International Journal of Serious Games

ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

Exploring the Impact of Player Traits on the Leaderboard Experience in a Digital Maths Game

André Almo[1], Mariana Rocha[1], Attracta Brennan[2] and Pierpaolo Dondio[1]

[1]*Technological University Dublin, School of Computer Science, Dublin, Ireland*
[2]*University of Galway, Computer Science Department, Galway, Ireland*
{andre.almo, mariana.rocha, pierpaolo.dondio} @tudublin.ie; {attracta.brennan}@universityofgalway.ie

Keywords:

Game-Based Learning
Game Design
Maths Anxiety

Received: May 2024
Accepted: November 2024
Published: November 2024
DOI: 10.17083/ijsg.v11i4.794

Abstract

Digital Game-Based Learning (DGBL) uses digital games to enhance engagement and learning, but its efficacy is linked to game features. Leaderboards are a commonly used feature to increase motivation through competition, improving engagement and learning outcomes. However, infinite leaderboards, displaying all players in an ordered ranking based on their scores, can demotivate players depending on their performance and characteristics. This study investigated primary school students' experience with an infinite leaderboard during a digital maths game intervention. 1389 Irish students participated in a 6-week programme with the game 'Seven Spells,' which featured an infinite leaderboard. Player traits and opinions about the game and its leaderboard were gathered via questionnaires and surveys. Leaderboard enjoyment was influenced by players' position on the leaderboard and maths anxiety levels. Maths anxious players disliked the leaderboard more than non-anxious players, even after controlling for their position on the leaderboard. How much players liked to play against each other was also found to be a significant factor predicting the enjoyment of the leaderboard. There was also a small correlation between leaderboard enjoyment and overall game enjoyment. These insights exemplify players' characteristics that should be considered when incorporating infinite leaderboards into maths games to avoid negative impacts on gaming experiences.

1. Introduction

Digital game-based learning (DGBL) is defined as "the use of games within an existing lesson, classroom, or other instructional contexts where the intent is at least as much to learn rather than to (exclusively) have fun" (p. 144) [1]. Applying games in the classroom is distinct from gamification, which involves adding game elements in non-game contexts, such as previously existing educational activities [2]. Games, on the other hand, must be interactive and have original rules, goals and quantifiable measures of progress [2]. In educational contexts, it is common to use 'serious games', designed with an additional purpose other than fun [3, 4, 5].

DGBL provides an opportunity to harness students' fascination with technology and computer games whilst delivering educational content [6, 7]. The efficacy of game-based learning has been widely demonstrated [8]. DGBL supports enjoyment, active learning, motivation and engagement; as a result, its application has a positive effect on learning, which can lead to better performance [8, 9, 10, 11]. A recent meta-analysis analysing the impact of school-based DGBL interventions when compared to traditional instruction [12] suggested that game-based approaches led to medium positive effects on overall learning and cognitive learning outcomes as well as having a small influence on affective and motivational learning outcomes [12].

As this study focuses on a digital maths game, it's important to review the evidence of how DGBL has affected this subject area thus far. Critical to everyday life, maths is used in numerous activities, from cooking and measuring to shopping and budgeting [13]. Maths also underpins technological, engineering and scientific knowledge [14, 15]. Studies show a direct relationship between executive functions (i.e. planning, analysing and solving problems, managing time) and mathematical competency [16]. Despite maths being a core element of educational systems [17] and its importance to daily life, academic and professional success [18], it is common for not only children, but also adults to experience difficulties and anxiety with this subject [13, 19]. Maths Anxiety (MA) is defined as feelings of stress, apprehension, and/or anxiety when one is faced with maths, including in classroom settings [20, 21, 22].

Digital games have been shown to support students' maths skills in areas such as geometry, algebra and arithmetic procedures, contributing to higher learning gains in comparison to traditional teaching [13, 23]. Learning through games can support problem-solving and critical thinking, and can help students comprehend abstract mathematical concepts [24]. A recent meta-analysis investigated the impact of digital games on students' learning achievements in different STEM subjects [25]. The authors included 33 studies published in the last ten years and reported an overall moderate effect size of 0.629 for maths. [26] suggests that learning through games positively promotes engagement and increases students' self-confidence, in addition to fostering positive attitudes towards maths. DGBL is also regarded as a potential intervention to integrate creativity within maths teaching [27, 10], foster positive views on maths and mitigate maths anxiety, despite the evidence to date being inconclusive [28]. However, the efficacy of DGBL relies on incorporating game features supporting enjoyment, active learning and engagement, all of which are interconnected [8, 10]. The choice of such features can influence a player's game experience [27] and, consequently, their learning experience. The use of game features to promote challenge and competition, for instance, has been shown to foster engagement and result in a deeper knowledge of the learning content [29, 30]. Moreover, the social aspect of competition is frequently perceived to motivate and engage players [31]. One of the most common approaches to promote competition, goals and performance-based feedback in gamified educational settings is the use of leaderboards [32, 33]. Leaderboards display game scores achieved by players, enabling more immediate and direct player-to-player comparison in a given group [32, 33, 34].

Leaderboards have been shown to be effective in enhancing a player's engagement and self-reported levels of motivation [32, 33, 34, 35]. However, leaderboards do not affect all players equally. In a classroom setting, a leaderboard's display of student achievement can increase social comparison and peer pressure to the point of becoming detrimental to some players' game experience, especially those unable to cope with higher levels of competition, social anxiety and/or stress; for these, leaderboards are a potentially harmful mechanic [15, 36, 34]. A study on the use of a leaderboard in a maths game team competition found that although the leaderboard tended to motivate the students, team rank, team commitment and enjoyment of the game predicted leaderboard motivation, with players from less successful teams being less motivated by the presence of the leaderboard [32]. Similarly, previous research has shown that students positioned in the lower ranks of the leaderboard can feel inadequate and may not respond positively due to a negative comparison of themselves against those players who are higher placed in the ranks [33]. In a 2013 study, students considered the leaderboard to be a demotivating factor, acknowledging that the growing gap between the highest-scoring students and the rest of the class discouraged them from scoring more points in the game [37]. A 2024 study also showed a negative effect of a leaderboard on student motivation [38]. Featuring a longitudinal quasi-experimental design conducted with higher education students, the aforementioned study applied a leaderboard to an asynchronous online course. The results showed that this feature decreased motivation, further highlighting the careful consideration needed when including a leaderboard in classrooms [38]. It seems reasonable to assume that in a classroom environment, introducing a leaderboard and encouraging social comparison could lead to negative outcomes such as embarrassment and a decrease in confidence for at least a subset of students. This is still however, an open question. For instance, a recent study collecting students' heart rate variability — as a measure of stress — when participating in a gamified activity with a leaderboard found no significant association between this physiological measurement and leaderboard position [34].

Identifying factors that shape players' views on the leaderboard is a key research focus. Aside from their leaderboard rank, other factors may affect a player's perception of this game feature. One study found gender to be a key factor, showing that male students engaged more than their female counterparts with the leaderboard, often checking it and comparing scores with opponents in an e-learning setting [39]. Personality traits may also impact how players engage with and evaluate the leaderboard [40, 41]. As an example, one study found that a player's trait competitiveness can be correlated to their position on the leaderboard [41]. Extraversion and introversion can also be relevant; in a study on a gamified Learning Management System at a public university, all extroverted students enjoyed the leaderboard, whereas just over half of the introverted students favoured it [40].

It is important to note however, that the mixed results regarding the effects of leaderboards might be connected not only to students' characteristics and performances, but also to the leaderboard design [42]. Infinite (or absolute) leaderboards are the most frequently used type of leaderboard in educational settings [32, 42]. In this design, the leaderboard displays the scores of players in an ordered ranking, from the top-scoring players (at the top of the leaderboard) to the lowest-scoring ones (at the bottom of the leaderboard), and players can view the positions of all other players.

This study analyses students' experience with an infinite leaderboard in a digital maths game. An observational study was conducted with 1389 Irish primary school children who participated in a 6-week DGBL programme, playing the digital maths game 'Seven Spells' [44,64]. The goal was to identify the factors predicting players' enjoyment of an infinite leaderboard. Based on prior research, players' perceptions of the leaderboard feature—whether positive or negative—may be influenced by their in-game performance (e.g., their ranking on the leaderboard) and emotional factors, such as anxiety.

More specifically, maths anxiety (MA) rather than generalised anxiety was used in this study due to the maths context of the games played during the school intervention. MA represents a more specific and accurate predictor of student performance and behaviour in maths tasks [43]. Moreover, since MA may relate not only to maths itself but also to the social experience of completing maths activities [18, 22], there is a potential interplay between MA and social comparison pressures promoted by the leaderboard. Our hypotheses were that liking the leaderboard would be positively associated with a player's position on the leaderboard (i.e. their game performance), and negatively associated with their level of MA.

In addition to these two main predictors, a set of covariates was also considered: gender, age, maths and literacy abilities, and game-related variables such as players' enjoyment of 'Seven Spells' and their video game habits.

More specifically, this study addressed the following research questions:

– RQ1A. What are the significant academic, demographic, emotional and game experience factors that can predict whether a player will report a positive or negative experience with an infinite leaderboard in a digital maths game?

– RQ1B. What level of accuracy can be achieved by a model predicting a player's opinion of the leaderboard, using a combination of demographic data and cognitive and emotional traits of players?

– RQ2. Is the level of enjoyment of the leaderboard associated with the overall enjoyment of the Seven Spells game?

– RQ3. What aspects of the leaderboard lead to higher or lower enjoyment of this feature?

Our hypotheses for RQ1A have been already discussed. Regarding RQ1B, our aim was to quantify the accuracy of a model predicting whether players would like the leaderboard. Machine learning models were used to address this question, with the goal of maximizing prediction accuracy rather than explaining the relationships between input factors and the players' enjoyment of the leaderboard—relationships already investigated in RQ1A. The rationale behind RQ1B was to provide an automatic tool that may be used to predict players' preferences regarding the leaderboard in game-based learning environments and adaptively switching on or off this game element to maximise players' game experience.

Regarding RQ2, our hypothesis was that the player's level of enjoyment with the infinite leaderboard would be significantly and positively associated with their overall enjoyment of the maths game. If verified, this would provide evidence regarding the importance and impact of a leaderboard in educational games. Finally, our hypothesis for RQ3 was that players' comments about the leaderboard could offer insights into what aspects of this feature are potentially connected to leaderboard enjoyment. The answers to the above research questions can help game designers and developers consider to what extent leaderboards should be used in a learning context and how to design or adapt leaderboards based on player characteristics.

2. Methods

2.1 Study design

Data collection was performed from January 2022 to March 2024 as part of 'Happy Maths' [44,64], a research project investigating the effects of game-based learning on primary school students' numerical cognition and maths anxiety (MA). Irish primary schools were contacted via e-mail and invited to participate in the Happy Maths programme. Schools received a protocol detailing the game used, visit frequency and duration, study aims, and an outline of the data to be collected from students. Only those children whose guardians permitted participation were included in the study. In compliance with GDPR, teachers distributed a list of randomly generated usernames to the children. The teachers were also responsible for linking each username to the child's age, gender, and score on the standardised national mathematics and literacy test (STen scores).

2.1.1 Participants

A total of 1389 students from 78 classrooms of 25 Irish primary schools participated in this study. Participants ranged in age from 8 to 11 years old, corresponding to 3rd, 4th, 5th, and 6th classes of the Irish educational system. Of the 78 classes included in the study, 106 students were from 3rd class; 515 from 4th class; 481 from 5th class; and 287 from 6th class.

2.1.2 Game description

Throughout the study, students engaged with the digital maths game, Seven Spells [44,64] developed by the authors. In this game, players enter a fictional martial arts academy, learning the 'ancient art of number fighting'. The goal of the game is to capture number cards using maths skills such as arithmetic operations (addition, subtraction, multiplication, and division), knowledge of mathematical concepts (even and odd numbers, prime numbers, numeric tables, intervals, greater and smaller, etc.) and an ability to combine and manipulate numbers.

During the study, students played Seven Spells in three different game modes: a *solo* game mode (i.e. without an opponent); a *vs. human* game mode (i.e. against another human player) and a *vs. CPU* game mode (i.e. against a computer-controlled character). All the classes followed the same sequence of sessions, as described in Table 1.

Table 1. Study design

Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Questionnaires Game session (solo) + leaderboard	Game session (solo) + leaderboard	Game session (vs. CPU)	Game session (vs. human)	Game session (vs. human)	Game session (solo) + leaderboard Feedback survey

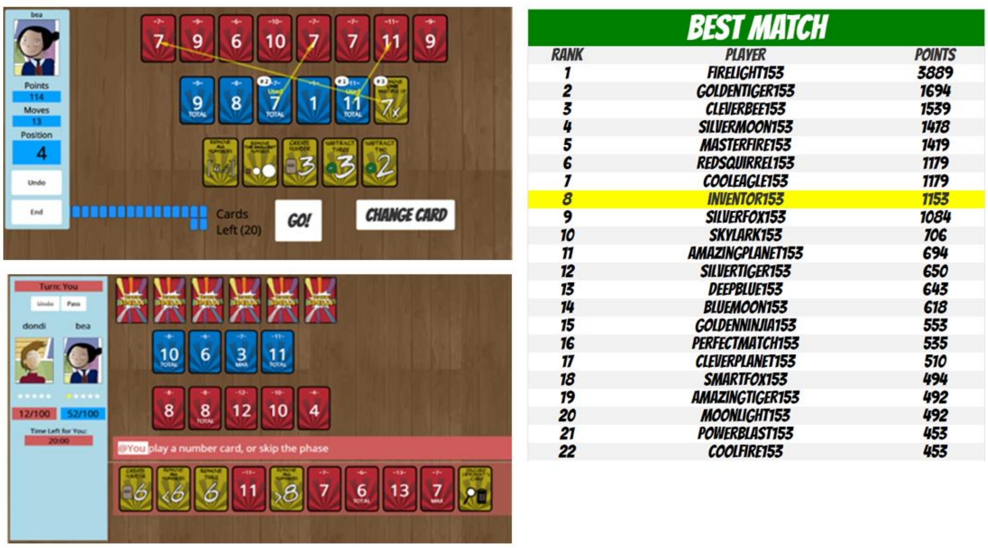

Figure 1. The Seven Spells *solo* game mode (top left), the *vs. human* game mode (bottom left) and the solo game mode leaderboard (right)

This study focused on the leaderboard present in the *solo* game mode. The other two games did not have a leaderboard. The *solo* game mode used an infinite design class leaderboard [32, 42]. This class leaderboard showed all the players in the class in an ordered ranking according to their scores. The class leaderboard was also displayed on the classroom whiteboard during the in-class game sessions. It was continuously updated in real-time, to reflect the evolving scores and changes in position rankings as the games progressed.

2.1.3 Intervention and variables collected

This study adopted an observational approach, in which one-hour weekly DGBL sessions were delivered in-class for six consecutive weeks.

During Session 1, information on the students' gaming habits was collected, in addition to cognitive and non-cognitive measurements. The students were asked how much they enjoyed playing video games and how much time they spent playing video games daily. They were also asked to complete the Modified Abbreviated Maths Anxiety scale (mAMAS) [21], a validated psychometric scale for primary school children (aged 8 to 13 years old) designed to measure their level of MA. The questionnaire consists of a 5-point Likert scale with nine questions, where the students were asked to express their feelings about maths in situations involving maths learning and testing. Using the Irish national standardised tests for maths and English, two cognitive measures were collected for each student: their level of maths ability and their literacy ability [45]. These scores were provided on a STen scale from 1 to 10 (STen scores). The gender and the class grade of each student were also collected.

After applying the questionnaires, ‘Seven Spells’ was presented through a detailed tutorial explaining the game’s rules, objectives and mechanics, in addition to an illustrative guided match.

After the end of the sixth session, when the three game modes had been played, students answered a short feedback survey about their game experience. This survey included a question asking the students whether they enjoyed the presence of the leaderboard in the *solo* game mode. There were four possible answers to this question: “I did not like the leaderboard”; “I kind of liked the leaderboard”; “I liked the leaderboard”; “I loved the leaderboard”. The answers to this question represented our outcome variable.

Additionally, students could provide comments in the blank space provided. The students were also asked to rate the *solo, vs. human* and *vs. CPU* game modes on a scale from 1 to 10. The ratings assigned to these game modes were introduced as key control variables. Indeed, even if the leaderboard was not present in the *vs. CPU* and *vs. human* game modes, players’ game experience with these two modes could have influenced their perceived experience about the leaderboard, present in the *solo* game mode. Therefore, we considered the ratings assigned to the *vs. human* and *vs. CPU* games as covariates to better control the relationship between our two main predictors (MA and in-game performance) and liking the leaderboard.

Moreover, there were further reasons to check if there was an association between the rating of the *vs. human* game and liking the leaderboard. In the *solo* game mode, the leaderboard introduced a competitive form of social comparison focusing on ranking players within a class. Playing against classmates in the *vs. human* game mode was also a form of social comparison (a more direct peer comparison rather than a class-wide comparison) that could have suited competitive players or players open to playing the game socially rather than individually. Therefore, since competition and social comparison were present in both the *solo* and the *vs. human* game modes, our hypothesis was that a positive association between the rating of the *vs. human* game mode and liking the leaderboard could exist, on the basis that players that liked (disliked) playing with a classmate also liked (disliked) to be compared with all the other classmates on the leaderboard.

This is also why this study tested for an association between the ratings for the different game modes (*solo, vs. human* and *vs. CPU*) and the level of enjoyment of the leaderboard, exploring to what extent the overall enjoyment of the game modes was related with the players’ experience with the leaderboard (RQ2).

The description of the variables collected in this study is presented in Tables 2, 3 and 4. Out of the 1389 records, data on leaderboard preference and game mode ratings (solo; vs. human; vs. CPU) were provided by 1257 students. However, only 984 records were fully complete.

Table 2. Description of the variables included in the study

Variable	Description
MA	Score in the modified Abbreviated Mathematics Anxiety Scale (mAMAS) [21]
Maths STen score (MS)	The students' result in the national mathematics standardised test
Literacy STen score (LS)	The students' results in the national literacy standardised test
Class year	The class year of the student
Age	Age of the student
Gender	The gender of the student (M/F)
Position on the leaderboard	Normalised position of the student on the class leaderboard. The variable is in the range [0,1], where 1 is the top position in the leaderboard
Opinion on video games	To what extent does the player like video games in general (Likert scale from 1 to 5)
Scores for each game mode	The 1 to 10 rating assigned by each player to the game experience for each game mode
Opinion on the leaderboard	Whether the students liked the leaderboard or not. Four possible choices ("I loved the leaderboard" / "I liked the leaderboard" / "I kind of liked the leaderboard" / "I did not like the leaderboard")

Table 3. Numerical study variables (N=984 for MA, MS, LS, Pos; 1257 for RateHum, RateCPU, RateSolo)

Variable	Description	Mean	SD	Range
MA	Maths Anxiety score (mAMAS)	18.6	7.69	9-45
MS	Maths Score	5.96	2.07	1-10
LS	Literacy Score	5.81	2.04	1-10
Pos	Normalized Rank of a player in the class leaderboard (1= highest rank)	0.5	0.3	0-1
RateHum	Rate the *vs. human* game version	9.15	1.55	1-10
RateCPU	Rate the *vs. CPU* game version	7.38	2.22	1-10
RateSolo	Rate the solo game version	6.88	2.36	1-10

Table 4. Categorical study variables (N=984 for Class Yr, Gender, LikeVideo, PlayTime; 1257 for Y). The variable Class Yr refers to levels in the Irish school system: 3rd class refers to 8-year-old children, 4th class refers to 9-year-old children, 5th class refers to 10-year-old children and 6th class refers to 11-year-old children)

Variable	Description	Distribution
Class Yr	Class grade	6.84% (3rd), 36.60% (4th), 35.08% (5th), 21.48% (6th)
Gender	Binary variable (M, F)	F: 48.37%, M: 51.63%
LikeVideo	Do you like video games? Likert scale (1=love 5=hate)	1: 60.35%, 2: 25.22%, 3: 9.54%, 4: 1.76%, 5: 3.14%
PlayTime	How much do you play video games daily?	Never: 6.72%, ≤ 1 hr: 49.57%, 1-2 hr: 37.00%, ≥ 2 hr: 6.72%
Y	Did you like the leaderboard?	Loved: 34.93%, Liked: 45.19%, Kind of liked: 16.87%, Did not like: 3.02%

2.2 Quantitative data analysis

Data analysis was carried out using R version 4.2.3 [46]. The distributions of students' preferences for the leaderboard and the scores assigned to each game mode were analysed using the Shapiro-Wilk normality test; due to this study's larger sample size, skewness and kurtosis were also analysed. In the case of a normal distribution, a t-test was to be used to evaluate statistical differences between groups. Conversely, if the distribution deviated from normality, a non-parametric test, the Kruskall-Wallis test, was chosen. In addition to comparing means, possible correlations between students' opinions about the leaderboard and the scores given to each game mode (RQ2) were investigated through Pearson or Spearman rank correlation, depending on the data's distribution.

In order to answer research questions RQ1A and RQ1B, different strategies were followed. RQ1A required explainable models that could report the association between predictors and the target variable. Therefore, we used an ordinal regression model where the target variable was represented by the four possible answers to the 'Did you like the leaderboard?' question. As RQ1B focused on model accuracy we decided to address it using machine learning (ML). The decision to use ML techniques stems from their advantages over traditional statistical methods in the context of predictive modelling [47]. Traditional statistical methods are limited by assumptions of linearity, normality, and multicollinearity, while the ML algorithms used in this study (decision tree, random forest, and XGBoost) are capable of capturing complex and non-linear relationships among variables without requiring such assumptions [47]. Additionally, ML models are more flexible in handling high-dimensional data and can automatically account for interactions between predictors that might be difficult to identify using traditional methods [47]. The size of our dataset and the number of input features excluded the use of deep-learning techniques. Hence, we decided to focus on tree-based algorithms (i.e. decision tree, random forest and XG boost), which are known for their generally high performance. Finally, the decision to use ML aligns with current trends in educational research and expectations of ML use in educational settings, where predictive accuracy is often prioritised for practical applications, such as personalised learning interventions [48].

The predictors used in the ML models are shown in Tables 3 and 4. As can be seen from Table 4, the present dataset was imbalanced, since the majority of players (80.12%) either liked or loved the leaderboard ("I liked the leaderboard"; "I loved the leaderboard"), and a minority (19.88%) had a more moderate opinion or disliked the leaderboard ("I kind of liked the leaderboard"; "I did not like the leaderboard"). To address the imbalance between the levels of the target variable and avoid an overly optimistic accuracy value, the SMOTE [49] oversampling technique was applied to balance the dataset.

Finally, in order to test the ML models, the full dataset was split into two subsets, 80% of the instances for training the models and 20% for testing the models.

2.3 Players' comments

Players' comments were extracted from the feedback survey completed in the sixth session. Only those comments containing the keyword 'leaderboard' were selected. Feedback concerning the leaderboard is offered as subjective evidence. Although these comments consisted of narrative data and anecdotes, they provided valuable insights from the players, augmenting the quantitative analyses conducted in the study. In order to better visualise the content of players' comments, a world cloud was created showing the most commonly-occurring words. Comments were pre-processed before analysis following a text-mining pre-processing pipeline. First, the term 'leaderboard' was excluded as it was the primary keyword in the initial search query and could dominate the comments. Tokenisation was then applied to break down the text into individual units, such as words, sentences, or sub-words, facilitating the identification of common topics in future analyses [50]. The next step involved removing stop words; i.e. common English words like "the" and "is" and then generating a word cloud to visualise the most frequently used terms by players. Finally, a stemming process was performed to simplify words to their root forms by removing prefixes and suffixes. This step prepared the dataset for word frequency analysis, resulting in bar plots representing the distribution of word frequencies. The analysis used Python language, adopting the Pandas, Wordcloud, Matplotlib and Seaborn libraries [50, 51, 52, 53, 54]. The integration of students' comments with quantitative analyses of their content constitutes a mixed-methods approach. This approach facilitates a thorough examination of students' opinions on the leaderboard, blending subjective narratives with objective frequency analysis for a comprehensive understanding.

3. Results

3.1 Descriptive statistics

While the dataset contained complete data for 984 players, 1257 players scored the three game modes (*solo*, *vs. human* and *vs. CPU*) and rated the infinite leaderboard feature according to their enjoyment. The target variable Y contained 37 negative outcomes ("Did not like"; 3.02%), 213 more moderate answers ("Kind of liked"; 16.87%), and 1007 positive outcomes ("Loved" (34.93%) and "Liked" (45.19%); 80.12% in total), showing that players generally enjoyed the leaderboard. Table 3 outlines the numerical study variables, providing information on the mean, standard deviation (SD), and range (minimum and maximum values). Meanwhile, Table 4 presents the study's categorical variables, offering an overview of the distribution of levels for each collected variable.

The MA variable represents the level of maths anxiety of each student measured by the score on the mAMAS scale. MS — Maths Score — refers to their score on the national mathematics standardised test (Maths STen score). The variable Pos measured the position of a player on the class leaderboard normalised in relation to each class, so that the top position in each class has a value of 1 whilst the bottom position has a value of 0. RateHum, RateCPU and RateSolo refer, consecutively, to the scores given to the *vs. human, vs. CPU* and the *solo* game modes. As indicated in the RateHum variable in Table 3 (which shows the score given to the *vs. human game* mode), the players enjoyed playing this game mode, with an average score of 9.15 out of 10. This variable is strongly right-skewed, with 61.46% of players giving a score of 10 and only 2.24% of players assigning a score less than 5 (Figure 2A). Contrary to the overwhelmingly positive opinions on the *vs. human* game mode, the ratings given to the *vs. CPU* and *the solo* game modes were much lower, with an average of 7.38 and 6.88 respectively (Table 3; Figure 2B; Figure 2C). The variable Class Yr displays the players' school year. As can be seen, the majority of players were from 4th class (36.60%) and 5th class (35.08%). There was also a significant number of players from 6th class (21.48%), and a small minority from 3rd class (6.84%). Gender shows that 48.37% of the players were female, and 51.63% male. The variable measuring how much a player liked playing video games (Likevideo) shows a strong right-skew (similar to RateHum). 4.9% of the players reported that they disliked or did not enjoy playing video games while the majority (60.35%) indicated that they loved playing video games. Regarding PlayTime, most played for less than an hour (49.57%) or between 1 and 2 hours (37.00%) per day. Finally, the enjoyment of the leaderboard in Seven Spells also showed a skewed distribution, with most players having loved (34.93%) or liked the feature (45.19%) whilst only 3.02% did not like the leaderboard (Table 4).

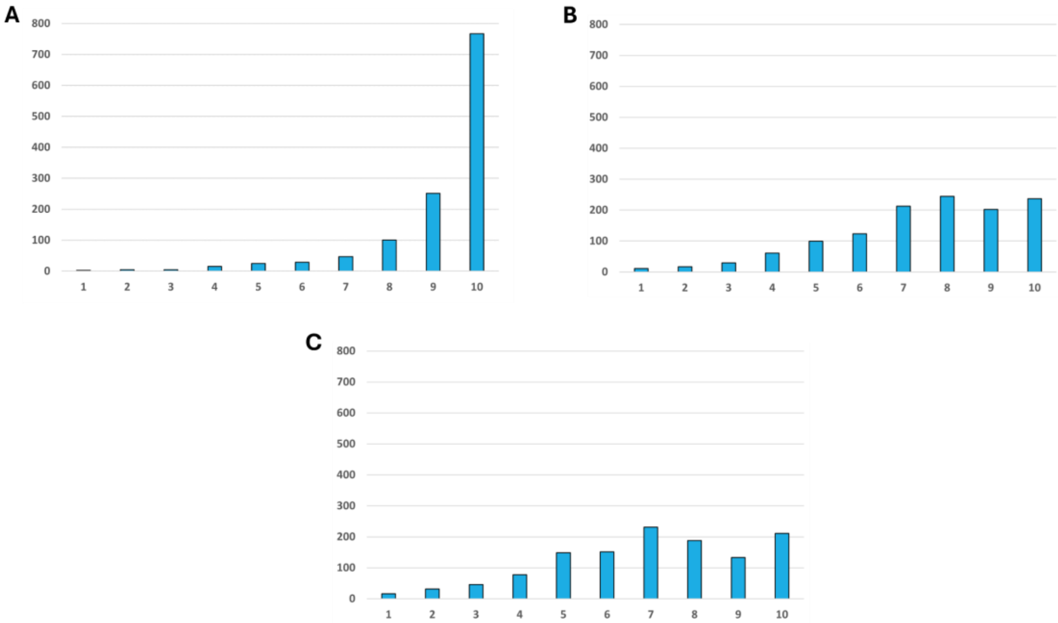

Figure 2. Distribution of scores given to the three game modes. A: *vs. human game* mode. B: *vs. CPU game* mode. C: *solo game mode*.

3.2 Research Question 1A: What are the significant factors predicting whether a player will report a positive or negative experience with the leaderboard?

The target variable Y i.e. enjoyment of the leaderboard, was modelled as an ordinal variable with four levels: "did not like the leaderboard", "kind of liked the leaderboard", "liked the leaderboard" and "loved the leaderboard". Table 5 shows the correlations between the variables considered. The variables 'leaderboard position' (Pos), 'Maths Anxiety' (MA) and the ratings assigned to the game modes displayed the highest correlation with the target variable Y.

Regarding correlations between predictors, there was a high correlation between Maths Scores (MS) and Literacy Scores (LS) (ρ=0.64), and between time spent playing videogames habitually (TimeP) and the liking of video games (LikeV), with ρ=0.51.

Table 5. Pearson correlations between the variables considered in the study. The significance levels are as follows: * = <0.05 , ** = <0.01, *** = <0.001

	Pos	LikeV	Rate CPU	Rate Solo	Rate Hum.	PlayT	C. Yr	MA	LS	MS
Y	0.26 ***	-0.10 **	0.18 ***	0.14 ***	0.33 ***	0.075 *	0.02	-0.17 ***	0.05	0.12 ***
Pos		-0.11 **	0.18 ***	0.10 **	0.10 **	0.11 **	0.03	-0.21 ***	-.31 ***	0.50 ***
LikeV			-0.05	-0.04	-0.05	0.52 ***	0.01	0.04	0.04	-0.03
Rate CPU				0.34 ***	0.21 ***	-0.02	0.08	-0.12 ***	0.11 **	0.13 ***
Rate Solo					0.12 ***	0.03	0.02	-0.08 *	0.01	0.06
Rate Hum.						0.00	0.05	-0.14 ***	0.06	0.05
PlayT							0.06	0.06	-0.09 *	0.00
Cl.Y								-0.09 *	0.06	0.02
MA									-0.25 ***	-0.34 ***
LS										0.64 ***

The first step of our analysis involved checking if the two main predictors — leaderboard position and maths anxiety (MA) — were significantly associated with the target variable. Table 6 reports the output of three ordinal regression models where the target variable Y is the enjoyment of the leaderboard. Models M0a and M0b show how MA and position on the leaderboard were significant predictors individually, establishing that there was a relationship between each of these predictors and the enjoyment of the leaderboard. Model M1 shows how the two predictors remained significant when present together. Leaderboard position (Pos) was more significant than MA. As expected, MA was negatively associated with Y while Pos was positively associated. The value of AIC for each model is also reported as a measure of goodness of fit.

Table 6. Baseline ordinal regression models. Target variable: Y (like the leaderboard). All the variables have been normalised using z-scores. The intercepts are omitted.

Model	Predictors	B	Pr (> |z|)	AIC
M0a	MA	-0.17	0.0006	3016
M0b	Pos	0.35	<0.0001	2986
M1	MA	-0.13	0.007	2979
	Pos	0.33	<0.0001	

We then considered the role of the covariate 'Maths Score' (MS). The variable had a high correlation with the position on the leaderboard (ρ=0.5). This was expected, as a significant portion of game performance in a maths game is likely a result of the player's mathematical abilities. Based on this, we hypothesised that there could be an effect of maths abilities on the leaderboard enjoyment, but this effect could be mediated by the leaderboard position of the player. Figure 3 and Table 7 show the result of a mediation analysis testing this hypothesis [51]. The position on the leaderboard (Pos) was a full mediator of the relationship between maths abilities (MS) and leaderboard enjoyment (variable Y).

Table 8 presents the full ordinal regression with all the covariates present.

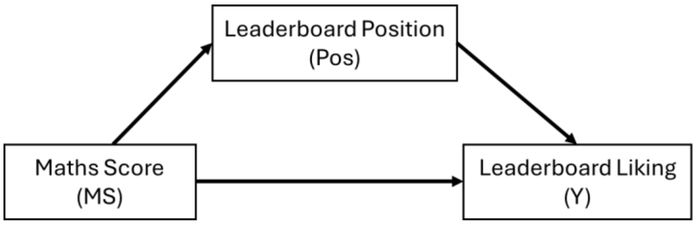

Figure 3. Mediator analysis: the variable Pos (leaderboard position) fully mediated the effect of maths scores (variable MS) on the liking of the leaderboard (Y).

Table 7. The position on the leaderboard as a mediator between maths abilities (MS) and liking of the leaderboard (Y). Model 1 is a linear regression model, while models 2 and 3 are ordinal regression models.

	Model	Target	Predictors	β	Pr (> |z|)	AIC or R^2
1.	MS →Pos	Pos	MS	-0.5	**<0.0001**	24.1 (R^2)
2.	MS →Y	Y	MS	0.16	**<0.0001**	3010 (AIC)
3.	MS+Pos →Y	Y	MS	-0.004	0.75	2989 (AIC)
			Pos	0.35	**<0.0001**	

Table 8. Full ordinal regression model. Target variable: Y (enjoyment of the leaderboard). All the variables, except Gender and Class Year have been normalized using z-scores.

Predictor	B	Pr (> |z|)
Gender = M	**0.30**	**0.027**
PlayTime	0.04	0.63
Class Year	-0.014	0.73
MA	**-0.19**	**0.009**
Pos	**0.40**	**<0.0001**
LS (literacy score)	-0.09	0.28
Rate Human Game	**0.57**	**<0.0001**
Rate Solo Game	0.12	0.08
Rate CPU Game	**0.18**	**0.012**
Intercepts		
did not like | kind of	0.81	0.32
kind of | liked	3.19	<0.0001
liked | loved	5.46	<0.0001
Observations = 984, AIC = 1820		

The covariate 'video game liking' (LikeV) was excluded because it was highly correlated with the variable 'time spent playing video games habitually' (TimeP). TimeP was retained, as it demonstrated better predictive power than LikeV based on the model's AIC. Students' gender, MA, position in the leaderboard (Pos) and the rates given to the *vs. human* and *vs. CPU* game modes were considered significant factors. The most significant variables were the players' position in the leaderboard and the score given to the *vs. human* game mode — both of which exhibited a positive relationship with leaderboard enjoyment — as well as players' MA, which displayed a negative association with how much they liked the leaderboard. The AIC decreased significantly from 2980 in the model with only the two main predictors, Pos and MA, to 1820 in the updated model.

3.3 Research Question 1B: What level of accuracy can be achieved by a model predicting a player's opinion of the leaderboard?

Table 9 shows the performances of the three trained ML models, including their respective accuracy, recall, precision, and F1-score. The performance indicators of Table 9 were obtained after upsampling the minority classes using the SMOTE algorithm (as described in 2.2). Therefore, the four classes of the target variable had equal frequency: the dataset used to train and test the models was perfectly balanced and the levels of accuracy presented in Table 9 have to be compared with a baseline random chance of 25%. All three models had an accuracy significantly higher than random. However, more complex tree-based models such as XGBoost and Random Forest significantly outperformed decision tree models. The best-performing model was XGBoost, with an accuracy of 79.7%. Considering that this was a four-class classification and that the dataset used to train and test the models was perfectly balanced (by the SMOTE algorithm), this accuracy is considered to be significantly high.

Table 9. Performance (accuracy, precision, recall and F-1 score) of the Machine Learning models (Decision Tree, Random Forest and XGBoost) in a 4-level multiclass classification.

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	49.2	50.4	48.5	49.4
Random Forest	74.6	75.9	73.1	74.4
XGBoost	79.7	79.5	80.0	79.7

For XGBoost, the best-performing model, a confusion matrix is also presented in Table 10. Table 11 reports the precision and recall by each of the four classes. The quality of the predictions is not homogeneous across the four classes, since it is more accurate to predict when a user will not like the leaderboard than the opposite.

Table 10. Confusion matrix for the multiclass XGBoost model.

		Reference value			
		Did not like	**Kind of liked**	**Liked**	**Loved**
Predictions	**Did not like**	72	2	4	1
	Kind of liked	4	77	5	12
	Liked	4	7	68	8
	Loved	8	8	8	62

Table 11. Precision and recall for each of the four classes (XGBoost model)

	Did not like	Kind of Liked	Liked	Loved
Precision	91.4	78.5	78.2	72.1
Recall	81.8	81.9	80.0	74.7

3.4 Research Question 2: Is the level of enjoyment of the leaderboard associated with the scores players assigned to each game mode?

Normality tests were performed to examine the distribution of scores given to each game mode. Table 12 shows the result of the Shapiro-Wilk tests conducted for the four groups of leaderboard enjoyment. All tests yielded statistically significant results ($p \leq 0.05$), indicating that the distribution of scores for all game modes did not adhere to a normal distribution within any of the groups.

Table 12. Shapiro-Wilk test results for the distribution of game mode scores considering four groups (Do not like/Kind of like/Like/Love the leaderboard).

Group	Do not like	Kind of like	Like	Love
vs. human	W = 0.889 p = 0.0015	W = 0.7243 p < 0.0001	W = 0.6216 p < 0.0001	W = 0.523 p < 0.0001
vs. CPU	W = 0.9065 p = 0.0045	W = 0.9411 p < 0.0001	W = 0.9178 p < 0.0001	W = 0.8724 p < 0.0001
Solo	W = 0.9394 p = 0.0441	W = 0.9619 p < 0.0001	W = 0.9448 p < 0.0001	W = 0.9013 p < 0.0001

Table 13 shows the kurtosis for the four groups of leaderboard enjoyment. Positive values indicate distributions with heavier tails, while negative values indicate flatter distributions with lighter tails. There is a high variation in kurtosis, indicating differences in the spread of the scores within each group. The lower values for the *vs. CPU* and the even lower values for the *solo* game mode indicate that there was a more even distribution of scores, so there was less variation in the scores students gave to these game modes. The high kurtosis values for the scores of the *vs. human* game mode, however, suggests a concentration of high scores with few moderate ratings: this indicates that students who kind of liked, liked and loved the leaderboard mainly assigned high scores to this game mode.

Table 13. Kurtosis results for the distribution of game mode scores considering 4 groups (Did not like/Kind of like/Like/Love the leaderboard).

Group	Did not like	Kind of like	Like	Love
vs. human	-0.5505	4.316	8.785	11.28
vs. CPU	-1.298	0.007147	0.7182	1.007
Solo	-0.8801	-0.5014	0.3244	-0.1737

Table 14 displays the skewness — how asymmetric the data distribution is, with 0 being a perfectly symmetrical distribution — for the four groups of leaderboard enjoyment. The negative values across all groups indicate that most players assigned higher scores to all three game modes, although there is variation across game modes and levels of leaderboard enjoyment. Of note, players who kind of liked, liked and loved the leaderboard assigned higher scores to the *vs. human* game mode.

Table 14. Skewness results for the distribution of game mode scores considering 4 groups (Do not like/Kind of like/Like/Love the leaderboard).

Group	Do not like	Kind of like	Like	Love
vs. human	–0.7041	-1.978	-2.671	-3.101
vs. CPU	-0.2745	-0.6498	-0.9072	-1.086
Solo	-0.3305	-0.2868	-0.6498	-0.7152

Considering the overall results of the normality tests, non-parametric statistical tests were chosen to analyse this data.

In order to answer RQ2 (i.e. whether or not liking the leaderboard is correlated to the overall rating given to the game modes), a Kruskall-Wallis test was performed to compare the mean ranks of the scores given to the three game modes across the four leaderboard enjoyment groups (Table 15 and Figure 5). The significant p-values across all game modes indicate that students' enjoyment of the leaderboard potentially affected the ratings of each game mode. The H statistic, displayed in the first column, indicates the degree of difference between the leaderboard enjoyment groups. The larger H value found for the *vs. human* game mode suggests a more pronounced variation in the scores among the four leaderboard enjoyment groups, indicating that students' feelings about the leaderboard have a stronger impact on how this game mode is rated. Additionally, the scores given to all game modes were significantly higher in the 'love the leaderboard' group.

When analysing game modes separately, the players who loved the leaderboard gave average scores of 9.49, 7.85 and 7.37 to the *vs. human, vs. CPU* and the *solo* game modes, respectively. For the 'like the leaderboard' group, the average scores were 9.19, 7.34 and 6.81. For students who only 'kind of liked' the feature, average scores were of 8.74, 6.80 and 6.22. Finally, the 'do not like the leaderboard' group assigned average scores of 6.81, 6.05 and 6.22 to the three game modes. If the scores of the three game modes are aggregated, players who loved the leaderboard gave an overall average rating of 7.37 (SD = 2.37); players who liked the leaderboard, 6.81 (SD = 2.17); players who kind of liked the leaderboard, 6.22 (SD = 2.42), the same average score of players who did not like the leaderboard (SD = 2.89).

Table 15. Kruskall-Wallis results for the distribution of game mode scores considering four groups (Do not like/Kind of like/Like/Love the leaderboard). The average ratings given to each game mode are also shown.

Group	H	Average score
vs. Human	$H = 82.59$; $p < 0.0001$	9.493 (Love); 9.192 (Like); 8.736 (Kind of Like); 6.811 (Do not like)
vs. CPU	$H = 48.11$; $p < 0.0001$	7.849 (Love); 7.344 (Like); 6.802 (Kind of Like); 6.054 (Do not like)
Solo	$H = 42.42$; $p < 0.0001$	7.374 (Love); 6.806 (Like); 6.217 (Kind of Like); 6.216 (Do not like)

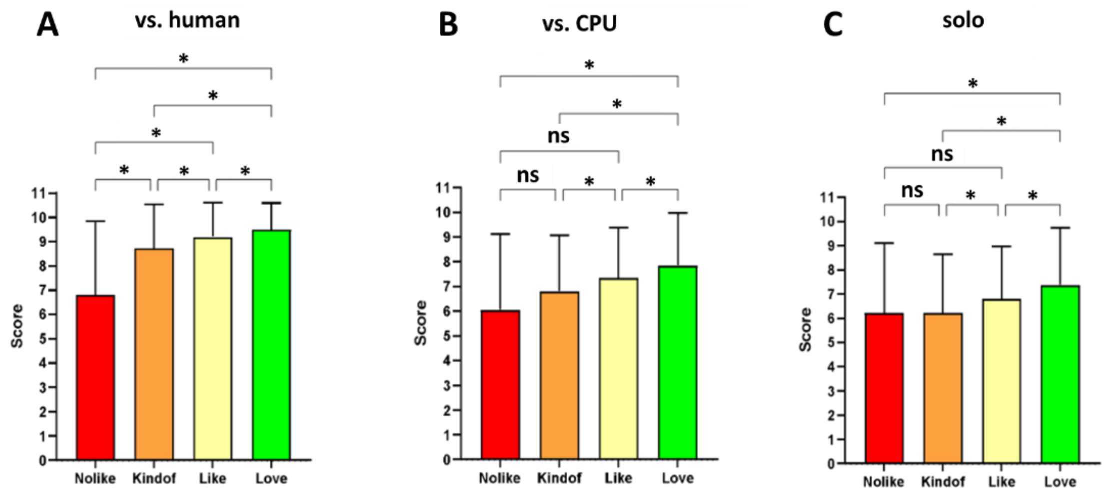

Figure 5. Kruskall-Wallis results for the A: *vs. human*; B: *vs. CPU*; C: *solo* game modes. Asterisks indicate significant differences.

In addition to the Kruskall-Wallis results in Figure 5, the spread of scores given to the different game modes across different levels of leaderboard enjoyment is also displayed through the violin plots in Figure 6.

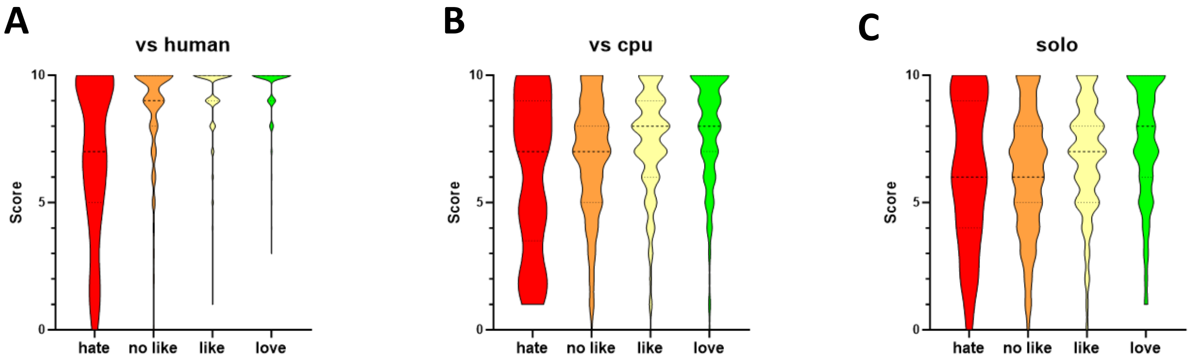

Figure 6. Violin plots displaying the data distribution of scores assigned to the A: *vs human;* B: *vs. CPU;* C. *solo* game modes.

In order to investigate possible correlations between leaderboard and game mode enjoyment, a Spearman's rank correlation was performed for each game mode, considering the 'Love', 'Like', 'Kind of like' and 'Do not like' groups as categorical variables (Table 14).

Table 16. Spearman's rank correlation results for the game mode scores considering four groups (Do not like/Kind of like/Like/Love the leaderboard). The sum of squared differences (S), p-value and Spearman's rho (ρ) are reported.

Group	**S**	**p-value**	**Ρ**
vs. Human	249914422	$p < 2.2e-16$	0.2359358
vs. CPU	263037204	$p = 2.759e-12$	0.1958155
Solo	267242871	$p = 6.921e-11$	0.1829575

The p-values were low and statistically significant (<0.05), indicating a strong likelihood that the relationship between liking or disliking the leaderboard and the scores given to the game across all three game modes is not due to random chance. The Spearman's rho values (ρ) were positive (but low) for all game modes, suggesting a positive and statistically significant, but weak correlation between liking the leaderboard and the score given to the game modes. A weak positive correlation was observed between liking the leaderboard and assigning higher ratings to the *vs. human* game mode. This relationship was even weaker for the *vs. CPU* game mode, and weakest for the *solo* game mode. This indicates that while the leaderboard might play a role in the player's overall enjoyment of the game, it is not a crucial aspect of their general experience, as some players enjoyed the leaderboard but not the game overall. In contrast, some players disliked the leaderboard feature, whilst rating the game modes positively.

3.5 Research Question 3: What aspects of the leaderboard lead to higher or lower enjoyment of this feature?

Player feedback regarding the leaderboard (gathered in the sixth and final session of the Happy Maths programme) offered additional insights. There were a total of 33 comments about the leaderboard (23 positive and 10 negative), and 19 comments about the competitive aspect of the digital maths game (16 positive and 3 negative). Some of the comments that exemplify these experiences include: "The game is very fun because I can practise to be on the top of the leaderboard"; "I like playing against my friends and I like the leaderboard"; "I love that I can see how I am improving on the leaderboard"; "I liked winning and being on the leaderboard"; "I liked the leaderboard and the points; you can score really high". Among the negative comments regarding the leaderboard and the competitive aspect of Seven Spells, are "I didn't like that the leaderboard compares how good you are at maths"; "The leaderboard can be embarrassing if you are last"; "I don't like the leaderboard"; "I hate the way people always brag that they are on the top of the leaderboard". Figure 7 shows the word cloud illustrating the most frequent words used in players' comments.

Figure 7. Wordcloud illustrating the most frequent words used by players when commenting about the leaderboard

Figure 8's bar plot displays the percentage of the most commonly used words by players when discussing the leaderboard. Following a stemming process, words are displayed in their root forms, excluding prefixes and suffixes.

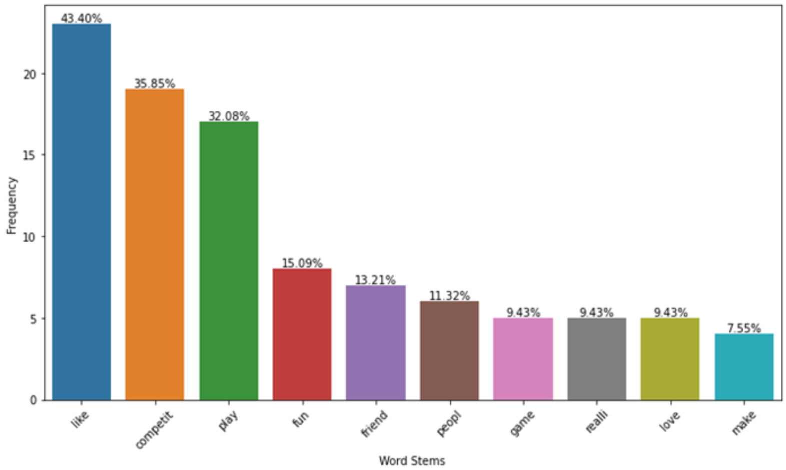

Figure 8. Percentages of players' most used word roots when commenting about the Seven Spells leaderboard.

4. Discussion

In this study, we sought to identify which player traits and demographic information can predict player's experience with an infinite leaderboard in a digital maths game (RQ1A) and the accuracy to which that experience can be predicted using different machine learning models (RQ1B). We also evaluated the relationship between the players' enjoyment of the leaderboard and their enjoyment of the three different game modes of the digital maths game 'Seven Spells' (RQ2). Finally, we analysed qualitative, anecdotal evidence from players' comments regarding the leaderboard with the aim of identifying possible aspects of the leaderboard which could be connected to its enjoyment or lack thereof (RQ3).

Regarding the first research question, the most significant factors predicting whether or not players liked the leaderboard were their position on the leaderboard, the score given to the *vs. human* game mode (both positively associated) and their MA score (negatively associated). Our hypotheses regarding the role of maths anxiety and the leaderboard position were respected.

As expected, a player's position on the leaderboard was a strong predictor of their enjoyment of this feature. Indeed, it is reasonable to expect that players enjoyed seeing their names at the top of the leaderboard, whilst contrarily did not enjoy displaying their poor performance to their classmates. The findings are in line with previous research suggesting how a player's position on the leaderboard is directly linked to how much they feel motivated by this game mechanic [32, 33]. The importance of leaderboard performance in defining how much players' liked this feature was reflected in their comments. Players who enjoyed the leaderboard and the competitive aspects of the digital maths game frequently mentioned winning, achieving high scores and appearing at the top of the leaderboard. Conversely, some players reported a negative experience with the leaderboard feature and competitiveness, mentioning that the leaderboard was embarrassing for those who rank lower and that well-performing players 'brag' about their position on the leaderboard.

Excluding the position on the leaderboard, the results indicated that factors unrelated to game performance are also important in predicting a player's experience with the leaderboard feature. MA was a significant factor even in the presence of the other covariates. This suggests that anxious students may react negatively to game elements that highlight competitiveness and social comparison, such as the leaderboard feature, even if their scores, mathematics knowledge, and literacy levels are comparable to those of non-anxious players. The accentuation of social comparison by the leaderboard can trigger the negative effects of MA, which could lead not only to a negative experience with the leaderboard feature, but more than likely, to a detrimental experience with the entire digital game, as suggested by the association between liking the leaderboard and the average ratings assigned to the different game modes. A well-known negative effect of MA is the avoidance of maths tasks [22, 18], which can, in a DGBL approach, lead to less engaged players, thereby undermining the potential of the game as a learning tool. Indeed, leaderboards following an infinite/absolute design could possibly trigger those negative effects of MA that game-based interventions are trying to mitigate. Our results support previous studies where social comparison was shown to be a significant predictor of MA [56] and where MA was discovered to be related not only to the manipulation of numbers, but also to the social experience of doing mathematics tasks in the classroom in front of peers [18]. This is in line with [57], who suggested that a number of in-class behaviours (e.g. embarrassment in front of their classmates) are related to a student's maths anxiety.

Among the covariates, the rating assigned to the *vs. human* game mode, *vs. CPU* game mode and gender were significant factors.

The rating players assigned to the *vs. human* game mode was the strongest predictor. The players that liked the leaderboard tended to also be the players that liked to play the *vs. human* game mode. This may indicate that the enjoyment of the *vs. human* game mode, marked by direct competition between players, might have acted as a proxy to how much students liked playing competitively, which could be indirectly related to the infinite leaderboard, in which players' scores are openly compared through an ordered ranking.

Gender was a significant factor, even if only at the 0.05 level (indicating 95% confidence in its influence) but not at the 0.01 level (which requires 99% confidence). This result is in line with most published studies, suggesting how boys tend to like competition more than girls [58, 59]. However, it is interesting to stress that this predictor did not have a large effect, a result that aligns with a subset of studies highlighting that girls are as competitive as boys at school in physical and playful activities [60, 61].

Neither the mathematics nor the literacy scores for the Irish national standardised tests were significant predictors of players' experience with the leaderboard when considering the other variables. Indeed, a main effect of math abilities was observed (Table 6, Model 2), but it was fully mediated by leaderboard position, suggesting that math abilities influenced leaderboard enjoyment only when they translated into game performance.

Students' gaming habits, specifically how much students liked video games and how much they played, were also not significant factors in predicting whether they liked the leaderboard when accounting for the influence of other variables. Game habits may not be good predictors of leaderboard enjoyment since students may have different motivations for enjoying and playing video games in general, which might not include competitiveness. Players may prefer specific video games that do not involve competition, or enjoy other aspects of playing, such as relaxing, experiencing a game's story or exploring the game world, instead of being motivated to play due to the presence of competition elements, such as leaderboards. Gaming habits might also not be relevant due to the context of the study: while students may enjoy playing video games, the experience of engaging with a serious game for practising maths in the classroom may be different than playing video games for pure entertainment.

Class year was not a significant factor in liking the leaderboard. Although MA and awareness of social comparison increase with age [43], it may be that these effects are already included in other factors (such as the MA score itself), thus diminishing the importance of age in predicting the leaderboard experience.

Overall, our results reinforce the need for educators and game designers to consider how the introduction of competitive elements might affect anxious and poorly performing students. While features such as leaderboards are generally well-liked and have the potential to motivate students, increasing their engagement with the game [32, 33, 35], they may be detrimental to the experience of more anxious students. Additionally, as shown by the importance of the player's position in predicting how much the leaderboard is liked, this study further confirms that a leaderboard may discourage students who are not performing well, thus decreasing their engagement with the serious game [32]. Consequently, the use of an infinite leaderboard in a game-based approach depends on the characteristics of the target audience; classes with a higher prevalence of anxiety may benefit from the use of a ‘relative leaderboard’ or from the removal of the leaderboard feature from the game. Relative — or ‘no-disincentive’ — leaderboards limit the information that players can access; the players can only view the positions of those players directly above and below them, rather than the positions of all players [42]. This may avoid negative feelings related to leaderboard position.

Regarding RQ1B, the machine learning models showed that liking the leaderboard could be predicted with an accuracy of 79.7%. This result is regarded as very significant, since the predictive problem was a 4-class classification task, and the accuracy was obtained with a fully balanced dataset. In the future, these models could be used to automatically make the game adapt itself according to the player’s profile and scores. In the context of the leaderboard, a model like this could make a digital educational game include, remove or modify the leaderboard feature (e.g. changing its design between infinite and relative) to improve player experience depending on the audience’s characteristics.

Concerning RQ2, our analysis showed a pattern that liking/disliking the leaderboard was significantly associated with the score given to the three game modes of the Seven Spells game. These correlations, however, were weak for all game modes: the strongest correlation was observed in the *vs. human* game mode ($\rho = 0.205$), followed by the *vs. CPU* game mode ($\rho = 0.149$), and finally, the *solo* game mode (0.138). While this finding may imply a potential impact of the leaderboard on the players’ game experience, the weak correlations indicate that other factors might play more prominent roles in determining how students evaluate the different game modes. Still, as one of many factors that may somehow influence students’ experience, educational game designers should carefully consider when and to what extent an infinite leaderboard should be present in an educational context, as it might negatively influence players’ engagement and enjoyment of the game.

Finally, the analysis of players’ comments (RQ3) indicated that its competitive aspect is potentially the main driver behind players’ preferences regarding this feature. Future work examining a greater number of comments and different leaderboard designs may shed further light on the reasons for players’ enjoyment of distinct types of leaderboards.

The current work, however, has some relevant limitations. Firstly, this study only considers students’ experiences with an infinite leaderboard. While players’ enjoyment of the infinite leaderboard was influenced by mathematics anxiety and leaderboard position, this may not be the case for the relative leaderboard. Since this study’s design includes a single ‘infinite leaderboard’ condition, a data-driven comparison is not possible in this paper. Another issue lies in the feedback survey, which asked students about their enjoyment of the infinite leaderboard feature. This was done through a 4-item Likert scale: ‘I did not like the leaderboard’; ‘I kind of liked the leaderboard’; ‘I liked the leaderboard’; ‘I loved the leaderboard’. The ‘I kind of liked the leaderboard’ answer is a suggestive item with a slightly positive sentiment, and thus may have reduced the validity of the ‘leaderboard enjoyment’ construct. However, it did not undermine our analysis, since the responses were treated as a 4-level ordinal variable.

5. Conclusions

This paper examined factors that predict player experience with an infinite leaderboard in a digital maths educational game. Whilst leaderboards are one of the most common elements in game design, their effect is debated, with some studies showing that leaderboards can motivate players and other studies highlighting their negative impact on both the player’s enjoyment of the game and their educational experience by increasing peer pressure [32, 33, 34, 35, 36, 63].

The present results suggest how player experience with the infinite leaderboard is not only driven by their game performance and leaderboard position but also by other individual traits, in line with previous research on leaderboards and DGBL [42]. Maths-anxious students were sensitive to the presence of a leaderboard and the social competition dynamic introduced by it. MA was one of the main factors that predicted a player's negative experience with the leaderboard, even after controlling for the player's position on the leaderboard, gender, age, and numerical abilities. Even if the results were weighted by the minority of players who reported a negative or more moderate opinion regarding the infinite leaderboard, this minority represented a non-negligible 19.88% of the participant cohort. Thus, while the leaderboard may usually be enjoyed by players, students who do not enjoy it may have a more negative experience with the DGBL approach.

This study exemplifies the impact of specific individual player traits on their experience with the infinite leaderboard feature in a digital mathematics card game. The use of DGBL approaches and specific game features may bring opportunities to engage students, but challenges may also arise depending on the target audience. Consequently, considering game features and students' characteristics before implementing educational games is an important step that should not be overlooked.

Finally, we showed how liking the leaderboard can be predicted with an accuracy of about 80% by tree-based ML models. These models could be used in the future to create a more adaptive experience, removing or modifying the leaderboard feature to better match players' profiles and in-game performance.

1. Acknowledgments

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.

2. Conflicts of interest

The authors declare that they have no conflicts of interest relevant to the content of this research paper. No financial, personal, or professional interests that could be construed to have influenced the work have been identified by any of the authors.

3. References

[1] R. Van Eck, "SAPS and Digital Games: Improving Mathematics Transfer and Attitudes in Schools," in *Mathematics Education in the Digital Era*, vol. 4, Dordrecht, Netherlands: Springer, 2015, pp. 141–173. doi: doi.org/10.1007/978-94-017-9517-3_9.

[2] K. Becker, "What's the difference between gamification, serious games, educational games, and game-based learning?", *Academia Letters*, article 209, 2021. doi: doi.org/10.20935/AL209.

[3] A. Theofylaktos, G. Lampropoulos, and K. Siakas, "Digital Game-Based Learning and Serious Games in Education", *International Journal of Advances in Scientific Research and Engineering*, vol. 4, no. 12, pp. 139-144, Dec. 2018. doi: doi.org/10.31695/IJASRE.2018.33016.

[4] F. Laamarti, M. Eid, and A. El Saddik, "An overview of serious games", *International Journal of Computer Games Technology*, Oct. 2014, doi: doi.org/10.1155/2014/358152.

[5] R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer, *Serious games: Foundations, concepts and practice*, 1st ed., Switzerland: Springer Cham, 2016.

[6] N. Tahiri and M. E. Alami, "A new evaluation technique through serious games for children with ASD", *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 1, pp. 202–217, 2020, doi: doi.org/10.3991/ijet.v15i11.12843.

[7] S.A. Al Majali, "The Digital World for Children and its Relationship with Personality Disorders: Exploring Emerging Technologies", *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 1, pp. 213–221, Jan. 2020, doi: doi.org/10.3991/ijet.v15i01.11412.

[8] J. Byun and E. Joung, "Digital game-based learning for K–12 mathematics education: A meta-analysis", *Sch. Sci. Math.*, vol 118, nos. 3-4, pp. 113–126, Apr 2018.
[9] E.N. Castellar, A. All, L. Marez, and J.V. Looy, "Cognitive abilities, digital games and arithmetic performance enhancement: A study comparing the effects of a math game and paper exercises", *Comput. Educ.*, vol. 85, pp. 123–133, May 2015, doi: doi.org/10.1016/j.compedu.2014.12.021.
[10] S.P. Hwa, "Pedagogical change in mathematics learning: Harnessing the power of digital game-based learning". *J. Educ. Technol. Soc.*, vol. 21, no. 4, pp. 259–276, 2018.
[11] S. Vanbecelaere, K. Van den Berghe, F. Cornillie, D. Sasanguie, B. Reynvoet, and F. Depaepe, "The effects of two digital educational games on cognitive and non-cognitive math and reading outcomes", *Comput. Educ.*, vol. 143, p. 103680, 2020, doi: doi.org/10.1016/j.compedu.2019.103680.
[12] N. Barz, M. Benick, L. Dörrenbächer-Ulrich, and F. Perels, "The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis", *Rev. Educ. Res.*, vol. 94, no. 2, pp. 193–227, Apr. 2024, doi: doi.org/10.3102/00346543231167795.
[13] U. Träff, L. Olsson, R. Östergren, and K. Skagerlund, "Development of early domain-specific and domain-general cognitive precursors of high and low math achievers in grade 6", *Child Neuropsychol.*, vol. 26, no. 8, pp. 1065–1090, Nov. 2020, doi: doi.org/10.1080/09297049.2020.1739259.
[14] J. Watkins and E. Mazur, "Retaining students in science, technology, engineering, and mathematics (STEM) majors", *J. Coll. Sci. Teach.*, vol. 42, no. 5, pp. 36–41, May 2013.
[15] L. Sun, H. Ruokamo, P. Siklander, B. Li, and K. Devlin, "Primary school students' perceptions of scaffolding in digital game-based learning in mathematics," Learn., Cult. Soc. Interact., vol. 28, no. 4, p. 100457, Dec. 2021, doi: doi.org/10.1016/j.lcsi.2020.100457.
[16] M.C. Passolunghi, I.C. Mammarella, G. Altoe, "Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades," *Dev. Neuropsychol.*, vol. 33, no. 3, pp. 229–250, 2008, doi: doi.org/10.1080/87565640801982320
[17] M.M. Kim, R.L. Andrews, and D.L. Carr, "Traditional versus integrated preservice teacher education curriculum: A case study," *J. Teach. Educ.*, vol. 55, no. 4, pp. 341–356, Oct. 2004, doi: doi.org/10.1177/0022487104266778.
[18] S.S. Wu, M. Barth, H. Amin, V. Malcarne, and V. Menon, "Math anxiety in second and third graders and its relation to mathematics achievement". *Front. Psychol.*, vol. 3, p. 162, Mar. 2012, doi: doi.org/10.3389/fpsyg.2012.00162
[19] K. Morsanyi, B.M.C.W. van Bers, T. McCormack, J. and McGourty, "The prevalence of specific learning disorder in mathematics and comorbidity with other developmental disorders in primary school-age children". Br. J. Psychol., vol. 109, no. 4, pp. 917–940, Nov. 2018, doi: doi.org/10.1111/bjop.12322.
[20] L. Buratta, M. Piccirilli, G.A. Lanfaloni, S. Ilicini, C. Bedetti, S. Elisei, "Mathematics anxiety and cognitive performance in adolescent students," *Psychiatr. Danub.*, vol. 31, suppl. 3, pp. 479–485, 2019.
[21] E. Carey, F. Hill, A. Devine and Szucs, D., "The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children", *Front. Psychol.*, vol. 8, p. 11, Jan. 2017, doi: doi.org/10.3389/fpsyg.2017.00011.
[22] F. Hill, I.C. Mammarella, A. Devine, .S. Caviola, M.C. Passolunghi, and D. Szűcs., "Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity". *Learn. Individ. Differ.*, vol. 48, pp. 45–53, Oct. 2016, doi: doi.org/10.1016/j.lindif.2016.02.006.
[23] U. Tokac, E. Novak, and C. Thompson, "Effects of game-based learning on students' mathematics achievement: A meta-analysis". *J. Comput. Assist. Learn.*, vol. 35, no. 3, pp. 407–420, 2019, doi: doi.org/10.1111/jcal.12347.
[24] B.D. Homer, C. Raffaelle, and H. Henderson, "Games as playful learning: Implications of developmental theory for game-based learning," in *Handbook of game-based learning*, MIT Press, 2020, pp. 25–52.
[25] L. Wang, B. Chen, G. Hwang, J. Guan, and Y. Wang, "Effects of digital game-based stem education on students' learning achievement: a meta-analysis," *Int. J. STEM Educ.*, vol 9, no. 1, pp. 1–13, Mar. 2022, doi: doi.org/10.1186/s40594-022-00344-0.
[26] D. Gil-Doménech and J. Berbegal-Mirabent, "Stimulating students' engagement in mathematics courses in non-stem academic programmes: A game-based learning". *Innov. Educ. Teach. Int.*, vol. 56, no. 1, pp. 57–65, Jan. 2017. doi: doi.org/10.1080/14703297.2017.1330159.
[27] P.S. Moyer-Packenham, C.W. Lommatsch, K. Litster, J. Ashby, E.K. Bullock, A.L. Roxburgh, J.F. Shumway, E. Speed, B. Covington, C. Hartmann, J. Clarke-Midura, J. Skaria, A. Westenskow, B. Macdonald, J. Symanzik and K. Jordan. "How design features in digital math games support learning and mathematics connections". *Comput. Human Behav.*, vol. 91, pp. 316–332, Feb. 2019, doi: doi.org/10.1016/j.chb.2018.09.036.
[28] P. Dondio, V. Gusev, and M. Rocha: "Do games reduce maths anxiety? A meta-analysis", Comput. Educ., vol. 194, p. 104650, Feb. 2023, doi: doi.org/10.1016/j.compedu.2022.104650.

[29] J.L. Tan, D.H. Goh, R.P. Ang, and V.S. Huan, “Participatory evaluation of an educational game for social skills acquisition,” *Comput. Educ.*, vol. 64, pp. 70–80, Mar. 2013, doi: doi.org/10.1016/j.compedu.2013.01.006.
[30] A. Rubin-Vaughan, D. Pepler, S. Brown,. and W. Craig, "Quest for the golden rule: An effective social skills promotion and bullying prevention program". *Comput. Educ.* vol. 56, no. 1, pp. 166–175, Jan. 2011, doi: 10.1016/j.compedu.2010.08.009.
[31] E.A. Boyle, T.M. Connolly, T. Hainey, and Boyle, J. M., "Engagement in digital entertainment games: A systematic review", *Comput. Human Behav.*, vol. 28, no. 3, pp. 771–780, May 2012, doi: doi.org/10.1016/j.chb.2011.11.020.
[32] M. Ninaus, S. De Freitas, and K. Kiili, “Motivational potential of leaderboards in a team-based math game competition,” in *Proc. 9th Int. Conf. Games and Learning Alliance (GALA)*, Laval, France, Dec. 2020, pp. 242–252, doi: doi.org/10.1007/978-3-030-63464-3_23.
[33] S. Park, and S. Kim, “Leaderboard design principles to enhance learning and motivation in a gamified educational environment: Development study”, *JMIR Serious Games*, vol. 9, no. 2, p. e14746, Apr. 2021, doi: doi.org/10.2196/14746.
[34] M. Schlömmer, T. Spieß, and S. Schlögl, “Leaderboard Positions and Stress—Experimental Investigations into an Element of Gamification”, *Sustainability*, vol. 13, no. 12, 6608, 2021, doi: doi.org/10.3390/su13126608
[35] S. Balci, J.M. Secaur, and B.J. Morris, “Comparing the effectiveness of badges and leaderboards on academic performance and motivation of students in fully versus partially gamified online physics classes”, *Educ. Inf. Technol.*, vol. 27, no. 6, pp. 8669–8704, Oct. 2022, doi: doi.org/10.1007/s10639-022-10983-z.
[36] S.A.A. de la Porte, “Exploring the inner workings of the leaderboard for use in educational contexts”. M.S. thesis, Dept. of Inf. and Comput. Sci., Utrecht Univ., Utrecht, The Netherlands, 2022. [Online]. Available: https://studenttheses.uu.nl/handle/20.500.12932/42867.
[37] S. Nicholson, “Exploring gamification techniques for classroom management” presented at the *Games+ Learning+ Society 9*, Madison, WI, USA, pp. 21–27, Jun. 12 - 14, 2013.
[38] N. Do, T. Jin, L.N. Meredith, and R. N. Landers, “A longitudinal quasi-experiment of leaderboard effectiveness on learner behaviors and course performance”, *Learning and Individual Differences*, vol. 116, 102572, October 2024, doi: doi.org/10.1016/j.lindif.2024.102572.
[39] N. Shabihi, F. Taghiyareh, and M.H. Faridi, “The relationship between gender and game dynamics in e-learning environment: An empirical investigation” in *Proc. Eur. Conf. Game-Based Learn.*, Oct. 2019.
[40] M. Denden, A. Tlili, F. Essalmi, M. Jemni, N. Chen, D. Burgos, “Effects of gender and personality differences on students’ perceptions of game design elements in educational gamification”. *Int. J. Hum. Comput. Stud*, vol 154, p. 102674, 2021, doi: doi.org/10.1016/j.ijhcs.2021.102674
[41] C.E. Höllig, A. Tumasjan, and I.M. Welpe, “Individualizing gamified systems: The role of trait competitiveness and leaderboard design”, *J. Bus. Res.*, vol. 106, 2018. doi: doi.org/10.1016/j.jbusres.2018.10.046.
[42] S. Bai, K.F. Hew, M. Sailer, and C. Jia, “From top to bottom: How positions on different types of leaderboard may affect fully online student learning performance, intrinsic motivation, and course engagement”, *Comput. Educ.*, vol. 173, p. 104297, Dec. 2021, doi: doi.org/10.1016/j.compedu.2021.104297.
[43] A. Dowker and A.S., Looi, "Mathematics anxiety: What have we learned in 60 years?" *Front. Psychol*, vol. 7, p. 508, 2016, doi: doi.org/10.3389/fpsyg.2016.00508.
[44] “The Happy Maths Project” Happy Maths. www.happymaths.games (accessed Mar. 11, 2024).
[45] ERC Test Department, “Primary paper tests.” Educational Research Centre. Accessed Mar. 11, 2024. [Online.] Available: www.tests.erc.ie/primary-paper-tests.
[46] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, version 4.2.3, Vienna, Austria, 2023. [Online]. Available: https://www.R-project.org/
[47] D. Bzdok, N. Altman, and M. Krzywinski, “Statistics versus Machine Learning”, *Nat. Methods*, vol. 15, pp. 233 – 234, 2018, doi: doi.org/10.1038/nmeth.4642.
[48] S. Hilbert, S. Coors, E. Kraus, B. Bischl, A. Lindl, M. Frei, J. Wild, S. Krauss, D. Goretzko and C. Stachl, “Machine learning for the educational sciences”, *Rev. Educ.*, vol. 9, p. e3310, 2021. doi: https://doi.org/10.1002/rev3.3310.
[49] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, “SMOTE: Synthetic Minority Over-Sampling Technique”, *J. Artific. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: doi.org/10.48550/arXiv.1106.1813.
[50] E. Loper, and S. Bird. “NLTK: The Natural Language Toolkit”, *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, Jul. 2002, pp. 63-70, doi: doi.org/10.3115/1118108.1118117
[51] W. McKinney, “pandas: powerful Python data analysis toolkit”, 2011.
[52] A. Mueller, “WordCloud for python documentation”, 2020.

[53] Matplotlib Development Team, “Matplotlib: a 2D Graphics Environment”, 2020.
[54] M.L. Waskom, “seaborn: statistical data visualization”, *Journal of Open Source Software*, vol. 6 (60), 2021, pp. 3021, doi: doi.org/10.21105/joss.03021
[55] R.M. Baron., and D.A. Kenny. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *J. of Personality and Social Psychol.*, vol 51(6), Dec. 1986, pp. 1173-1182, doi: doi.org/10.1037//0022-3514.51.6.1173
[56] A. Erdoğan, Ş. Kesici, and I., Şahin, "Prediction of high school students’ mathematics anxiety by their achievement motivation and social comparison", *Elem. Educ. Online*, vol 10, no. 2, pp. 646-652, 2011.
[57] C.D. Jackson and R. J. Leffingwell, “The role of instructors in creating math anxiety in students from kindergarten through college,” *Math. Teach.*, vol. 92, no. 7, 583–586, 1999.
[58] K. Eriksson and P. Strimling, “Gender differences in competitiveness and fear of failure help explain why girls have lower life satisfaction than boys in gender equal countries”, *Front. Psychol.*, vol. 14, 2023, doi: doi.org/10.3389/fpsyg.2023.1131837
[59] M. Niederle, and L. Vesterlund, “Do Women Shy Away From Competition? Do Men Compete Too Much?”, *The Quart. J. of Econ.*, vol. 122, Issue 3, Aug. 2007, pp. 1067–1101, doi: doi.org/10.1162/qjec.122.3.1067
[60] A. Booth and P. Nolen, "Choosing to compete: How different are girls and boys?", *J. Econ. Behav. Organ.* vol. 81, no. 2, pp. 542–555, Aug. 2012, doi: https://doi.org/10.1016/j.jebo.2011.07.018.
[61] A. Dreber and V.E., Ranehill, E., "Outrunning the gender gap—boys and girls compete equally," *Exp. Econ.*, vol 14, no. 4, pp. 567–582, Dec. 2011, doi: 10.1007/s10683-011-9282-8.
[62] C.J. Ferguson and C.K. Olson, “Friends, fun, frustration and fantasy: Child motivations for video game play”, *Motiv. Emot.*, vol. 37, pp. 154-164, Mar. 2012, doi: doi.org/10.1007/s11031-012-9284-7
[63] K. Werbach and D. Hunter, *For the win: How game thinking can revolutionize your business*, 1st ed., Philadelphia, PA. USA: Wharton Digital Press, 2012.
[64] Almo, André, et al. "Seven Spells and Peer Tutoring: a Collaborative Mathematics Game Experience." *ECGBL 2022 16th European Conference on Game-Based Learning*. Academic Conferences and publishing limited, 2022.3