at Open Science Framework (OSF), together with the evaluated data and the script used to generate the results that will be discussed throughout the paper [6].

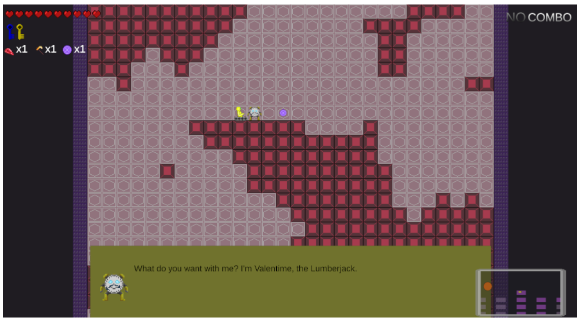

Figure 1. A screenshot of the game's dungeon. The player is the yellow character, the red mage is the enemy, the gray actor is a Non-player character, and the purple gem is a collectible item.

To win, the player must find a *Triforce*, an item instantiated far from the player's initial room. Still, the player can explore additional rooms, complete quests, defeat enemies, and collect items. The loss condition is achieved when the player receives enough damage to lose all its life points.

As mentioned in Subsection 2.4, the procedural content of levels, rules, and narrative was orchestrated by *Overlord*, using player profiling and content adaptation.

3.3 Gameplay metrics collection

Gameplay metrics are measures of player behavior, which can be collected during any game development stage Magy Seif El-Nasr [2013]. In this study, we collected in-game metrics implicitly through gameplay logs. The log was stored once the player finished a play section by winning or exiting the game.

All collected metrics are presented in Table 1. We measure the number of enemies defeated, how much health the player lost, how many rooms they visited (with and without repetition), how many keys were found and locks were opened,

[4] `https://unity.com/` (Accessed July 3, 2025)

[5] `https://github.com/LeonardoTPereira/Overlord-Project` (Accessed July 3, 2025)

[6] `https://osf.io/jqf39/?view_only=33bc172e52e145eaa85807edd451d497` (Accessed July 3, 2025)

how many quests were completed (from each type of available quest), how many hits they applied to enemies without taking damage (combo), how much time they needed to finish the dungeon, and if they had finished successful or "died" in the dungeon. These data were collected through integer values and boolean ones.

3.4 Yee's player profiling

For the player profiling, we collected data explicitly by sending a 12-question questionnaire to the player, sent the first time the player opened the game. The pre-test questionnaire questions and their description are presented in Table 2.

We based the pre-test questionnaire on the work of Yee and Ducheneaut [2018], for its large acceptance in academia and industry. Nick Yee's work also influenced our player profiling method, in which he identified that players have distinct motivations for playing [Yee, 2006]. We considered the following motivation factors:

- **Achievement**: interest in maximizing game goals and building up the strongest character;
- **Creativity**: interest in character customization and in-game self-expression;
- **Immersion**: interest in meaningful narratives and character immersion;
- **Mastery**: interest in seeking challenges, complex problems, and strategy.

Each profile has a numeric weight for each motivation factor. After the player answers the pre-test once, the profile is set by calculating Equation 1 and setting the weight of 4 for the highest, 3 for the second, 2 for the third, and 1 for the lowest-rated motivator.

$$\mathbf{A} = Q8 + Q9 + 5 - Q12 \tag{1}$$
$$\mathbf{C} = Q6 + Q7 + 5 - Q12 \tag{2}$$
$$\mathbf{I} = Q10 + Q11 + 5 - Q12 \tag{3}$$
$$\mathbf{M} = Q3 - 3 + Q4 + Q5 \tag{4}$$

The questionnaire comprised 12 items: two about general gaming experience (Q1–Q2, excluded from profiling), three for Mastery (Q3–Q5), two for each remaining profile (Q6–Q11), and one (Q12) penalizing non-Mastery profiles.

4 Results

We shared the game on social media and among the researchers' networks. We did not collect any personal information. Each participant was presented with the consent term in Figure 2 and proceeded only if agreed. Figure 3 illustrates how the questionnaires are presented on the game screen. Participants were requested to answer the pre-test questionnaire the first time they played the game. Furthermore, they were asked to respond to the post-test questionnaire each time they ended a dungeon. None of the questions in the questionnaires were mandatory, and participants were free to leave the game without any harm.

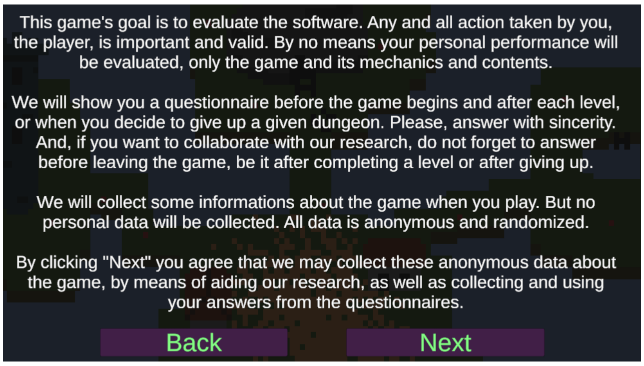

Figure 2. Game screen of research explanation and player consent. The screen explains the research goal (i.e., evaluate the system) and data collection method (i.e., questionnaires and gameplay metrics). It also asks if the player agrees to continue the process.

All data were scaled between the range of 0 to 1, considering their natural bounds as 0's or 1's, except for the variable *Rooms Visited*, in which its value was divided by the dungeon's total rooms (which is the backtracking ratio), and then scaled. This approach is helpful for comparing distinct variables with distinct bounds, especially for the variable *Max combo*, where there is no pre-defined upper bound. These data are shown in Table 3.

From the data, we identified that players' combo of the *best players* was, on average, higher than *regular players*, as evidenced by the considerable standard deviation (0.232) for the variable *Max Combo*. The same principle was found for *completionist players*, with a higher value of completed quests than regular players and a high standard deviation close to 0.2.

Data not related to quests presented the highest standard deviation. The data of *Lost health, Lock Usage Rate,* and *Enemy Kill Rate* had a standard deviation next to 0.4, and *Key Collected Rate* close to 0.3. These data indicate their importance in our used clustering algorithm, given their high *entropy*. The explored data as well as the script used to generate all the following results are present in our already mentioned OSF's repository[7]

4.1 Principal Component Analysis

To better understand the data's entropy, we reduce the dataset's dimensionality in two dimensions using Principal Component Analysis (PCA). As a result of PCA, the first component has 57% of the explained variability, and the second has 9%, with a total of 65% of the data variability explained.

In Table 4, we present the contribution (eigenvalues)[8] for each variable, given its component (row). The first component (which has the highest explained variability) is mainly formed by the information of the variables: *Lock Usage Rate*, *Enemy Kill Rate*, *Map Completion*, *Key Collected Rate*, and *Lost Health*. The second component is formed mainly by: *Completed Immersion Quests*, *Completed Report Quests*, *Key Collected Rate*, *Map Completion*, *Lock Usage Rate*,

[7] `https://osf.io/jqf39/?view_only=33bc172e52e145eaa85807edd451d497` (Accessed July 3, 2025)

[8] The highest eigenvalues indicate the most important variables for the corresponding PCA component.

Table 1. Collected gameplay variables. The table shows the data collected during gameplay sections. It also describes the data type and its relation to players' interactions in the game.

Variable	Description
Has Finished	A boolean indicating whether the player found the Triforce or not.
Has Died	A boolean indicating if the player died.
Lost Health	An integer indicating how much health the player lost during a play session.
Max Combo	An integer accounting for the maximum number of times the player has hit enemies without losing health.
Rooms Visited	An integer indicating how many rooms (with repetition) were visited.
Unique Rooms Visited	An integer indicating how many unique rooms were visited.
Keys Found	An integer indicating how many keys the player found.
Doors Unlocked	An integer indicating how many doors the player unlocked.
Enemies Defeated	An integer indicating how many enemies the player defeated.
Completed Quests	A list of integers containing the number of times a quest type was completed.
Time to Finish	An integer indicating the seconds the player spent to complete the level.

Table 2. Pre-test questionnaire. The table shows all questions used for the pre-test questionnaire, used for player profiling. All questions are in a 5-point Likert scale format, ranging from *Strongly Disagree* (1) to *Strongly Agree* (5), except for Q3.

ID	Question
Q1	I am an experienced player.
Q2	I am an experienced player in the action-adventure genre.
Q3	In which difficulty do you usually play? (Options: Very easy, Easy, Medium, Hard, Very Hard)
Q4	I like playing games where I can explode, crush, destroy, shoot, and kill.
Q5	I like playing games where I can fight using close combat skills and evade fast attacks.
Q6	I like playing games where I can explore the game world and uncover secrets and mysteries.
Q7	I explore all the places, elements, and characters of the virtual world.
Q8	I complete all quests, including those that aren't necessary to finish the game.
Q9	I like playing games where I can collect rare items and hidden treasures.
Q10	I like playing games where I can build friendships between game characters and work toward a common goal.
Q11	I like playing games where I can immerse myself in the role of the character and make meaningful decisions.
Q12	I usually only do what is necessary to pass a level or complete a quest.

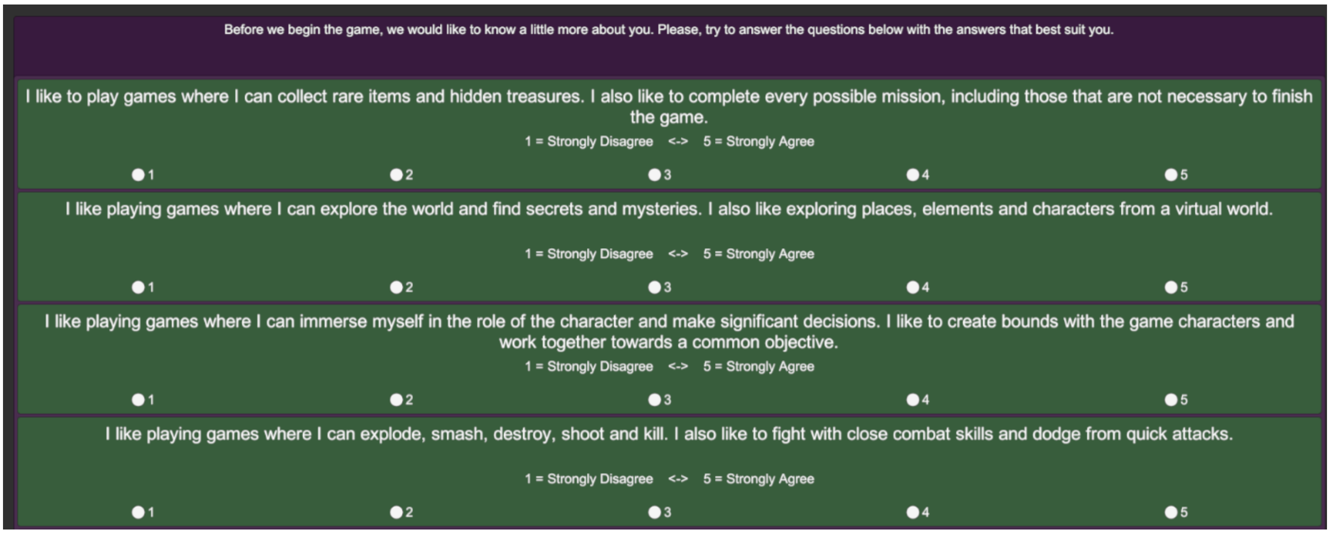

Figure 3. Game screen of a questionnaire. The figure shows the pre-test questionnaire presented to the player in the first game it plays.

Table 3. Descriptive statistics for each gameplay variable, after min-max scaling. Table adapted from Pereira *et al.* [2024b].

Data	Max Combo	Completed Report Quests	Completed Listen Quests	Completed Achievement Quests	Lost Health	Completed Explore Quests	Completed Exchange Quests
Mean	0.176	0.101	0.101	0.131	0.527	0.126	0.0857
Std. Dev.	0.232	0.249	0.188	0.222	0.375	0.2156	0.174
Min.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max.	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Data	**Completed Kill Quests**	**Completed Mastery Quests**	**Completed Give Quests**	**Time To Finish**	**Completed Creativity Quests**	**Completed Immersion Quests**	**Completed Go To Quests**
Mean	0.109	0.109	0.081	0.223	0.117	0.1397	0.078
Std. Dev.	0.199	0.199	0.194	0.190	0.196	0.234	0.170
Min.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max.	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Data	**Completed Gather Quests**	**Completed Read Quests**	**Lock Usage Rate**	**Key Collected Rate**	**Map Completion**	**Room Revisit Rate**	**Enemy Kill Rate**
Mean	0.125	0.067	0.628	0.774	0.694	0.235	0.580
Std. Dev.	0.228	0.183	0.414	0.323	0.358	0.187	0.390
Min.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max.	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Completed Achievement Quests, *Completed Gather Quests*, and *Completed Give Quests*.

These results indicate that information about *locks*, *keys*, *enemies*, *map completion*, and health are relevant for player profiling through gameplay collection. This data group relates to *creativity* (linked to exploration) and *mastery* motivation factors. It also relates to *achievement* motivation, for players' motivation to complete the entire dungeon.

Furthermore, with a minor but significant contribution, it is relevant to collect quest-related data, which in our work relates to the data of quest completion of *achievement*, *immersion*, and *creativity* quest types. This second group relates to *immersion* and *achievement*.

This shows that data related to the four motivation factors have significant relevance in explaining the variability aspect of the PCA, indicating its usefulness for player profiling through implicit collection of the aforementioned gameplay data.

4.2 Gameplay metrics clustering

We applied clustering algorithms in the gameplay data to compare its classifications with the pre-test questionnaire results.

We used the *K-Means* algorithm setting eight clusters, as we wanted to correlate with the eight player profiles identified from the pre-test questionnaire. The identified clusters are shown in Figure 4. The image at the bottom shows a scatter plot of each data, considering PCA, the player's profile (by color), and the played level's profile (by marker). The graph enables the visualization of the impact of distinct dungeon configurations over the gameplay data of each player profile. As shown, the clustering did not divide players by profile precisely. Still, some emerged groups are relevant to account.

The clusters C3 and C7 were mainly composed of the yellow profile (A=4, C=1, I=3, M=2), which contained mostly the level profile of (M=3, I=4, A=2, C=1). C3 contained mostly the level profile of (M=3, I=1, A=4, C=2), and C7 the levels of (M=4, I=1, A=3, E=2). Cluster C0 has the most data from the level profile of (M=4, I=1, A=3, E=2) as C7. However, it has the most players with high mastery motivation (mostly M=4).

The clusters C4 and C5 have a significant number of players with high mastery motivation (M=3 and M=4) but vary in level profiles. C6 has few data and consists of players with high mastery (M=4) who played low mastery levels (M=1). Furthermore, C5 has players with high mastery but playing mid-level mastery (M=2 and M=3). A high diversity of player and level profiles is found in clusters C1, C2, and C4.

We also used the *Spectral Clustering* algorithm, setting 8 clusters. The results are similar to the K-means, as shown in Figure 5. The Spectral's C0, C1, C4, and C5 are mainly equal to K-Means' clusters C1, C6, C3, and C2, respectively. Spectral's C3 includes more level profiles of (M=3, I=4, A=2, C=1) from the same player's profiles than K-Means' C0. Spectral's C2 has most of the levels profile of (M=4, I=1, A=3, C=2) and contains most combat difficulty and players with high *mastery* motivation, furthering the profile described by K-Mean's C0. Spectral's C6 and C7 are mainly a segregation of K-Means' C5, separating easier dungeons in Spectral's C6 and harder dungeons in C7.

Our clustering illustrates the impact of the level played on the player's gameplay, regardless of the player's profile. Furthermore, we notice that some levels focused on combat (high mastery) could make it difficult for players to clear the level. Further, given the importance of the variables of *Player Health* and *Enemy Kill Rate* in the variety in the PCA algorithm, it is noted that the level's difficulty impacted all gameplay data.

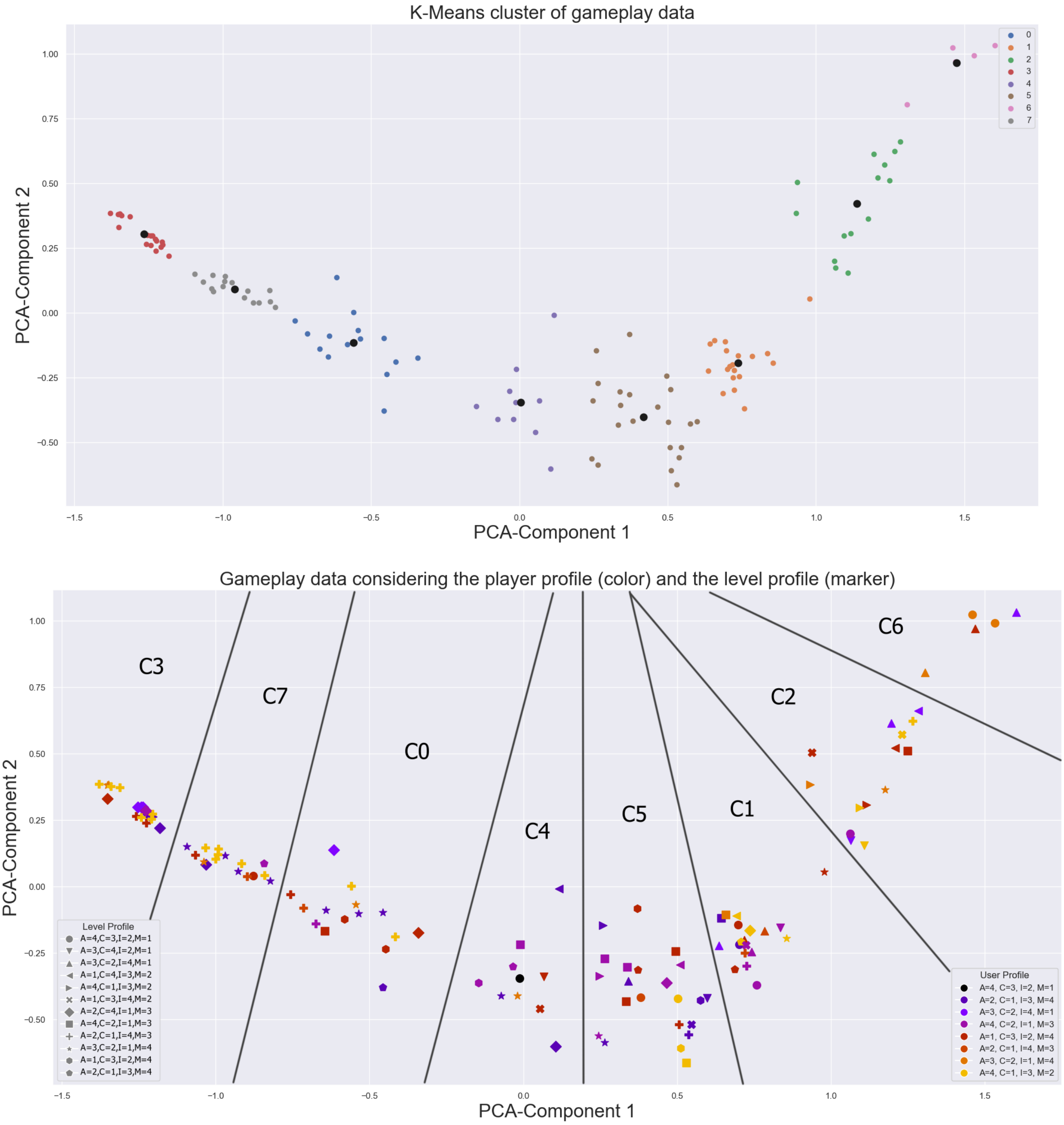

Figure 4. Scatter plot showing the generated clusters by the K-Means algorithm (top), and the gameplay data marked by player and level profile, with each cluster space marked by the lines. Figures obtained from Pereira *et al*. [2024b].

Table 4. Each variable's contribution (eigenvalues) to each principal component of the PCA. Table adapted from Pereira *et al*. [2024b].

Component	Max Combo	Completed Report Quests	Completed Listen Quests	Completed Achievement Quests	Lost Health	Completed Explore Quests	Completed Exchange Quests
0	0.02	0.14	0.12	0.16	0.34	0.15	0.10
1	0.21	0.33	0.19	0.27	0.09	0.20	0.18
Component	**Completed Kill Quests**	**Completed Mastery Quests**	**Completed Give Quests**	**Time To Finish**	**Completed Creativity Quests**	**Completed Immersion Quests**	**Completed Go To Quests**
0	0.12	0.12	0.11	0.18	0.14	0.18	0.10
1	0.08	0.08	0.24	0.10	0.21	0.36	0.16
Component	**Completed Gather Quests**	**Completed Read Quests**	**Lock Usage Rate**	**Key Collected Rate**	**Map Completion**	**Room Revisit Rate**	**Enemy Kill Rate**
0	0.15	0.09	0.44	0.31	0.38	0.18	0.40
1	0.25	0.19	0.29	0.31	0.29	0.09	0.10

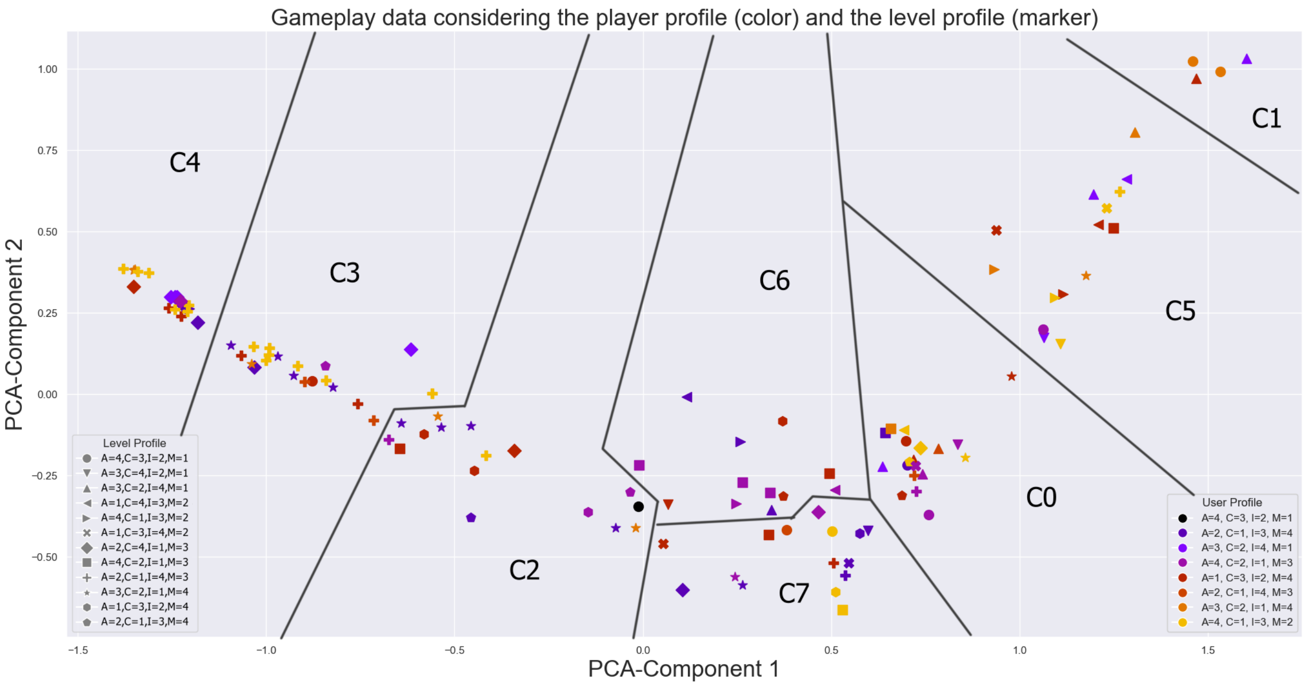

Figure 5. Scatter plot showing the gameplay data marked by player and level profile, with each cluster space marked by the lines, now generated by the Spectral Cluster algorithm (top). Figure obtained from Pereira *et al*. [2024b].

4.3 Player profile classifier

Although our clustering analysis did not provide a clear division between player profiles and gameplay metrics, we continued our data exploration through classification Zaki *et al.* [2020]. We tested different classifiers from Python's *scikit-learn*[9]. To help us check the quality of different families of supervised learning methods for classification, we selected methods from different classes of supervised learning algorithms to explore their performance with our data: the Gaussian Naive Bayes (GNB) Classifier[10]; the Decision Tree (DT) Classifier [11]; the Stochastic Gradient Descent (SGD) Classifier[12]; the Multi-layer Perceptron (MLP) Classifier[13]; the Histogram-based Gradient Boosting (HGB) Classification[14]; and, finally, the Support Vector Machine (SVM) Classifier[15].

Our predictor variable was the player's profile, which was the string representing the array of weights for each profile and their respective letter (e.g., A=4, C=1, I=3, M=2). Then, we discarded all variables related to the pre- and post-tests, using only the gameplay metrics as training variables, and using the Min-Max Scaler[16] to transform all of them to the 0-1 range. Therefore, the training data comprised the 22 columns: Max Combo; Completed Quests (for all types separately, summing 13 columns); Time to Finish; Total Lost Health; Key Collection; Lock Usage; Map Completion; Room Revisited Rate; Enemy Kill Rate; and Success Rate.

We conducted two tests with our classifier: a train-test split of 30%, and a 5-fold cross-validation score analysis. We then evaluated the accuracy, precision, recall and f1-score for the train-test split, and the mean and standard deviation of the cross-validation. Initial results were promising compared to chance, with some classifiers achieving an accuracy of 0.222. That is, considering that we had 8 different player profiles, a random guess would score 0,08 points in accuracy. This means that the classifiers can extract some knowledge from the gameplay metrics. But these results were still unacceptable for a real-life scenario, as 80% of gameplay data would be incorrectly classified and give players a wrong profile.

These issues are probably caused by our lack of data, since machine learning algorithms usually need thousands and even hundreds of thousands of data to have an increased accuracy. A recent research particularly shows how having a reduced dataset with close to one hundred instances (like ours) can significantly impact classification results Althnian *et al.* [2021]. Therefore, we used Data Augmentation strategies to enhance our dataset and test if this may lead to better results Ekwaro-Osire *et al.* [2025]. As far as we know, this is the first study using data augmentation for player-profile classification in the literature. So we explored different strategies and will further discuss the results, so others may benefit.

We select two statistical data augmentation techniques: noise and scaling, both maintaining the predictive columns. In the former, until the given "n" new instances are created, we iterate through the original instances and, for each instance, we give each of the 22 training columns' value a 50% chance to "mutate" it to a random value, by adding a value sampled from a normal distribution, centered in 0 and scaled using the column's standard deviation. Concerning the latter, for the same given "n" new instances, we iterate through the original instances and, for each, multiply the original value by a scaling factor.

Finally, we execute the same train-test split and cross-validation methods to evaluate the previous classifiers, with 6 different data augmentation setups, always applying the over an integer multiple of the original dataset, stated as n for simplification, and N will represent the Noise Data Augmentation technique and S the scaling technique, followed by the scaling factor: (1) $1xN$ (augments the original dataset with n new instances with the noise data augmentation); (2) $1xS - 1.05$ (augments the original dataset with n new instances by scaling the original data by a 1.05 factor); (3) $2xN$; (4) $2xS - 1.05 + 0.95$ (augments the original dataset with n instances by scaling with a 1.05 factor and another n new instances with a 0.95 factor); (5) $1xS - 1.05 + 1xN$; (6) $6xN + 4xS - 1.05 + 0.95 + 1.1 + 0.9$. The original dataset will be referenced as $baseline$ for clarity. Tables 5 through 10 present the results from the classifiers, highlighting the best classifiers for each data augmentation strategy in bold.

As shown in Table 5, all augmentation techniques improved accuracy over baseline, with HGB outperforming SGD and SVM (the baseline's top performers). As these two methods are sensitive to hyperparameters Zaki *et al.* [2020], tuning may help them achieve better results on the augmented dataset.

Table 5. Classifiers' accuracy using the 30% train-test split approach, with seed 42.

Setup	GNB	DT	SGD	MLP	HGB	SVM
Baseline	0.083	0.056	**0.222**	0.083	0.167	**0.222**
1xN	0.056	0.222	**0.333**	0.167	0.292	0.208
1xS-1.05	0.069	0.097	0.264	0.125	**0.333**	0.222
2xN	0.056	0.093	0.111	0.148	**0.259**	0.204
2xS-1.05+0.95	0.111	0.097	0.194	0.222	**0.333**	0.306
1xS-1.05+1xN	0.098	**0.308**	0.175	0.147	0.273	0.154
6xN+4xS-1.05+ 0.95+1.1+0.9	0.088	0.192	0.086	0.112	**0.332**	0.216

Tables 6, 7, and 8 presents, respectively, the classifiers' precision, recall, and f1-score, corroborating and better explaining the results found while analyzing the accuracy. HGB is the best technique for most cases, providing an out-

[9] `https://scikit-learn.org/stable/` (Accessed July 3, 2025)

[10] `https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB` (Accessed July 3, 2025)

[11] `https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier` (Accessed July 3, 2025)

[12] `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html#sklearn.linear_model.SGDClassifier` (Accessed July 3, 2025)

[13] `https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html#sklearn.neural_network.MLPClassifier` (Accessed July 3, 2025)

[14] `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingClassifier.html#sklearn.ensemble.HistGradientBoostingClassifier` (Accessed July 3, 2025)

[15] `https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC` (Accessed July 3, 2025)

[16] `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html` (Accessed July 3, 2025)

standing precision for the 10x augmented dataset (bottom row in Table 6. This means that the model is, even at its best, misclassifying over half of the players. Nonetheless, the recall table mirrors the results from the accuracy, meaning that the model's specificity and sensitivity are the same, being able to have a similar amount of correct classifications spread through all profiles. Finally, Table 8 has the f1-score, and it mirrors the accuracy and recall results, with a somewhat poor score even for the best case (below 0.5). However, we do see an increase with data augmentation.

Table 6. Classifiers' precision using the 30% train-test split approach, with seed 42.

Setup	GNB	DT	SGD	MLP	HGB	SVM
Baseline	0.014	0.057	0.127	0.079	**0.128**	0.049
1xN	0.028	0.267	0.199	0.09	**0.271**	0.043
1xS-1.05	0.027	0.073	**0.201**	0.099	0.196	0.049
2xN	0.003	**0.307**	0.272	0.035	0.178	0.041
2xS-1.05+0.95	0.035	0.126	0.11	0.14	**0.306**	0.093
1xS-1.05+1xN	0.065	**0.256**	0.25	0.158	0.219	0.024
6xN+4xS-1.05+0.95+1.1+0.9	0.026	0.193	0.139	0.324	**0.421**	0.047

Table 7. Classifiers' recall using the 30% train-test split approach, with seed 42.

Setup	GNB	DT	SGD	MLP	HGB	SVM
Baseline	0.083	0.056	**0.222**	0.083	0.167	**0.222**
1xN	0.056	0.222	**0.333**	0.167	0.292	0.208
1xS-1.05	0.069	0.097	0.264	0.125	**0.333**	0.222
2xN	0.056	0.093	0.111	0.148	**0.259**	0.204
2xS-1.05+0.95	0.111	0.097	0.194	0.222	**0.333**	0.306
1xS-1.05+1xN	0.098	**0.308**	0.175	0.147	0.273	0.154
6xN+4xS-1.05+0.95+1.1+0.9	0.088	0.192	0.086	0.112	**0.332**	0.216

Table 8. Classifiers' f1-score using the 30% train-test split approach, with seed 42.

Setup	GNB	DT	SGD	MLP	HGB	SVM
Baseline	0.023	0.048	0.13	0.067	**0.117**	0.081
1xN	0.029	0.192	**0.245**	0.11	0.234	0.072
1xS-1.05	0.03	0.07	0.153	0.094	**0.236**	0.081
2xN	0.006	0.087	0.073	0.056	**0.197**	0.069
2xS-1.05+0.95	0.045	0.106	0.135	0.162	**0.312**	0.143
1xS-1.05+1xN	0.052	**0.274**	0.134	0.079	0.202	0.041
6xN+4xS-1.05+0.95+1.1+0.9	0.026	0.179	0.075	0.075	**0.303**	0.077

Overall, considering a train-test split analysis, we observe that data augmentation helped increase the quality of almost all models, specially for the Decision Tree and the Histogram-based Gradient Boosting techniques. Nonetheless, we then conducted a 5-fold cross-validation analysis, which is more robust and reduces variability. The results are presented in tables 9 and 10, and they show a high mean and low standard deviation for the best cases, leading to a more positive outcome for the models.

For the Cross Validation metrics, we observe that MLP was the best one for the baseline, while HGB was best for most cases, with Decision Trees being the best alternative for

Table 9. Classifiers' 5-fold cross validation score's average accuracy.

Setup	GNB	DT	SGD	MLP	HGB	SVM
Baseline	0.16	0.243	0.227	**0.361**	0.152	0.193
1xN	0.307	0.517	0.197	0.306	**0.551**	0.265
1xS-1.05	0.323	**0.779**	0.177	0.676	0.766	0.286
2xN	0.294	0.704	0.247	0.325	**0.776**	0.311
2xS-1.05+0.95	0.278	**0.774**	0.236	0.601	0.741	0.311
1xS-1.05+1xN	0.336	0.817	0.242	0.521	**0.909**	0.313
6xN+4xS-1.05+0.95+1.1+0.9	0.323	0.877	0.268	0.494	**0.947**	0.354

Table 10. Classifiers' 5-fold cross validation score's accuracy's standard deviation.

Setup	GNB	DT	SGD	MLP	HGB	SVM
Baseline	0.049	0.071	0.042	**0.067**	0.045	0.057
1xN	0.05	0.099	0.071	0.089	**0.09**	0.023
1xS-1.05	0.056	**0.15**	0.037	0.106	0.129	0.044
2xN	0.013	0.108	0.039	0.053	**0.06**	0.035
2xS-1.05+0.95	0.056	**0.109**	0.087	0.17	0.135	0.061
1xS-1.05+1xN	0.046	0.154	0.043	0.111	**0.073**	0.022
6xN+4xS-1.05+0.95+1.1+0.9	0.013	0.084	0.042	0.068	**0.031**	0.033

2 cases ($1xN$ and $2xS - 1.05 + 0.95$). The best results came from the data augmented 10 times over the baseline, that is, the 119 original gameplay metrics, augmented 6 times with noise (714 new instances), and with 119 new instances for four different scaling setups: multiplying the original training instances by 0.9, 0.95, 1.05, and 1.1 (476 total new instances). Therefore, the best training set contained 1309 instances of gameplay data, and had an average accuracy of almost 95%, with a standard deviation of 0.031 between folds.

5 Discussion

We analyzed gameplay data gathered from 15 different players, which played 119 gameplay sessions in total. They played levels generated using 12 different profile inputs from a content orchestrator that can procedurally generate dungeons, their rooms and locked-door puzzles, their quests, and the enemies placed inside them, both regarding placement, status parameters, and behaviors.

We first extracted information about our data's entropy and explainability, presented in Section 4.1, showing us that information about locks, keys, enemies, map completion, and health were the most relevant to differentiate play-styles. Said data is most important to creativity, mastery, and achievement play-styles, but immersion-related data was important on a minor scale. This difference may have been influenced by our game, that did not focus on immersion: the NPC dialogues were shallow, focused only on providing quests, and not deepening their characters or building a lore for the game's world. Therefore, not only is it important for developers to collect data from different gameplay aspects, but bias may be presented according to different gameplay focus for each game, and developers must study each game's data uniquely to understand how to understand their players.

For the second part of our analysis, in Section 4.2, we used the K-Means and Spectral Clustering algorithms, both showing that both the player profile and the dungeon's profile impact the gameplay data, meaning that the same player may behave differently according to the dungeon they play, and players with different profiles may play similarly in a given dungeon type. This may signify that for research on automatic player profiling given gameplay metrics, the dungeon's profile may also need to be used as input for an accurate machine learning predictor of player preferences and performance.

Our classifier results indicate that our sample was too small for good results, but with data augmentation, our gameplay metrics may lead to efficient classifiers. These may model players to match different combinations of player profiles using only their gameplay metrics. Our findings may allow developers to reduce their dependency on explicit data collection methods when modeling player behavior. This could reduce unwanted data collection interactions (such as profile selection or self-reporting questionnaires) that affect player experience. It also mitigates self-reporting biases. Thus, we highlight the importance of following ethical standards, respecting data protection norms, and transparency regarding data usage.

Player profiling is helpful for adaptive PCG, enabling more accurate player-adapted content, and leading to a possibly more engaging player experience. However, an inaccurate player profiling model can result in frustrating and disengaging content, negatively affecting player experience. Yet exploratory, our study provides a method that may increase player profiling accuracy by identifying relevant gameplay metrics relating to player profiles and suggesting the data imputation technique to increase accuracy. To the best of our knowledge, this is the first study to use data imputation to increase the accuracy of a player-profiling technique.

This approach could be particularly beneficial for indie and serious games with limited player bases. Unlike Melhart *et al*. [2019], who collected data from Ubisoft's Rainbow Six Siege, that reportedly has up to 70 million players[17], most indie and serious games have player bases in the hundreds or thousands. Our imputation technique may help them adapt their content for different users.

Accuracy is very relevant in this case, as failing to adapt content to the players may lead to boredom (if too easy) or frustration (if too hard), breaking their flow experience. However, if done correctly, this may lead players to experience different play-styles and observe how the contents adapt to their changes. Furthermore, an accurate content adaptation can increase long-term engagement, as it increases replayability and new experiences when playing with new styles. This is a common strategy in roguelike games, such as *Hades* and *The Binding of Isaac*, which use their variety of characters, items, and weapons and increasing difficulty to engage players in playing through the same levels even after hundreds of hours, and may be enhanced by adapting content orchestration.

Although the prospect of adaptation may increase immersion, it is important to be transparent with players about what content is adapted [Denisova and Cairns, 2019]. Moreover, players may not necessarily want a game that is constantly changing (for accessibility, preferences, or even speed-run related reasons), and a "static" version of the game, developed by human designers, should be present for players to select and play.

Another important issue is related to data privacy and respecting the user's will on sharing their data. Games that may use content-adapting techniques should always ask for the player's consent on using their data to adapt content and, by implementing this "static" version, they could also allow users who disagree on their data usage to still play the game.

Finally, our work may lead to a real-time feedback loop for content generators and orchestrators, enabling a virtually infinite re-playable experience, with the system adapting to changes in user preferences and skills. However, further studies using our classifier inside the game for a real-time feedback loop of PCG generation are needed, so we can verify if players liked and could perform well on content generated by their predicted profiles (after playing a first dungeon that will collect their gameplay data).

6 Limitations and Future Work

The study has some limitations and threats to validity that should be accounted for. However, they were mitigated to the best of our knowledge and available resources. The sample size used for this study is limited (n = 15), as mentioned in Pereira *et al*. [2024b]' work. We used data augmentation to mitigate it, but a larger sample is needed to draw further conclusions. Additionally, participants were invited only in Brazilian gaming communities, which may impact further generalizations of the results for other demographics. Thus, we present preliminary results that need to be applied to a larger and more diverse sample.

As we do not know the precise demographic information of the participants, pre-existing differences may have been introduced, such as the design of the game and persuasive elements appealing mostly to people already interested in gaming. This could introduce a selection bias and limit the adaptation's effectiveness for less engaged audiences. We collected data through pre-test questionnaires, which are susceptible to social desirability and response biases. We also assess gameplay data specific to our game, so studies with other games from the same genre, and games from different genres are necessary to measure generalization.

We did not consider in what order each level was played for a given player, so play-time and in-game experience may influence the results of the gameplay data.We recognize the importance of further investigations regarding emotional design, audio elements, and player modeling/profiling within content adaptation. Although an in-depth analysis of the relationship between the aforementioned elements is beyond the scope of this study, we consider it a valuable direction for future work [Isbister, 2016b; Nunes and Darin, 2024b].

Finally, we acknowledge the need to evaluate the player experience of procedurally orchestrated and profile-adapted content using a validated instrument [Borges *et al*., 2020], and the profiles should be evaluated with other validated questionnaires, such as Tondello *et al*. [2019]. In future work,

[17] `https://www.pcgamesn.com/rainbow-six-siege/player-count` (Accessed July 3, 2025)

we plan to use a validated instrument named *GAMEX*, for its wide acceptance and flexibility in use contexts [Eppmann *et al.*, 2018].

7 Conclusion

In this work, we extend the work of Pereira *et al.* [2024b], analyzing gameplay metrics compared to a player profiling method using PCA, clustering, and classification methods.

We collected gameplay data from 119 playthroughs of 15 anonymous players. The players were classified into 8 profiles using a pre-test questionnaire based on Yee and Ducheneaut [2018] player motivation factors. Players selected a playable dungeon from 12 options, each with a distinct level profile, with procedurally generated content of levels, narrative, and rules.

Using PCA in the gameplay data, we identified two components, the first with 57% and the second with 9% of explained variability. The first component was formed by data of *Lock Usage Rate*, *Enemy Kill Rate*, *Map Completion*, *Key Collected Rate*, and *Lost Health*. The second component is formed mainly by: *Completed Immersion Quests*, *Completed Report Quests*, *Key Collected Rate*, *Map Completion*, *Lock Usage Rate*, *Completed Achievement Quests*, *Completed Gather Quests*, and *Completed Give Quests*. These gameplay metrics were found relevant for understanding players' behavior without the need for explicit data collection (e.g., questionnaires, profile selection).

We could not identify player profiles using our player profiling methods from our clustering algorithms (K-Means and Spectral Clustering). However, we found *game difficulty* as an important level component for impact clustering. By using K-Means and Spectral Clustering, we identified categories related to levels and players: players who liked dungeon combat in hard or easy levels and players who disliked dungeon combat in hard or easy levels.

Despite the need for more experiments, our work provided some insights. The first insight is the significant effect of the dungeon's difficulty on the player profile's classification. To mitigate this, we plan to limit the dungeon's configurations for our experiments in player profiling, even if generality could be hindered. The second is the unexpected inability to identify players motivated by immersion or achievements from players' completed quests. The lack of the quest's complexity could explain the problem, and we plan to treat it and improve our quest's data collection and analysis.

We also trained a supervised algorithm to predict the player's profile using only their gameplay metrics. Using data augmentation, we were able to achieve a mean score of almost 95% accuracy using the Histogram-based Gradient Boosting technique, with low a standard deviation of 0.031. This means that it is possible, even with a small gameplay dataset, to correctly predict said profiles using only the gameplay data from the player's last played dungeon (or level). This paves the way for real-time content adaptation using content generators and enables game developers to rely less on self-reporting measures.

For future work, we plan to use classification algorithms to assess player profiling using only gameplay data, and use the results to predict a player's profile and feed our real-time procedural content orchestration algorithm to generate a new content and analyze players' opinions on the quality of generated content, the classification algorithm and both the player and the algorithm's performance. We also intend to strengthen the connection between emotional design, audio, and player profiling, as it may enhance this study's contributions.

Finally, we plan to test other player profiling methods, such as HEXAG [Tondello *et al.*, 2016], comparing the methodologies, their results when compared to gameplay metrics and assessing their use in content adaptation with procedural content generation.

Declarations

Acknowledgements

We thank the Editorial Committee of the event *Workshop sobre Interação e Pesquisa de Usuários no Desenvolvimento de Jogos* (WIPlay 2024) for inviting us to extend our work for this journal.

We also thank all the reviewers for their relevant insights, which considerably improved our work.

Acknowledgement of AI use

We used *Grammarly*, *LanguageTool*, and *ChatGPT* as AI tools to check for typos and increase writing structure. However, we do not generate any text using generative AI. All content and interpretations are the authors' own.

Funding

This research was funded by the National Council for Scientific and Technological Development (CNPq). This study was also financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) – Finance Code 001.

Authors' Contributions

Leonardo Tórtoro Pereira: conceptualization, data curation, formal analysis, investigation, methodology, software, and writing-review & editing.

T. Yuji Teoi: software, visualization, writing-original draft, and writing-review & editing.

Claudio F. M. Toledo: conceptualization and project administration.

Competing interests

All authors declare no conflict of interest.

Availability of data and materials

The data and scripts used for this research are publicly available at OSF[18]

[18] `https://osf.io/jqf39/?view_only=33bc172e52e145eaa85807edd451d497` (Accessed July 3, 2025)

Citation Diversity Statement

Here, we acknowledge our citation bias and quantify our cited authors' diversity. We consider the importance of citing diversity, given the relevance of citations in academia and the evident imbalance in gender citations [Dworkin *et al*., 2020; Zurn *et al*., 2020].

For quantifying our citation diversity, we used a public script[19] for analyzing our references according to the first and last authors, and its probability of they being a man or a woman [Zhou *et al*., 2022]. As a result, excluding self-citations to the first and last authors of our current paper, our references contain 12.53% woman (first)/woman (last), 22.56% man/woman, 22.71% woman/man, and 42.19% man/man.

The limitations of this method are:

- names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity;
- it cannot account for intersex, non-binary, or transgender people.

Using the same script, we predicted the racial and ethnic categories of the first and last authors. Excluding self-citations, our references contain 14.38% author of color (first)/author of color (last), 14.50% white author/author of color, 24.73% author of color/white author, and 46.38% white author/white author.

The method's limitations are:

- the mapping between names and databases with probabilities of a name being in a racial/ethnic group could be imprecise;
- it cannot account for Indigenous and mixed-race authors or those who may face differential biases due to the ambiguous racialization or ethnicization of their names.

We actively expect to improve our citation diversity in future works.

References

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., and Kurdi, H. (2021). Impact of dataset size on classification performance: An empirical evaluation in the medical domain. *Applied Sciences*, 11(2). DOI: https://doi.org/10.3390/app11020796.

Bernhaupt, R. (2015). *Game User Experience Evaluation*. Springer Publishing Company, Incorporated. DOI: https://doi.org/10.5555/2807325.

Borges, J. B., Juy, C. L., de Andrade Matos, I. S., Angelo Silveira, P. V., and Darin, T. d. G. R. (2020). Player experience evaluation: a brief panorama of instruments and research opportunities. *Journal on Interactive Systems*, 11(1):74–91. DOI: https://doi.org/10.5753/jis.2020.765.

Carneiro, N., Miranda, D., Pereira, G., Mendonça, G., and Darin, T. (2022). A systematic mapping on player's profiles: Motivations, behavior, and personality characteristics. *Journal on Interactive Systems*, 13(1):257–273. DOI: https://doi.org/10.5753/jis.2022.2572.

Cowley, B., Charles, D., Black, M., and Hickey, R. (2008). Toward an understanding of flow in video games. *Comput. Entertain.*, 6(2). DOI: https://doi.org/10.1145/1371216.1371223.

[19] https://github.com/dalejn/cleanBib/ (Accessed July 3, 2025)

de Lima, E. S., Feijó, B., and Furtado, A. L. (2021). Adaptive branching quests based on automated planning and story arcs. In *2021 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 9–18. DOI: https://doi.org/10.1109/SBGames54170.2021.00012.

Denisova, A. and Cairns, P. (2019). Player experience and deceptive expectations of difficulty adaptation in digital games. *Entertainment Computing*, 29:56–68. DOI: https://doi.org/10.1016/j.entcom.2018.12.001.

Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., and Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, 23(8):918–926. DOI: https://doi.org/10.1038/s41593-020-0658-y.

Ekwaro-Osire, H., Ponugupati, S. L., Al Noman, A., Bode, D., and Thoben, K.-D. (2025). Data augmentation for numerical data from manufacturing processes: an overview of techniques and assessment of when which techniques work. *Industrial Artificial Intelligence*, 3(1):1–19. DOI: https://doi.org/10.1007/s44244-024-00021-x.

Eppmann, R., Bekk, M., and Klein, K. (2018). Gameful experience in gamification: Construction and validation of a gameful experience scale [gamex]. *Journal of Interactive Marketing*, 43:98–115. DOI: https://doi.org/10.1016/j.intmar.2018.03.002.

Heijne, N. and Bakkes, S. (2017). Procedural zelda: a pcg environment for player experience research. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, FDG '17, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3102071.3102091.

Isbister, K. (2016a). *How Games Move Us - Emotion by Design*. MIT Press. DOI: https://doi.org/10.7551/mitpress/9267.001.0001.

Isbister, K. (2016b). *How games move us: Emotion by design*. Mit Press. DOI: https://doi.org/10.7551/mitpress/9267.001.0001.

Jacko, J. A. (2012). *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition*. CRC Press, Inc., Boca Raton, FL, USA, 3rd edition. DOI: https://doi.org/10.5555/2378709.

Kirchner-Krath, J., Altmeyer, M., Schürmann, L., Kordyaka, B., Morschheuser, B., Klock, A. C. T., Nacke, L., Hamari, J., and von Korflesch, H. F. (2024). Uncovering the theoretical basis of user types: An empirical analysis and critical discussion of user typologies in research on tailored gameful design. *International Journal of Human-Computer Studies*, 190:103314. DOI: https://doi.org/10.1016/j.ijhcs.2024.103314.

Koster, R. (2005). *A Theory of Fun for Game Design*. Paraglyph Press, Arizona. DOI: https://doi.org/10.5555/1207478.

Liapis, A., Yannakakis, G. N., Nelson, M. J., Preuss, M., and Bidarra, R. (2019). Orchestrating game generation. *IEEE Transactions on Games*, 11(1):48–68. DOI: https://doi.org/10.1109/TG.2018.2870876.

Liapis, A., Yannakakis, G. N., and Togelius, J. (2014). Computational game creativity. In *Proceedings of the fifth International Conference on Computational Creativity*.

Loria, E. and Marconi, A. (2018). Player types and player behaviors: Analyzing correlations in an on-the-field gamified system. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, CHI PLAY 2018, Melbourne, VIC, Australia, October 28-31, 2018*, pages 531–538. DOI: https://doi.org/10.1145/3270316.3271526.

Machado, M. C., Fantini, E. P. C., and Chaimowicz, L. (2011). Player modeling: What is it? how to do it? In *SBGames 2011 - Tutorials Track*.

Magy Seif El-Nasr, Anders Drachen, A. C. (2013). *Game Analytics: Maximizing the Value of Player Data*. Springer London. DOI: https://doi.org/10.1007/978-1-4471-4769-5.

Melhart, D., Azadvar, A., Canossa, A., Liapis, A., and Yannakakis, G. N. (2019). Your gameplay says it all: Modelling motivation in tom clancy's the division. In *2019 IEEE Conference on Games (CoG)*, pages 1–8. DOI: https://doi.org/10.1109/CIG.2019.8848123.

Microsoft (2020). Microsoft purchases 'minecraft'. Accessed in: 2025-01-13.

Nunes, C. and Darin, T. (2024a). Echoes of player experience: A literature review on audio assessment and player experience in games. *Proc. ACM Hum.-Comput. Interact.*, 8(CHI PLAY). DOI: https://doi.org/10.1145/3677069.

Nunes, C. and Darin, T. (2024b). Echoes of player experience: A literature review on audio assessment and player experience in games. *Proc. ACM Hum.-Comput. Interact.*, 8(CHI PLAY). DOI: https://doi.org/10.1145/3677069.

Pereira, L., Rodrigues, K., Toledo, C., and Teoi, T. (2024a). Understanding players to enhance their fun: how to extract player data and motivation factors for procedural content generation. In *Anais Estendidos do XXIII Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 37–42, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/sbgames_estendido.2024.241181.

Pereira, L., Teoi, T., and Toledo, C. (2024b). Towards evaluating a procedural content orchestrator gameplay data to differentiate user profiles. In *Anais do III Workshop sobre Interação e Pesquisa de Usuários no Desenvolvimento de Jogos*, pages 75–86, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/wiplay.2024.245483.

Pereira, L. T. (2022). *Arquitetura e desenvolvimento de um sistema de geração procedural de múltiplos conteúdos para jogos eletrônicos em tempo real*. PhD thesis, Universidade de São Paulo, São Carlos. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação.

Pereira, L. T., Viana, B. M. F., and Toledo, C. F. M. (2022). A system for orchestrating multiple procedurally generated content for different player profiles. *IEEE Transactions on Games*, pages 1–11. DOI: https://doi.org/10.1109/TG.2022.3213781.

Salen, K. and Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. The MIT Press. DOI: https://doi.org/10.5555/1215723.

Teoi, T., Pereira, L., and Toledo, C. (2024). Towards adapting a content orchestrator to a different gamegenre: Generating levels, rules, and narrative for diverse player profiles from a top-down adventure to a 2d platformer. In *Anais do XXIII Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 396–415, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/sbgames.2024.240282.

Togelius, J., Shaker, N., and Nelson, M. J. (2016). *Introduction*, pages 1–15. Springer International Publishing, Cham. DOI: https://doi.org/10.1007/978-3-319-42716-4_1.

Togelius, J., Yannakakis, G. N., Stanley, K. O., and Browne, C. (2011). Search-based procedural content generation: A taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3):172–186. DOI: https://doi.org/10.1109/TCIAIG.2011.2148116.

Tondello, G. F., Arrambide, K., Ribeiro, G., Cen, A. J.-l., and Nacke, L. E. (2019). "i don't fit into a single type": A trait model and scale of game playing preferences. In Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M., and Zaphiris, P., editors, *Human-Computer Interaction – INTERACT 2019*, pages 375–395, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-29384-0_23.

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., and Nacke, L. E. (2016). The gamification user types hexad scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '16, page 229–243, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/2967934.2968082.

Yannakakis, G. N., Spronck, P., Loiacono, D., and André, E. (2013). Player Modeling. In Lucas, S. M., Mateas, M., Preuss, M., Spronck, P., and Togelius, J., editors, *Artificial and Computational Intelligence in Games*, volume 6 of *Dagstuhl Follow-Ups*, pages 45–59. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany. DOI: https://doi.org/10.4230/DFU.Vol6.12191.45.

Yannakakis, G. N. and Togelius, J. (2018). *Modeling Players*, pages 203–255. Springer International Publishing, Cham. DOI: https://doi.org/10.1007/978-3-319-63519-4_5.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, 9(6):772–775. PMID: 17201605. DOI: https://doi.org/10.1089/cpb.2006.9.772.

Yee, N. and Ducheneaut, N. (2018). Gamer motivation profiling: Uses and applications. *Games User Research*, pages 485–490. DOI: https://doi.org/10.1093/oso/9780198794844.003.0028.

Zaki, M. J., Meira Jr, W., and Meira, W. (2020). *Data mining and machine learning: Fundamental concepts and algorithms*. Cambridge University Press.

Zhou, D., Stiso, J., Bertolero, M., Cornblath, E., Teich, E., Blevins, A. S., Murphy, K., Oudyk, K., Michael, C., Urai, A., Matelsky, J., Virtualmario, Camp, C., Castillo, R. A., Saxe, R., Dworkin, J., and Bassett, D. (2022). dalejn/cleanbib: v1.1.3.

Zurn, P., Bassett, D. S., and Rust, N. C. (2020). The citation diversity statement: a practice of transparency, a way of life. *Trends in Cognitive Sciences*, 24(9):669–672. DOI: https://doi.org/10.1016/j.tics.2020.06.009.