International Journal of Serious Games

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

Article

Latent Class Analysis of Gameplay Metrics from Youth Playing a Robotics Game

Lawrence M. Scheier[1], William Hansen[1], Jennifer Javornik[2], and Alex Stone[2]

[1]*Prevention Strategies, Greensboro, NC, USA;* [2]*Filament Games, Madison, WI, USA*
scheier@larsri.org; billhansen1949@gmail.com; jjavornik@filamentgames.com; stone@filamentgames.com

Keywords:

gameplay metrics
intensity
persistence
latent class analysis
automatic log files
robotics
youth

Received: April 2025
Accepted: July 2025
Published: July 2025
DOI: 10.17083/anb2q267

Abstract

Game metrics have become a staple component to understanding how players interact with various aspects of the game and whether they comprehend the game mechanics. Despite their intrinsic value, few studies have used this type of objective instrumentation to examine whether users display unique gameplay styles. We used latent class analysis with seven game metric indicators from a robotics game to ascertain whether there are distinct patterns of gameplay. We also validated gameplay styles using measures of persistence and intensity of play. Four gameplay styles were obtained including *Fully Engaged* (engaged multiple aspects of the game), *Engaged in Training* (drove the robot but did not prepare for matches or take tutorials), *Engaged in Building* (accomplished game objectives, met challenges, and took tutorial), and *Engaged in Driving* (only drove the robot). Persistence and gameplay intensity were both associated with class membership and the obtained classes differed in mean levels of these measures. This study is unique by using a person-centered approach with game metrics as opposed to lower resolution and less reliable self-reports from players. Findings are discussed in terms of ways game developers can utilize game metrics to improve robotics game design and enhance game mechanics.

1. Introduction

Robotics and Scientific Reasoning Skills

Digital game-based learning has made great strides with respect to demonstrating improved motivation, behavior change, and learning outcomes [1], [2]. These accomplishments have been extended to include digital games as a means of teaching STEM-related skills including scientific reasoning, logical, and analytic skills or what is termed higher-order thinking [3], [4]. Several meta-analyses have now documented the pedagogical benefits of teaching computational thinking and STEM preparation using digital game-based learning [5], [6]. Digital games provide unique educational opportunities; they can be designed to be self-paced, learner-centered, provide immediate feedback on performance, and are engaging and motivating to play [7], [8].

A key aspect of STEM-based digital games is that they simulate real-world complexity and can be used to teach complex concepts (i.e., Newton's laws of motion) and higher-order thinking skills that are required in scientific reasoning [4]. This can include generating, testing, and refining hypotheses, deduction and inference, and presentation skills that encourage teamwork and collaboration. The use of digital games in educational settings is in keeping with Vygotsky's cultural-historical approach [9], which suggests that meaning and learning is obtained through activity, or what is termed “enacted or situated learning” [10]. In this context, the user can actively frame questions, pursue different lines of reasoning, and test their ideas in a safe environment with quests, challenges, leaderboards, badges, and rewards indicating the player’s proficiency. This type of educational approach is in keeping with the National Research Council’s emphasis on learning science by doing [11]. This latter goal requires authentic experiences reflecting participatory, immersive learning [12] and is closely aligned with a constructivist perspective [13].

Robotic games are one area that has received increased attention for their use in learning STEM skills. In the current study, we focus on a robotics game, *Robot Champions*™ which combines the opportunity to gain scientific reasoning skills with principles of engineering and robotics construction in a competitive multi-player digital game environment. In the game, players design, build, and program their own robots to compete in various challenges and tournaments. By allowing players to experiment with different robot components and design strategies, the game introduces them to basic engineering principles and concepts of physics, mechanics, and technology. As players progress, they encounter increasingly complex challenges that require them to apply mathematical reasoning, coding skills, and a deeper understanding of how their robot designs will perform under different conditions. Additionally, *Robot Champions*™ fosters a sense of accomplishment and curiosity by providing immediate feedback through gameplay, such as winning challenges or improving robot performance. This can motivate players to further explore STEM topics and acquire real-world problem-solving skills. The game's community aspects, including collaboration and competition with other players, also help to build a supportive environment where learning and sharing knowledge are encouraged. The game design concepts behind *Robot Champions*™ are supported by both individual studies [14], [15] and meta-analyses [16], [17], [18] that show the benefits of using educational robotics digital games to teach complex concepts, cooperative learning, problem solving, and basic engineering principles that are essential components of STEM education. This is not only true of robotics games but also digital games more generally, which can support development of cognitive skills relevant to STEM [16], [19], [20].

2. Literature Review

2.1 Game Metrics

One means of assessing the staying power of a game relies on using gameplay metrics as an objective means to determine whether players enjoy the game and will persist in their playing behaviors. Gameplay metrics are derived from in-game measures of performance, the latter often called behavioral telemetry or instrumentation data [21], [22]. Data obtained from automatic log files track players’ moment-to-moment in-game behaviors and game states. They provide a means to assess different experiential facets of gameplay and provide a glimpse of the interior world of players and how they process and react to game states.

Automatic log files that capture player-game interaction data can reveal to game developers whether players stick with the game (a proxy for enjoyment or intrigue), whether the game mechanics (rules, goals, and logic) are attractive, and whether players successfully navigate the various game challenges or if they get stuck at a particular place. Gameplay metrics are a supplement to other means of quantifying user experience including usability and playtesting (i.e., using think-aloud techniques), surveying players to assess motivation, biometrics such as eye tracking or galvanic skin response, and direct or video observation of gameplay [23]. Taken together, all this information provides a useful base from which to make decisions about game quality, marketing potential, and whether the game requires modifications [24]. Games that are too difficult or too easy can be modified either during the build and development process or post-release based on gameplay metrics.

Typical gameplay metrics parallel the user’s experience during a game and can include number of keystrokes or button presses, duration of play, challenges attempted and made, rewards, level completion times, errors committed on a specific level, and time on task [25]. Other metrics can include churn rate (stopping play in a defined period) [26], duration of a single game session, lag between sessions, game progression, and retention or stickiness (the number of players who continue to engage in the game over a defined time). In some cases, game metrics can include spatial analysis with geographic information systems (GIS), owing to the ability of the game telemetry to track where players spend time in a virtual environment through analysis of their navigational paths [22]. Drachen [21] also suggests that game metrics can be "genre or game-specific" reflecting the activities needed to advance in the game. For instance, a first person shooter game can track which weapons are used and how many shots are taken, player deaths, and NPC kills, while a racing game can track which car is selected, track performance times, and where and when crashes occur. This can also be extended to include spatial analysis [27] and heatmap visualizations, which distinguish where players spend time and where certain events occur in a virtual game [28].

2.2 Characterizing Gameplay

With few exceptions, most of the research addressing game metrics has examined the relationships between two or more game metrics (e.g., keystrokes and duration of play) using correlational techniques, including some form of linear regression or data reduction using factor analysis. In these typical variable-centered approaches, the resulting statistical parameters (e.g., a beta weight or factor loading) describes the average behavior of the entire population. This overlooks the fact there may be distinct subgroups of youth that play video games based on unique styles of play. In game design terms, the collective nature of how players engage a game, and the way they move through the game addressing the game's challenges and mechanics is what has been termed "play-personas" [29]. From a statistical point of view, rather than obtaining a single parameter, each distinct subgroup may have its own set of parameters that describe how they engage the game and move through the various game challenges. For instance, one type of persona may involve skipping the tutorial and getting right to the game to determine whether it is intriguing and challenging. Conversely, a different gameplay strategy may involve using the tutorial to learn about all the different possibilities that can occur in the game and better understand the game's logic and mechanics. These differences, which are important from a game design point of view, are not often examined in detail. The unique playstyles of different "personas" can be revealed using game metrics, which represent the interface of game mechanics and how players interact with and engage the game.

Mixture modeling presents an alternative approach to quantify unique styles of gameplay or what is termed "subgroup heterogeneity." One type of mixture model is Latent Class Analysis (LCA) [30], which provides a means to detect qualitatively unique classes of behavior. The classes are homogeneous with respect to a set of behaviors creating mixtures that are uniquely composed within the larger population of events. The goal of LCA is to derive the most parsimonious set of mixtures that can explain the different patterns of gameplay. Membership in each class is based on estimated posterior probabilities using the joint marginal distributions.

To date, several studies of gameplay behaviors have applied mixture modeling approaches like LCA; however, these studies have not focused on objective game metrics per se, but rather utilized players' self-reported game involvement. For instance, Faulkner et al. [31] examined game playing classes of behavior in a relatively large sample of Canadian high school youth. Likewise, Colder-Carras and Kardefelt-Winther [32] used LCA to study gaming-related problems in a large sample of European youth. Chang et al. [33] examined subtypes of internet gaming disorder in Chinese youth during COVID-19 and Bowman and Chang [34] examined subtypes of gamers in Taiwanese youth based on motivations for playing games (e.g., competing, teaming, role playing creating, entertainment). Myrseth and Notelaers [35] examined gaming disorder in a sample of youth taken from the Norwegian National Registry, and Siste et al. [36] examined problematic gaming behaviors in a sample of Indonesian youth. In all of these examples, the authors were able to extract multiple subtypes of behaviors that could distinguish players' self-reported gameplay activities and their motivations. Lacking from these studies, however, is a true and accurate accounting of their progress through the game that can be used to qualify whether there are objectively unique patterns of gameplay.

2.3 Gameplay and Persistence

One concern that we address is whether a player's gameplay style is related to persistence. Persistence generally refers to a willingness on the part of the player to continue with gameplay both within sessions and across time. Persistence is an important metric because it shows that players find the game intriguing and are willing to stick with playing in a determined manner. In psychological terms, persistence refers to the willingness to continue trying to solve a task despite difficulty [37]. Psychological theories of motivation consider persistence vital to action and the direction of behavior [38] and a major component of self-regulation and self-control [39]. Persistence is considered important developmentally [40], valued as a form of temperament [41], and tied to a person's self-efficacy [42]. In terms of gameplay, persistence is used as a proxy for player engagement and has been used to indicate whether players are learning [43]. To achieve the goals or objectives of a serious game a player must demonstrate some level of persistence including mastery of game skills [43].

In an LCA framework, once subgroups (classes) are defined, validation requires some type of external marker that can be used to characterize the different elements of gameplay. A handful of studies have tackled the issue of assessing persistence in terms of gameplay metrics. Ventura et al. [44] developed a measure of gameplay persistence based on the amount of time a player spent on solving physics problems and the number of restarts on the problem (all calibrated by the problem's degree of difficulty). The authors also correlated this score with self-reported persistence and an online computerized task to assess persistence. DiCerbo [45] assessed persistence in a cohort of children playing the virtual game *Poptropica*™ using log files that contained completion (e.g., number of quest events completed) and time indicators (e.g., time on task). She modeled persistence as a latent factor with six indicators (three separate quests), produced a well-fitting model, a reliable scale score (α =.87), and showed an age-graded effect.

2.4 The Present Study

Despite efforts to use classification techniques related to gameplay, to our knowledge no study has used mixture modeling approaches applied to in-game metrics. This represents a gap in the literature because there are most likely tremendous differences with how players approach a game (e.g., whether they skip the tutorial or engage certain components of the game) as well as differences in player-game interactions. All this information can be useful to game developers to determine whether a game is captivating to the target audience and determine precisely what features of the game are most attractive. Additional value comes from the ability to use in-game metrics to obtain greater insight into and quantify learning mechanics [23].

To better understand the experiential side of how youth play this game, we apply mixture modeling using LCA to game metric data obtained from youth playing the online video game *Robot Champions*™. Self-reports are subject to social desirability [46] and even in very advanced user-testing procedures it is extremely difficult for a player to recall precisely what happened and what they liked or disliked at a particular juncture of the game. The advantage of using in-game metrics rather than self-reported gameplay is the objective nature of these measures as they capture moment-to-moment in situ gameplay free from any bias associated with self-reports.

In addition to using mixture modeling to extract unique classes of youth playing the game, we conduct a structural validation by examining the association of persistence and intensity with class membership. Both measures provide insight into the length of time a player will engage in playing the video game, reflecting how hard they will work at the task, and the pace or intensity of their play, the latter factoring in how many events a player engages with respect to a defined time frame. In addition to using these measures to characterize class membership, we also examine class-specific means for persistence and intensity corresponding to the different gameplay styles. For both persistence and intensity, higher scores indicate the player's gameplay style is more engaged.

To summarize, very few studies have used game metrics to derive unique "styles of play" rather the focus has been on unique styles of players. The shift in focus is important from a design point of view as styles of gameplay can be indicative of functional aspects of the game and reveal what game elements are working that is attractive to players. The ability to discern unique styles of gameplay is an advantage of mixture modeling approaches, which look to identify unobserved patterns in how players approach a game. Given the exploratory nature of the study, we hypothesize the following: (1) the LCA will reveal definable play personas reflected in unique patterns of gameplay. Related to this, class structure cannot be specified *a priori;* however, we expect no less than 3 and possibly 4 or more unique styles of gameplay; (2) the distribution for persistence will be positively skewed. As noted, there is relatively little research on persistence in gameplay. We anticipate that most players will lack persistence both across time and within gameplay episodes. This rests on an understanding that most players who log into the Roblox platform to play Robot Champions™ will be experiencing the game *de novo.* While tutorials included in the game may serve to orient players generally, many players will nonetheless experience a steep learning curve that may contribute to frustration.

3. Methods and Material

3.1 Game Setting

Robot Champions™ is a serious educational game designed to introduce young people (ages 13 to 17) to the core concepts related to robotics with an eye cast toward helping them develop an interest in and skills related to STEM. Players manipulate and construct robots from a menu of classic robot components (e.g., gears, wheels, body armor, armature) and learn about basic engineering principles, material science, and physics including balance, mass, acceleration, gearing and pulleys, force, and other facets of Newton's laws of motion that can be learned from robot construction. Players compete individually or as a team against other players in timed competitions. The game is hosted on *Roblox* (https://www.roblox.com/), an online sandbox game platform and game creation system that allows users to program and play games created by themselves or other users.

Robot Champions™ includes five challenges:

- *Boarding Party* is an intermediate level cooperative game in which players must construct a ramp that goes up a mountain using blocks and ramps. Players succeed when they collaborate and communicate to make it out of a steep valley.
- *Bot Ball* is a competitive game designed for beginners. The game is played by building robots that can either kick or block a giant soccer ball with the objective to put the ball in their opponent's goal.
- *Parmesan Pipeline* is a competitive game designed for players who have intermediate robot design skills. This game is inspired by industrial realities. Players pack and ship wheels of cheese in a warehouse. To succeed, players need to build a robot that can precisely manipulate objects and coordinate the cheeses on a conveyor belt.
- *Skatepark Scramble* is a competitive game designed for beginning robot builders. Two teams build robots that compete to collect special orbs as they appear on a virtual skatepark map. Robots that are fast and agile have a greater chance of success.
- *Yeet the Sheep* is a cooperative game also designed for beginning robot developers. In this game, players must sort sheep by creating a robot that can scoop or grab a sheep and toss it into the appropriate virtual holding pen.

3.2 Dataset Creation

We obtained log files for 59 contiguous days of game play from February and March 2024. Data were downloaded from Mixpanel software in CSV format. Mixpanel is a time series data warehouse that ingests game-based analytics events and allows them to be efficiently queried. All data included a player's unique ID number, the date and time of execution, and the event that transpired. Thus, if a player logged into a game more than once, each time a game started, the game was uniquely identified along with any corresponding events. Not available from Mixpanel were demographics including players' ages, genders, or countries of origin.

The *Roblox* platform's server records a variety of event-based actions. Each action (e.g., driving a robot, entering a queue to compete in a match, taking a tutorial) is recorded as an event. Table 1 shows the variables available from the gameplay metrics as part of the server log files. When a player starts a practice session, queues for or finishes a match, the specific challenge is identified by an event trigger and is timestamped recording the date and time when it occurred.

Table 1. Mixpanel lexicon and variables used to formulate analysis of game metrics

Mixpanel Name	Entity Name	Entity Description
Game.Joined		User joined the game server
Game.CompletedObjective	Complete an Objective	The user completed an objective
Session.TutorialStarted	Start a Tutorial	A tutorial was started
Session.TutorialSkipped	End a Tutorial	User skipped or ended a tutorial
Game.EnteredDriveMode	Enter Drive Mode	User started driving their robot
Challenges.QueuedForMatch	Queue for a Match	Queued for a challenge
Challenges.StartedMatch	Start a Match	User successfully started a match
Challenges.StartedPractice	Queue for Practice	User started practice mode
Game.Leaving		User left the game server

Game.Joined and Game.Leaving are both used as "gating" measures. An episode of gameplay was required to have no fewer than three steps to be included in the analysis.

3.3 Assessing Gameplay Metrics

Using players' ID numbers, start and end times as well as each recorded event, we created several gameplay metrics. For each player, we tabulated the number of days in which a player logged in (*Days*), and the total number of games played (*Games*). Using the event data, we calculated the number of events logged for each game played (*Steps*). The duration of play for each login (*Duration*; end time minus start time) for each game. *Intensity* or pace of gameplay was calculated as the number of *Steps* completed during the time of play divided by *Duration*.

3.4 Latent Class Analysis Methods

LCA is a model-based technique and considered a categorical analogue to factor analysis [30], [47]. It works under the premise that several qualitatively distinct modes of gameplay can be detected in the larger population based on game indicators. These unique patterns of gameplay are considered latent classes or distinct mixtures that represent unobserved heterogeneity in the population (i.e., unique play personas). In the current context, we sought to identify patterns of gameplay using in-game event data (see Table 1).

In an LCA framework, parsimony is desired and the statistical fit of a model with k classes is evaluated against a model with $k+1$ classes using the sample data. Fit can improve, for instance, by the extraction of additional classes that can account for meaningful patterns of gameplay in the data. Different model fit indices are used to evaluate overall model fit including the Akaike Information Criterion (AIC) [48], Bayesian Information Criterion (BIC) [49], Entropy [50], and the Log-likelihood statistical fit index (LL). In the case of information criteria, smaller values are considered better. One indication of superior fit of one model over another is that as more classes are extracted there should be a modicum of shrinkage in the information criteria. The LL statistic reflects the likelihood of observing the empirical data given the set of parameter estimates (the logarithm of the LL is used so that higher values closer to 0 indicate better fit). Entropy is a standardized measure that reflects the chaos or uncertainty of the model classification based on estimated posterior probabilities. Entropy ranges from 0 to 1, with values closer to 1 denoting better classification certainty. A nonsignificant p-value for the Lo-Mendell-Rubin (LMR) statistic for a k-class solution provides support for a neighboring k-1 class solution. The LMR is a likelihood-based test modified for a mixture model by using an adjusted asymptotic distribution [51]. Because no gold standard exists for determining fit, it is therefore crucial to consider whether the model makes sense in terms of theory, previous research, interpretability, parsimony, and class separation, the latter indicating the classes are clearly distinguished in a meaningful way [52].

Once class membership is established and the best fitting model chosen, the model can be conditioned on covariates, which helps to characterize class membership. The integration of a mixture model with a variable-centered approach utilizes multinomial logistic regression to demonstrate the association of latent class membership with specific covariates. Odds ratios indicate the likelihood of belonging to one mode of gameplay (i.e., class) compared to a reference (usually the largest) class. As a follow-up to conditioning the LCA model on covariates, we modeled the same covariates as “distal outcomes” [53]. This integration of person- and variable-centered approaches models class membership as observed multinominal subgroups and examines mean (intercept) differences (using pairwise comparisons between styles of play) in measures of persistence (i.e., *Games*, *Steps* and *Duration*) and *Intensity*.

The Mplus statistical software was used to model class structure [54]. Two classify-analyze procedures including the R3STEP procedure [55] and the BCH procedure [56] were used to model auxiliary measures and distal outcomes, respectively. In the R3STEP procedure, once a case is assigned to their respective class using modal posterior probabilities, it is “fixed” in this class and a multinominal class membership variable is created. This keeps the covariates “structurally independent” of the measurement model.

In other words, the heterogeneity of the class structure is based solely on the joint distribution of the class indicators and kept separate from the effects of the auxiliary measures. R3STEP is a stepwise procedure, first estimating an unconditional model to compute the item response probabilities for modal class assignment by producing a parameter representing the average classification error. Cases are then assigned to their most likely class based on the latent class posterior probability distribution. With the model measurement parameters (i.e., thresholds expressed as logits) in place, and accounting for measurement error in the class assignment process, the final model is then conditioned by the covariates, adjusted for uncertainty in classification.

The BCH procedure uses a “weighted multiple group analysis” where each group becomes one of the latent categories from the LCA model and the groups are known and modeled as a multinominal fixed variable. BCH weights (the inverse of the matrix of classification errors) represent the measurement or classification error based on posterior probabilities and treated as a regression coefficient in the model that arises because there is uncertainty in class assignment (i.e., error probabilities). This locks in cases to their respective gameplay style class and their behavior can be compared using traditional regression or other variable-centered methods (i.e., ANOVA or t-tests). For the auxiliary measures, multinomial logistic regression is combined with the LCA model to estimate the effects of covariates on class membership. For the distal outcomes, post-hoc pairwise comparisons between classes is used to statistically contrast intercepts for each style of gameplay. Significant differences are based on a critical z-ratio with the parameter divided by its respective standard error.

The sample available for analysis is extremely large and may be overpowered for the LCA models as fit indices are sensitive to sample size. When considering the Law of Large Numbers (i.e., sample parameters will converge to the population values with larger datasets), we decided to draw four simple random samples of 20,000 gameplay cases without replacement. Each of the four datasets is then used to support one of the different model testing procedures. Based on the principles of random sampling, we expected only trivial deviations between samples in terms of the obtained parameter estimates. A power analysis using Monte Carlo simulation indicated the sufficiency of this sample size with up to 8-classes and covariate adjustments [57]. This decision is supported by Monte Carlo simulations that vary model parameters (i.e., number of indicators, response probabilities, covariate effect sizes) and obtain better replication and model convergence with larger samples [58].

4. Results

4.1 Description of Gameplay Behaviors

The dataset initially included data obtained from 240,638 players who logged 1,806,841 events. Cases were deleted when leaving the game was not recorded, which most likely occurred when players closed the game app without logging out of the game. Cases were also deleted when fewer than three steps were recorded or when the duration of play lasted less than 5 seconds (to eliminate BOTS and those just perusing the website but not engaging). As a result of eliminating these cases, the final dataset included data from 153,658 players who played 199,055 game sessions and engaged in 1,708,207 events. No cases had any missing data, as their gameplay was logged automatically on the server.

When players logged in to *Robot Champions*™ they had the option to pursue one of the five challenges (*Boarding Party*, *Bot Ball*, *Parmesan Pipeline*, *Skatepark Scramble*, and *Yeet the Sheep*) or to build and test drive a robot without specifically determining to play a challenge. Most of the time, when they chose a challenge to pursue, they played *Bot Ball* (75.8%). Less frequently, they played *Yeet the Sheep* (9.7%), *Skatepark Scramble* (6.0%), *Parmesan Pipeline* (4.4%), and *Boarding Party* (4.1%).

The most frequently reported event was to enter drive mode, completed in 77.4% of games. In slightly more than half of the games played (53.2%), players completed a defined objective. In about a third of the games (30.6%), players started a tutorial. In 12.1% of games, the tutorial was then skipped. Players queued to practice a challenge in 13.3% of the games. They queued to enter a match in 6.9% of the games but only started a match 2.3% of the time.

4.2 Gameplay Performance Outcomes

Days of play were positively skewed ($\bar{X}$ = 1.67, SD = 1.99, Skew = 6.30) with most players (89.7%) playing on only one day. Of those who played more than one day, 7.5% played on two days and 1.6% played on three days. Two players played 27 days. The number of games played ranged from 1 to 152. Like *Days*, *Games* was highly positively skewed ($\bar{X}$ = 2.68, SD = 7.70, Skew = 12.61) with the vast majority (83.9%) playing only once. Fewer players played two games (11.1%), three games (2.7%), or four games (1.0%).

Steps, like *Days* and *Games* was positively skewed, although generally less so ($\bar{X}$ = 8.48, SD = 7.42, skew = 3.23). About half of games played (51.5%) included six or fewer steps. During only 12.4% of games played did players complete more than 15 steps. *Intensity* (*Steps*/*Duration*) reflected pace of play and ranged between 0 and 0.57 ($\bar{X}$ = 0.043, SD = 0.040, Skew = 2.30). Larger *Intensity* scores reflect players taking more in-game actions in a shorter amount of time. *Intensity* and *Games* had a small inverse relationship (r = -.124, p <001).

Figure 1a shows the results for game duration for the group as a whole. *Duration* ranged from less than a minute (6 seconds or greater per the inclusion criteria) to a maximum of 23 hours and, like previous analyses of performance outcomes was high skewed ($\bar{X}$ =441.68 seconds, SD = 1232.71; Skew = 43.16). Accordingly, as shown in Figure 1b we used a log transformation for *Duration* ($\bar{X}$ =3.62, SD = 1.09; Skew = 0.348), which considerably smoothed the distribution and reduced the skew. Player activity lasted less than four minutes in slightly more than half of games (55.9%). In only 12.1% of games did player activity last more than 15 minutes. *Steps* and *Duration* were correlated (r = .526, p <001); the number of steps completed was generally greater when more time was spent playing the game.

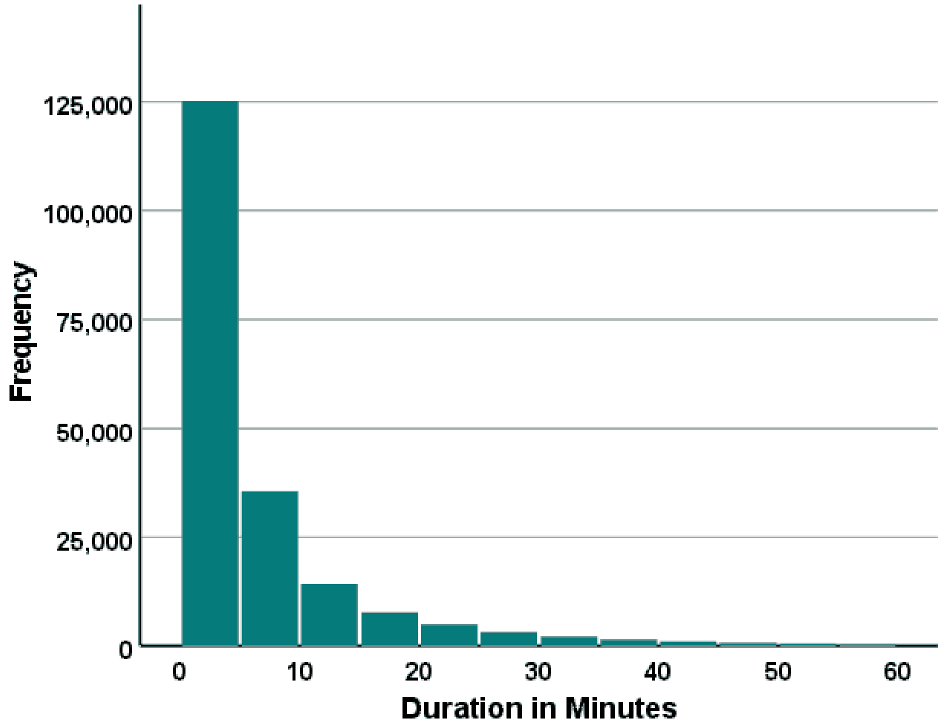

Figure 1a. Distribution of gameplay duration.

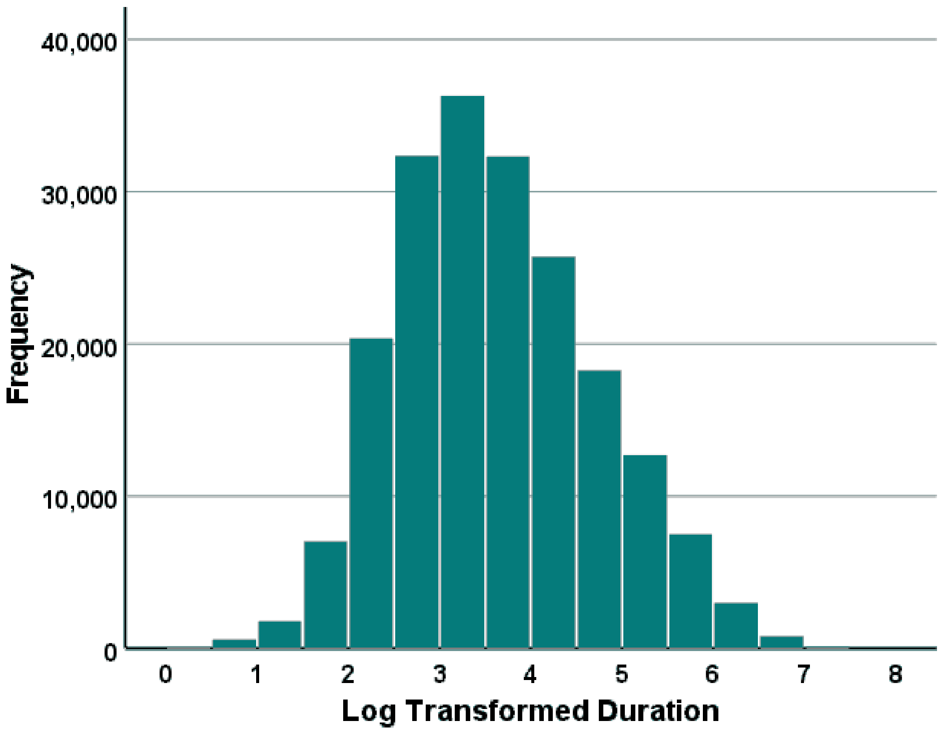

Figure 1b. Results of log transformation of duration.

4.3 Results of LCA Models

We tested multiple LCA models with anywhere from 2 to 8 classes. Table 2 shows the model fit indices obtained from these tests. Both the AIC and BIC shrunk appreciably with the addition of new classes, with the decrement slowing down between the 3- and 5-class models. Moving from the 3- to the 4-class model resulted in shrinkage of 2322 and 2259 information criterion points (AIC and BIC, respectively), while moving from the 4- to the 5-class model produced shrinkage of 1130 and 1067 points. Furthermore, increasing extraction of classes led to superior classification as demonstrated by the entropy statistic (except for the 5-class model). Other than the 2-class model, entropy was highest for the 3-class model (.982), followed by the 6-class (.978) and 4-class (.961) models. The LMR test was less informative as all the p-values were significant (indicating a model with *k-1* classes did not improve over a model with *k* classes).

Table 2. Model fit statistics for latent class analyses

Classes	LL (Deviance)	No. of free Parameters	AIC	BIC	Relative Entropy	A-LRT
2	-50,948	15	101,926	102,045	1.000	11,512***
3	-48,743	s23	97,532	97,713	0.982	4,355***
4	-47,574	31	95,209	95,454	0.961	2,309***
5	-47,001	39	94,079	94,388	0.943	1,132***
6	-46,592	47	93,279	93,650	0.978	806***
7	-46,299	55	92,708	93,143	0.920	579***
8	NA[1]	--	--	--	--	--

LL = Log-likelihood statistic; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LMR = Lo-Mendell-Rubin Adjusted LRT test for k vs k+1 classes. Relative Entropy is a summary measure of classification certainty once posterior class probabilities are obtained and can be computed for k > 1-class models. *** *p* <001
[1]NA = not estimable, best loglikelihood value not replicated, may result in a local maximum with a solution that may not be trustworthy.

Given the inconclusive support for a best-fitting model based solely on statistical criteria, we examined four additional elements of the models: (1) the latent class proportions based on the estimated posterior probabilities, (2) the degree of latent class separation, (3) the item response probabilities visualized with plots, and (4) the accuracy of classification. These classification diagnostics can help point toward the best fitting model with an eye cast toward finding meaningful and interpretable classes and for the purpose of replication.

After reviewing this information and deciding on how well the emergent classes defined modes of gameplay, we decided to focus on the 4-class model. Figure 2 shows the plot of item response probabilities for the 4-class model and should be viewed in conjunction with Table 3, which contains the item response probabilities. The plot shows four distinct classes each with a unique style of gameplay. There was evidence of latent class separation, with each of the classes having high endorsement but of different aspects of gameplay.

Table 3. Item response probabilities for the 4-class model

	Latent Class			
	1	2	3	4
	Fully Engaged	*Engaged in Training*	*Engaged in Building*	*Engaged in Driving*
Prevalence	6.38%	5.92%	29.00%	58.70%
Queue Match	**1.000**	0.006	0.022	0.021
Start Match	0.428	0.000	0.000	0.000
Queue Practice	**0.700**	0.000	0.291	0.052
Objective Done	**0.785**	0.311	**0.998**	0.475
Drive Mode	**0.999**	0.376	**1.000**	**0.944**
Skip Tutorial	0.155	0.535	0.370	0.000
Start Tutorial	0.369	**1.000**	**1.000**	0.000

The bold numbers represent probabilities exceeding .600
(i.e., 60% likelihood of being in this gameplay style if indicator is endorsed).

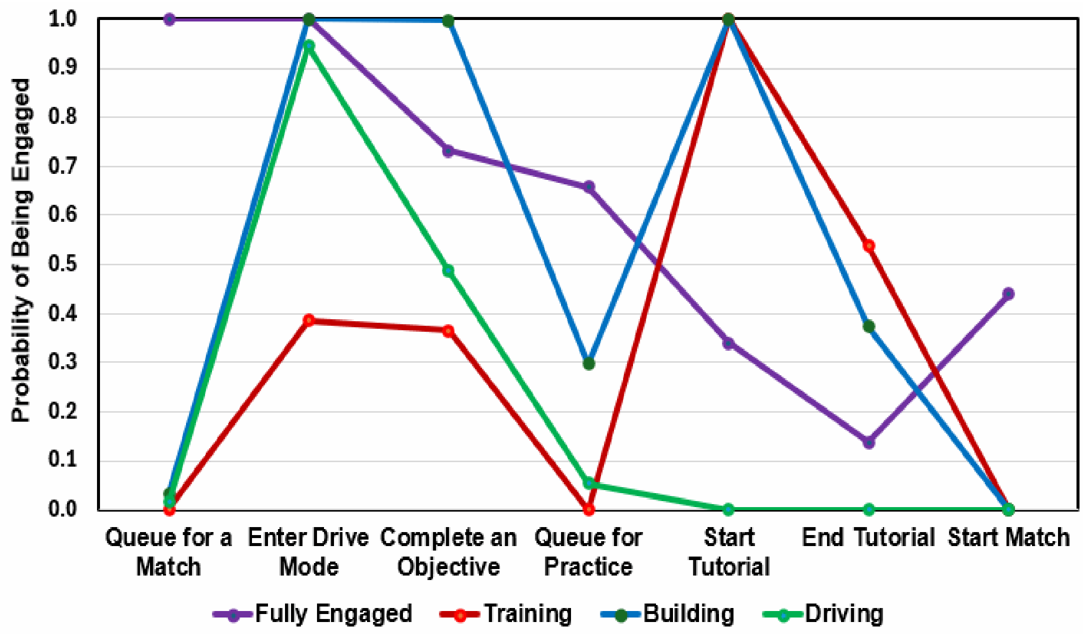

Figure 2. Graphic display of four unique classes of gameplay

We labeled Class 1 (6.38%) as "*Fully Engaged*" because in these games, players queued for matches and likewise queued for practice, completed objectives, and engaged the robot in drive mode. It should also be noted that there was minimal involvement in the tutorials for this gameplay style. Class 2 was labeled "*Engaged in Training*" (5.92%) because gameplay behavior in this class primarily involved starting the tutorial. One other item, albeit below the benchmark cutoff of .600, ended the tutorial (.535) suggesting that this style of gameplay used the tutorial intermittently or at best inconsistently. The third class, labeled "*Engaged in Building*" (29.00%) included gameplay that involved high participation in accomplishing objectives (.998), driving the robot (1.00), and utilizing the tutorials (1.00). Although other facets of the game indicating engagement were involved (e.g., queuing for practice), they were much less likely to characterize gameplay for this class. The fourth and largest class (58.70%) was labeled "*Engaged in Driving*" as this mode of gameplay involved only driving the robot (i.e., robots could be built and saved between games).

As a second factor in our decision to accept the 4-class model, the accuracy of assigning gameplay cases to their respective classes was relatively high (0.923 for Class 1, 0.993 for Class 2, 0.974 for Class 3, and 0.995 for Class 4). Conversely, the same classification metric showed that class-specific accuracy for off-diagonal mismatches was relatively low, ranging from a low of .001 for *Engaged in Building* classified as *Engaged in Training* to a high of .117 for *Engaged in Training* classified as *Engaged in Building*. These numbers stand in contrast to a much lower level of classification for the 3- and 5-class models and higher levels of misclassification as evidenced by the off-diagonal mismatches (available from 1st author). Finally, other factors that go into the decision to accept one model over another include the degree of separation between classes with the extraction of an additional gameplay style. In other words, does the addition of a class depict a unique gameplay style or just a fine discrimination that produces a relatively small class? In the current context, the addition of a gameplay style in the 5-class model did not produce a meaningful and distinctive style of play.

4.4 Conditioning Class Membership

We next conditioned class membership by adding covariates to the model. This process was conducted in two separate steps using the R3STEP procedure in the Mplus software. We first modelled *Intensity* (*Steps*/*Duration*) and *Games* as covariates using multinomial logistic regression with the 4-class model. Following this, we separated out the components of *Intensity* and modelled separately *Steps*, *Duration* (log transformed), and *Games*. Each model involves specifying the covariates as "auxiliary" variables and freezing gameplay styles into distinct and mutually exclusive categories. In a second set of models, we then modelled *Steps*, *Duration*, and *Games* as distal outcomes using the BCH procedure. The latter model provides a means to compare intercepts using pairwise comparisons for all the gameplay styles involved.

Table 4 shows the results of the multinomial logistic regression with the categorical latent factor regressed on the covariates. Model Sequence One contains the results when *Intensity* and *Games* were modelled. Model Sequence Two shows results for when *Intensity* was decomposed into *Steps* and *Duration*. When the odds ratios are greater than 1.0, they indicate a greater likelihood that the gameplay of the focal class had a higher score on *Intensity* or *Games* played compared to the reference class (*Engaged in Driving*). Likewise, for Model Sequence Two, a positive and significant odds ratio indicates the class being compared to the reference class had higher scores on *Steps*, *Duration*, and *Games*. For instance, for the comparison of Class 1 (*Fully Engaged*) to the reference class (*Engaged in Driving*), gameplay behavior was 89 times more likely to involve higher levels of *Intensity* ($p < .001$).

Table 4. Results of multinomial logistic regression predicting class membership

	Latent Class			
	Fully Engaged	*Engaged in Training*	*Engaged in Building*	*Engaged in Driving*
	Model Sequence One			
Intensity[1]	89.731***	1.358***	43.638***	ref
Games	1.001	0.946*	0.741***	ref
	Model Sequence Two			
Steps	2.414***	1.100***	2.252***	ref
Duration[2]	0.385***	0.311***	0.148***	ref
Games	1.004	0.983	0.888***	ref

[1]Intensity is the ratio of Steps/Duration. [2]Duration is log transformed. *p <05, ***p <001. Ref is the reference category for comparison. Odds ratios (OR) adjusted.

Model Sequence Two (lower portion of table) shows that for the same comparison (*Fully Engaged* versus *Engaged in Driving*), being *Fully Engaged* was 2.4 times more likely to involve increased numbers of steps being completed ($p < .001$) and 38.5% less likely to have a long duration of playtime ($p < .001$) with no significant difference in number of games played.

The final set of analyses used the BCH procedure to weight class membership based on the estimated posterior probabilities and assign gameplay styles to the different classes. Using pairwise comparisons each style of gameplay was contrasted with the others in terms of their mean values (intercepts) on *Steps*, *Duration*, and *Games*. The omnibus Wald test for the three outcome measures indicated there were significant intercept differences and we could proceed with the post-hoc pairwise comparisons: *Steps*, $\chi^2(3) = 5954.423$, *Duration*, $\chi^2(3) = 2835.374$, and *Games*, $\chi^2(3) = 476.678$, p's < .0001. Table 5 shows the results of the pairwise comparisons. As depicted, the *Fully Engaged* style of play had higher numbers of steps (3.706), played more games (0.608), and for a longer Duration (6.446) compared to the remaining three styles of gameplay. The *Engaged in Building* style of gameplay was next in steps (1.879), and Duration (5.096) but not games (0.211), which was the lowest. The *Engaged in Training* style of gameplay had the lowest number of steps (0.734), and Duration of play (4.403).

Table 5. Results of pairwise comparisons for latent classes predicting gameplay behaviors

			Mean Comparisons[1]	
Class Comparison	t-value	SE	Mean 1	Mean 2
Comparisons for *Steps* as a Distal Outcome				
Fully Engaged vs. Engaged in Training	18.189***	0.472	3.706	0.734
Fully Engaged vs. Engaged in Building	11.178***	0.482	3.706	1.879
Fully Engaged vs. Engaged in Driving	16.673***	0.472	3.706	0.981
Engaged in Training vs. Engaged in Building	-7.011***	0.106	0.734	1.879
Engaged in Training vs. Engaged in Driving	-1.515***	0.075	0.734	0.981
Engaged in Building vs. Engaged in Driving	5.495***	0.097	1.879	0.981
Comparisons for *Games* as a Distal Outcome				
Fully Engaged vs. Engaged in Training	2.307***	0.285	0.608	0.268
Fully Engaged vs. Engaged in Building	2.700***	0.255	0.608	0.211
Fully Engaged vs. Engaged in Driving	1.124***	0.264	0.608	0.443
Engaged in Training vs. Engaged in Building	0.393***	0.143	0.268	0.211
Engaged in Training vs. Engaged in Driving	-1.183***	0.157	0.268	0.443
Engaged in Building vs. Engaged in Driving	-1.575***	0.081	0.211	0.443
Comparisons for *Duration*[2] as a Distal Outcome				
Fully Engaged vs. Engaged in Training	2.093***	0.043	6.446	4.403
Fully Engaged vs. Engaged in Building	1.383***	0.031	6.446	5.096
Fully Engaged vs. Engaged in Driving	1.139***	0.030	6.446	5.335
Engaged in Training vs. Engaged in Building	-0.710***	0.036	4.403	5.096
Engaged in Training vs. Engaged in Driving	-0.955***	0.035	4.403	5.335
Engaged in Building vs. Engaged in Driving	-0.244***	0.016	5.096	5.335

[1]Means are standardized. [2]Negative value indicates the 2nd class had a higher mean value than the first class.
**p<01, ***p <001.

4.5 Gameplay Across Multiple Games

Of the 153,658 players, 299 played 12 or more games. Patterns of play were examined for these individuals. Overall, these players spent 78.7% of their gameplays driving robots, 14.2% of their gameplays were classified as *Fully Engaged*, 5.2% of their gameplays building their robots and just 1.9% of their gameplays focused on training. In general, except for how many sessions these players spent fully engaged and engaged in training, this matches the time spent observed in the entire sample.

Because the individual gamers played on multiple occasions, it was possible to assess how they spent their time across multiple game events. Nearly all these high gameplay individuals (99.3%) spent some time driving, a majority (69.9%) spent at least one gameplay session fully engaged, about half (48.5%) spent at least one gameplay session engaged in building their robots, and about one in five (19.1%) spent time solely engaged in training. However, analysis revealed that an additional 42.8% had started a tutorial as either part of being *Fully Engaged* or as part of being *Engaged in Building*, bringing the total percentage who started the tutorial to 51.2%. A final analysis showed that 8.4% of multi-game players engaged in all four unique classes of gameplay.

5. Discussion and Conclusions

Game metrics provide fine-grained measures of user experience that give game developers ways to obtain a sense of whether players like a game, find it engaging and pleasing as opposed to incredibly challenging or cumbersome. Compared to self-report, game metrics offer higher resolution and can afford a lens from which to see "inside the game" as the player engages the game mechanics. Metrics can also provide a useful barometer of players' willingness to progress in the game's challenges despite facing difficulties. The latter is construed in terms of persistence and can be used to gauge players' interest in the game (i.e., as a measure of "staying power"). If players encounter bottlenecks or sections of the game, they don't find intriguing, game developers can use this information to reformulate the game mechanics and gain clarity on how players approach the game (i.e., the drivers of gameplay). This is all part of finding out what intrigues players and what they find engaging as a prelude to continued game involvement.

In the current study, we modeled seven objective indicators of gameplay behavior using LCA to determine whether there are distinct gameplay styles that could be reliably identified. LCA is a mixture modeling approach that seeks to identify patterns in the data using probability theorems based on the joint marginal distributions. Rather than identify "individual" players that share common behavioral preferences and assigning them to unique classes, we identified common elements of playing a robotics game that form unique "styles of play." This shift in emphasis from player to persona was necessary given that players could engage the Roblox platform repeatedly over time and show different preferences depending on their game choices. We then validated the obtained gameplay styles using distinct measures of engagement including number of events that transpired in the game (i.e., challenges attempted), number of games played, and length of time engaged in the game. From these measures, we were able to compute a measure capturing "intensity of play," which reflects the degree of concentration and effort in how players engage or tackle the game. This type of high-quality data is unmatched in sheer size of the sample as well providing access to objective gameplay metrics that can be utilized to examine player engagement.

The combination of statistical fit indices and examining several other factors including class enumeration, model classification, and latent class prevalence suggested that the 4-class model fit the best and provided the clearest picture of uniquely different styles of gameplay. It is worth noting that other candidate models with fewer or more classes fit well based on the traditional model fit criteria. However, a careful inspection of these models indicated they did not present a clear picture of distinct modes of gameplay or with the extraction of additional classes produced small and unstable classes.

In the 4-class model, one relatively small class (*Fully Engaged*) was distinguished by a style of gameplay where players queued for matches and practice sessions, completed game objectives, and tested the robot's capabilities in drive mode. Tutorials were less of a concern and playing competitive matches was less emphasized. This gameplay mode reflected high levels of engagement in the robotics challenge and was further characterized by more events and longer play sessions compared to the other modes of play. The *Full Engaged* mode of play and its emphasis on matches and practice is crucial to understanding what features in a game motivates players. Players found in this class appear to be intrigued by the creative component of building and testing robots with less emphasis placed on actual competition. This mode of play apparently takes an incremental posture, slowly engaging all aspects of the game platform building and practicing to increase their learning objectives. It is worth speculating that this style of gameplay places a premium on "confidence" derived from repeated practice and testing rather than quickly jumping to competition and possibly losing in match play. A second and relatively small class of gameplay (*Engaged in Training*) was minimally involved in the game from a competitive standpoint and only utilized the tutorials to learn about the games. Further characterization of this gameplay included rarely queuing for matches, starting matches or practicing with their robot, and infrequently driving their robot, once designed. Training is typically an initial step in learning about a new experience, one that does not require frequent repetition. We expect that once trained generally (or for a particular challenge), players perceived little value to be gained from further training.

A third, and somewhat larger class (*Engaged in Building*) was a bit more involved and characterized by accomplishing game objectives, driving (ostensibly to test the robot), and starting the tutorial. In addition to these characteristics, this style of gameplay also included queuing for practice, albeit less frequently. For novice players, building a robot requires learning a wide range of specific engineering concepts related to robot design. Robot Champions™ somewhat simplifies the process in an age-appropriate manner; however, players must still master several technical skills. The game allows players a great deal of freedom to redesign their robotic creations, working through Newton's laws of motion in a trial-by-error fashion. For many, it may have been fulfilling to have the opportunity to spend time constructing and modifying their robots and was therefore worth spending time solely focused on this phase of gameplay.

The fourth and by far largest class (*Engaged in Driving*) was characterized by a singular focus limited to testing the robot in drive mode. Very limited events were associated with this gameplay style, with very limited queuing for matches, practicing, or utilizing the tutorials. Instead, this gameplay style was characterized by exploring the functional capabilities of robots by navigating the different environments (players could choose which game they wanted to build for their robotic challenge). The game offered the potential to refine robot builds (adding or subtracting features). Driving a robot affords opportunities to determine how it handles, navigates obstacles, turns, stops, and addresses challenges that may surface in a competition. While it is speculative on our part, it may be that driving one's robot is simply a uniquely enjoyable experience that players viewed as the single most "fun" aspect of Robot Champions™.

Further characterization of gameplay classes with measures of persistence and intensity provides a means to validate class membership as being uniquely different. This adds a layer of clarity because one would expect the most varied style inclusive of building, driving, practicing, and competing to be more event-driven, spend more time in the game meeting objectives, and concentrate more on what the game offers. When play was characterized as *Fully Engag*ed, gameplay was also characterized by a greater likelihood of being persistent and a relatively faster pace of gameplay. This stands in contrast to the *Engaged in Training* class, where gameplay was characterized by low persistence and a slower pace of gameplay. The other styles of gameplay, both *Engaged in Driving* and *Engaged in Building* had higher mean scores for *Steps* and *Duration* relative to the *Engaged in Training* but still lower than the *Fully Engaged* style of gameplay.

From these data, it is possible to conclude that not all players approach a game in the same way every time they log in to the Roblox platform to play Robot Champions™. In some cases, the tutorials represent a starting point while in other cases players shy from the tutorials while actively pursuing the game challenges and queuing for matches or driving the robot to assess its functionality. Either way, the seven indicators of gameplay were sufficient to discern several unique modes of gameplay. It is also worth noting that when all players were considered in the LCA model, accomplishing game objectives, queuing for practice and/or starting matches, and driving the robot were rarely combined with starting a tutorial. However, as seen when players who had logged in to multiple gameplay events, a full engagement style of play and participating in training was much more prevalent. It is perhaps the case that players' focus on building and driving robots is somehow preparatory for full engagement.

The non-competitive elements of gameplay (where gameplay consists only of building and testing robots) may simply be fun for many players. On the other hand, tutorials may only be seen as needed by players as either being a brief introduction to the game mechanics or needed for understanding specific elements of gameplay and can thus be referenced only sporadically. Research that sequentially tracks the activity of players from the time they initiate gameplay through full cycles of engagement may ultimately reveal how gameplay strategies evolve.

5.1 Implications for Game Developers

The type of modeling applied to log data in this study offers several pieces of information for game designers. For one thing, given a certain level of specificity (i.e., temporal and spatial resolution), game metrics can tell game developers how players interact with the game move by move, event by event yielding a crude but objective measure of how they interpret the virtual world [27]. This can include simple metrics such as progression (start and stop) and performance (build and drive) as well as more detailed information on their levels of engagement (specific actions they take and timestamps indicating player persistence in meeting challenges).

In the current context, players may struggle with building robots or getting them to execute movements in a way that conforms to engineering principles (e.g., balance, momentum, mass, acceleration). Robots may look cool and appear versatile for the task at hand but then tip over or not function properly. This can frustrate players to the point that they are initially enthusiastic but then quickly leave the platform once they encounter a setback. Tutorials will only help if they have educational value and provide some mystique or intrigue that the player wishes to engage further. There may be a benefit to including in-game metrics that note when robots fail to function. Linking error states to a tutorial cue (e.g., “Click here to see how to keep your robot from tipping over.”) may be a user-friendly way to increase tutorial participation. This can be extended to offer contextual hints within building interfaces that connect specific build components or design choices to success in a particular challenge (i.e., creating armature and rotational forces for throwing sheep into a pen). Adding more sandbox challenges to afford players opportunities to experiment with their robot build without time constraints or competitive pressures represents a good way to capitalize on the benefits of tutorials.

Scaffolding is another means of supplying tutorials in a natural or less didactic manner. The current study shows that tutorial participation was only pursued by two modes of gameplay. One style (*Engaged in Driving*) involved only driving the robot only while another mode characterized as “*Fully Engaged*” bypassed the tutorials completely. This may reveal the need for in-game (pop-up) tutorials that scaffold players one step at a time and increase engagement as players both have fun and acquire new skills. This approach can link learning and game mechanics in a way that builds player confidence early in the game experience. Other forms of scaffolding can involve mentoring new players and providing small token rewards at different stages rather than only after competition. Mentoring may be tied to persistence, especially in less engaged players who need more in-game support. Along these same lines, intermittent feedback on robot builds can serve as a wayfinding guide and “try it” prompts that appear during the robot build stages can be tied to in-game actions. Players can also engage in “test runs” where they can assemble a robot and test the addition of new parts without fully engaging practice or match competition. Here too, pop-ups and reminders tied to build strategies may encourage younger players to see through the complete robot build. Using trial or pre-built robots may also encourage further game participation as players can minimally invest and see outcomes more rapidly.

Skills-based matchmaking in the competitions may improve persistence especially to motivate players who lack experience in the build and test phases. This will help novice players grasp the fundamentals of the game and achieve favorable outcomes in the competition phases. This is tied to the benefits of reward structures, which psychological studies show is beneficial when doled out in even and consistent doses. Sharing robot designs coupled with opportunities to view other robot builds (and perhaps vote on them for style and efficiency) can promote learning through collaboration. Players can test out top-rated robots or featured “community” robots that afford players a chance to see how the robots fare in competition.

It is also important to note that the four gameplay styles are not mutually exclusive when examined across multiple trials. In other words, one style does not preclude another style if a player engages Robot Champions™ more than once. Here, skipped tutorials may happen once or twice, however, a player can return to play, engage in building, and resort to using a tutorial (this was characteristic of the *Engaged in Building* style). This is where pop-ups may be useful to encourage using the tutorial or revealing “slices” of the tutorial tied to whatever phase of build the player engages. Gating is a form of scaffolding where players can test more advanced features backed by tutorial knowledge, Trial and error runs that prepare a player through application of tutorial content is another way to encourage “testing robot builds” during practice sessions and in the absence of pressure from competition.

5.2 STEM Implications

Educational robotics has now emerged as a leading practical tool for teaching 21st century skills [19], [59] and preparing students for STEM by advancing technological competencies and computational skills [17], [60]. This can include computational thinking, scientific reasoning, and problem-solving skills. Recent meta-analyses reinforce that programs targeting these skills through robotics in general are successful as well as those that utilize digital game-based learning approaches [20], [61]. Blending robotics with digital games provides youth with innovative learning tools where they can acquire a host of important skills in a fun and engaging way. Although we did not study the relations between gameplay and STEM interests or motivation, this becomes an important milestone in the development of robotic games, especially given their emphasis on using technology and learning engineering principles. The success of video games rests on their being fun and enjoyable even if they are used in an educational context. This is perhaps why it is important to find out why players abandoned the game early, why average playing time was relatively low, and why players skipped tutorials that were designed to explain the game’s mechanics. With this information in hand, game developers can change various game mechanics, tailor features of the game to different skill levels or vary challenges to reinforce the game’s learning mechanics. It may be that players were frustrated with the game design as overly complicated given their limited knowledge of engineering principles and how to use the robot to achieve the desired outcome in competition.

Specific STEM skills that can be taught include developing a more systems thinking approach and applying decision-making skills that have functional value in situations that call upon knowledge of engineering and robotics. This might involve integration of the 7-steps of effective decision making [62] in tandem with teaching engineering principles and Newton’s laws of motion. For instance, players can be taught how to identify what their robot will need to function in the game challenge (i.e., what parts they need to build a robot that can meet the challenge requirements), how they can gather and incorporate the essential components based on a schematic design or a robotic image provided, identify alternatives to their build (e.g., view community builds), deliberate over how the robot will function in a real-world environment and given certain challenges, pilot test the robot in a non-competitive environment but with similar demands, select alternatives when the robot does not perform adequately, formally test the robot, including its control systems (e.g., mechanics and actuation sequences), rotational forces (e.g., torque, levers, pulleys), and providing age-graded tutorials that explain statics and dynamics, which is fundamental to robotics and STEM learning.

5.3 Limitations

There are several limitations associated with this study. First, we were not able to obtain any self-report measures of persistence or engagement. Adding external sources would help to validate whether we are really measuring persistence from an experiential point of view. This is important for the field given the need to combine contextual factors with gameplay metrics to really develop a sense of user experience. Without contextual data, we cannot address the “why” behind the “what” and know more about in-game factors that potentially motivates user engagement [26]. For instance, if a player skips the tutorial and goes directly to drive mode, we don’t know why they did not feel the need to queue for practice or engage a match to test their mettle against competitors. In addition, the various nomenclature we applied to characterize the different gameplay styles is nothing more than a convention. Inside each class is a mixture of user experiences that we need to learn how to tease apart. The heuristic value of the LCA model is to summarize these patterns in a way that makes theoretical or conceptual sense and that focuses attention of possible motivations for discrete episodes of gameplay. Ultimately, this information may be useful for improving game design.

It is also important to recognize that the unique gameplay classes extracted in the LCA are not individuals, pe se, but represent a multitude of gameplay events that are more like each other within rather than across classes. Traditionally, LCA produces “patterns” that include individuals rather than in-game events such as driving a robot or engaging a tutorial. In the current context, players could engage multiple games with a dissimilar mode of gameplay. Stated differently, a single player can be represented across time in multiple classes.

We were unable to conduct any subgroup analyses based on demographics including age or gender. This information is not part of the Mixpanel download available from Roblox currently. Given these data can at some point in time be made available, future studies may want to tease apart whether gender or age subgroups differ in their gameplay strategies and whether game metrics adequately and accurately capture these differences. For instance, older youth may have a more in-depth understanding of Newton’s laws of motion or engineering principles through physics coursework and therefore not require detailed instruction through tutorials. Their robot builds may be more successful straight from design to testing and require fewer practice sessions prompting them to engage in competition quicker. Only with detailed model comparisons using invariance tests can we tease apart the influence of age or gender.

Overall, the statistical fit indices were inconclusive in which model fit the sample data best. As expected, there was shrinkage in the information fit indices with the extraction of additional classes; however, other model-fit statistics did not definitively point to one model as superior to another. We thus combined the statistical information with classification diagnostics to select the 4-class model. Importantly, the four classes represent distinct gameplay styles (i.e., latent class separation) that is not evidenced in models with one more or one less class. The lack of conclusive statistical information suggests that the models may be affected by measurement issues (i.e., how well the individual game metrics point to distinct gameplay “styles”). This leads to classification uncertainty when measures have low reliability and results in lower entropy. One way to offset this is to include additional game metrics that can clearly point to distinct gameplay styles. We are currently working with the game developers to increase the size of the game metric variable pool. This will enable us to mine highly granular data and learn more about both player trends and game quality. Examples of the different type of metrics that could be informative (outside of demographics) include the time intervals spent on each event, which events (challenges) were encountered and in what order (sequentially). The latter metric can provide a window into what the player does first, second, third and so forth and at what point they disengage from the game. This information can be further dissected to see if parts of the game are not as engaging or well-designed interactively and informationally.

6. Conclusions

In-game metrics can be used to characterize the unique gameplay styles of players and their levels of engagement. The number of distinct occasions (events), duration of play, number of steps players to address a challenge take all reflect their gameplay persistence. Data from automatic log files that track the various actions taken by players is useful for understanding their in-game focus and how they approach gameplay. In the current context we found four distinct ways in which players approach a robotics game with each style reflecting a different degree of engagement. Future studies may want to follow the same player over time to determine whether their gameplay style changes with further game experience and learn more about what in-game factors (e.g., game mechanics, tutorials, visual graphics) influence their gameplay behaviors.

Overall, we conclude that in-game metrics related to persistence and intensity of gameplay offer game developers a unique window into how the games they develop perform in terms of player gameplay behavior. Many games offer such metrics. Indeed, the absolute volume of data available for analysis may well be overwhelming. However, it is only when games are analyzed in relation to patterns of players’ behaviors that game attractiveness can become clear. Methods and results described in this manuscript may provide useful information that program developers should consider adopting.

Acknowledgments

Lawrence M. Scheier and William B. Hansen received external funding from the National Institutes of Health, National Alcohol Abuse and Alcoholism Small Business Innovative Research Phase I grant (1R43AA031941-01 Hansen PI). Additional funding was provided by the North Carolina Department of Commerce, Office of Science, Technology, and Innovation, One North Carolina Small Business Program Matching Grant (NCBSTI–FY 24-25M).

Conflicts of interest

Alex Stone and Jennifer Javornik have a financial interest in Robot Champions. Players can play for free but may purchase game add-on items for which Filament Games benefits financially.

References

[1] R. Huang et al., "The impact of gamification in educational settings on student learning outcomes: A meta-analysis," Educational Technology Research and Development, vol. 68, pp. 1875-1901, 2020, https://doi.org/10.1007/s11423-020-09807-z.

[2] M. Qian and K. R. Clark, "Game-based Learning and 21st century skills: A review of recent research," Computers in human behavior, vol. 63, pp. 50-58, 2016, https://doi.org/10.1016/j.chb.2016.05.023.

[3] M. Ahmad, N. R. Mansor, R. A. Rashid, N. Ain, C. R. Zakaria, and C. M. Sung, "Implementation of digital games in advancing students' higher-order thinking skills: A review," in Journal of Physics: Conference Series, 2021, vol. 1793, no. 1: IOP Publishing, p. 012069, https://doi.org/10.1088/1742-6596/1793/1/012069.

[4] B. J. Morris, S. Croker, C. Zimmerman, D. Gill, and C. Romig, "Gaming science: the "Gamification" of scientific thinking," Frontiers in psychology, vol. 4, p. 607, 2013, https://doi.org/10.3389/fpsyg.2013.00607.

[5] D. B. Clark, E. E. Tanner-Smith, and S. S. Killingsworth, "Digital games, design, and learning: A systematic review and meta-analysis," Review of educational research, vol. 86, no. 1, pp. 79-122, 2016, https://doi.org/10.3102/0034654315582065

[6] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis," Computers & education, vol. 70, pp. 29-40, 2014, https://doi.org/10.1016/j.compedu.2013.07.033.

[7] M. Kalogiannakis, S. Papadakis, and A.-I. Zourmpakis, "Gamification in science education. A systematic review of the literature," Education sciences, vol. 11, no. 1, p. 22, 2021.

[8] M. Ullah et al., "Serious games in science education: a systematic literature," Virtual Reality & Intelligent Hardware, vol. 4, no. 3, pp. 189-209, 2022, https://doi.org/10.1016/j.vrih.2022.02.001.

[9] L. S. Vygotsky, Mind in society: The development of higher psychological processes. Harvard University Press, 1978.

[10] B. DeVane and K. D. Squire, "Activity theory in the learning technologies," in Theoretical foundations of learning environments: Routledge, 2012, pp. 242-267.

[11] National Research Council, Division of Behavioral Social Sciences Board on Science Education Committee on a Conceptual Framework for New K-12 Science Education Standards, A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. National Academies Press, 2012.

[12] S. Barab and C. Dede, "Games and immersive participatory simulations for science education: An emerging type of curricula," Journal of Science Education and technology, vol. 16, no. 1, pp. 1-3, 2007, https://doi.org/10.1007/s10956-007-9043-9.

[13] E. J. Hyslop-Margison and J. Strobel, "Constructivism and education: Misunderstandings and pedagogical implications," The Teacher Educator, vol. 43, no. 1, pp. 72-86, 2007, https://doi.org/10.1080/08878730701728945.

[14] M. Kandlhofer and G. Steinbauer, "Evaluating the impact of educational robotics on pupils' technical- and social-skills and science related attitudes," Robotics and Autonomous Systems, vol. 75, pp. 679-685, 2016, https://doi.org/10.1016/j.robot.2015.09.007.
[15] K. J. Newton, J. Leonard, A. Buss, C. G. Wright, and J. Barnes-Johnson, "Informal STEM: Learning with robotics and game design in an urban context," Journal of Research on Technology in Education, vol. 52, no. 2, pp. 129-147, 2020, https://doi.org/10.1080/15391523.2020.1713263.
[16] Y. Gui, Z. Cai, Y. Yang, L. Kong, X. Fan, and R. H. Tai, "Effectiveness of digital educational game and game design in STEM learning: a meta-analytic review," International Journal of STEM Education, vol. 10, no. 1, p. 36, 2023, https://doi.org/10.1186/s40594-023-00424-9.
[17] F. Ouyang and W. Xu, "The effects of educational robotics in STEM education: A multilevel meta-analysis," International Journal of STEM Education, vol. 11, no. 1, p. 7, 2024, https://doi.org/10.1186/s40594-024-00469-4.
[18] T. Sapounidis, S. Tselegkaridis, and D. Stamovlasis, "Educational robotics and STEM in primary education: a review and a meta-analysis," Journal of Research on Technology in Education, vol. 56, no. 4, pp. 462-476, 2024, https://doi.org/10.1080/15391523.2022.2160394.
[19] A. Eguchi, "Educational robotics for promoting 21st century skills," Journal of Automation, Mobile Robotics and Intelligent Systems, pp. 5-11, 2014, https://doi.org/10.14313/JAMRIS_1-2014/1.
[20] L.-H. Wang, B. Chen, G.-J. Hwang, J.-Q. Guan, and Y.-Q. Wang, "Effects of digital game-based STEM education on students' learning achievement: a meta-analysis," International Journal of STEM Education, vol. 9, no. 1, p. 26, 2022, https://doi.org/10.1186/s40594-022-00344-0.
[21] A. Drachen, "Behavioral telemetry in games user research," in Game user experience evaluation, R. Bernhaupt Ed.: Springer, 2015, pp. 135-165.
[22] A. Drachen and A. Canossa, "Towards gameplay analysis via gameplay metrics," in Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era, 2009, pp. 202-209, https://doi.org/10.1145/1621841.1621878.
[23] J. L. Plass et al., "Metrics in simuulation and games for learning," in Game analytics: maximizing the value of player data, A. Drachen, M. S. El-Nasr, and A. Canossa Eds.: Springer-Verlag, 2013, pp. 697-729.
[24] G. K. Chung, "Guidelines for the design and implementation of game telemetry for serious games analytics," in Serious games analytics: Methodologies for performance measurement, assessment, and improvement, C. S. Loh, Y. Sheng, and D. Ifenthaler Eds.: Springer-Verlag, 2015, pp. 59-79.
[25] A. Tychsen and A. Canossa, "Defining personas in games using metrics," in Proceedings of the 2008 conference on future play: Research, play, share, 2008, pp. 73-80, https://doi.org/10.1145/1496984.1496997.
[26] S. Kim, D. Choi, E. Lee, and W. Rhee, "Churn prediction of mobile and online casual games using play log data," PloS one, vol. 12, no. 7, p. e0180735, 2017, https://doi.org/10.1371/journal.pone.0180735
[27] A. Drachen and A. Canossa, "Analyzing spatial user behavior in computer games using geographic information systems," in Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era, 2009, pp. 182-189, https://doi.org/10.1145/1621841.1621875
[28] G. Wallner and S. Kriglstein, "An introduction to gameplay data visualization," in Game research methods, P. Lankoski and S. Björk Eds. Pittsburgh, PA, 2015, pp. 231-250.
[29] A. Canossa and A. Drachen, "Patterns of play: Play-personas in user-centred game development," in Proceedings of DiGRA 2009 Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory, 2009, https://doi.org/10.26503/dl.v2009i1.399.
[30] A. C. McCutcheon, Latent Class Analysis Sage Publications (Beverly Hills.[Google Scholar]). Sage, 1987.
[31] G. Faulkner, H. Irving, E. M. Adlaf, and N. Turner, "Subtypes of adolescent video gamers: A latent class analysis," International Journal of Mental Health and Addiction, vol. 13, pp. 1-18, 2015, https://doi.org/10.1007/s11469-014-9501-6
[32] M. Colder-Carras and D. Kardefelt-Winther, "When addiction symptoms and life problems diverge: A latent class analysis of problematic gaming in a representative multinational sample of European adolescents," European child & adolescent psychiatry, vol. 27, pp. 513-525, 2018, https://doi.org/10.1007/s00787-018-1108-1

[33] C. I. Chang et al., "Exploring subtypes and correlates of internet gaming disorder severity among adolescents during COVID-19 in China: A latent class analysis," Current Psychology, vol. 42, no. 23, pp. 19915-19926, 2023, https://doi.org/10.1007/s12144-022-03133-8
[34] N. D. Bowman and C. Chang, "Covariation among gaming motivations is correlated with anxiety and sociality: A latent class analysis," Entertainment Computing, vol. 45, p. 100546, 2023, https://doi.org/10.1016/j.entcom.2023.100546.
[35] H. Myrseth and G. Notelaers, "A latent class approach for classifying the problem and disordered gamers in a group of adolescence," Frontiers in Psychology, vol. 9, p. 2273, 2018, https://doi.org/10.3389/fpsyg.2018.02273
[36] K. Siste, E. Hanafi, L. T. Sen, R. Damayanti, E. Beatrice, and R. I. Ismail, "Psychometric properties of the Indonesian Ten-item Internet Gaming Disorder Test and a latent class analysis of gamer population among youths," Plos one, vol. 17, no. 6, p. e0269528, 2022, https://doi.org/10.1371/journal.pone.0269528
[37] M. C. Howard and M. P. Crayne, "Persistence: Defining the multidimensional construct and creating a measure," Personality and Individual Differences, vol. 139, pp. 77-89, 2019, https://doi.org/10.1016/j.paid.2018.11.005.
[38] A. Bandura, "Social foundations of thought and action," Englewood Cliffs, NJ, vol. 1986, no. 23-28, p. 2, 1986.
[39] C. S. Carver and M. F. Scheier, Attention and self-regulation: A control-theory approach to human behavior. Springer Science & Business Media, 2012.
[40] K. Drake, J. Belsky, and R. Fearon, "From early attachment to engagement with learning in school: the role of self-regulation and persistence," Developmental psychology, vol. 50, no. 5, p. 1350, 2014, https://doi.org/10.1037/a0032779.
[41] S. G. McClowry, "The temperament profiles of school-age children," Journal of Pediatric Nursing, vol. 17, no. 1, pp. 3-10, 2002, https://doi.org/10.1053/jpdn.2002.30929
[42] B. Jacobs, S. Prentice-Dunn, and R. W. Rogers, "Understanding persistence: An interface of control theory and self-efficacy theory," Basic and Applied Social Psychology, vol. 5, no. 4, pp. 333-347, 1984, https://doi.org/10.1207/s15324834basp0504_6
[43] V. J. Shute et al., "Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game," Computers & Education, vol. 86, pp. 224-235, 2015, https://doi.org/10.1016/j.compedu.2015.08.001.
[44] M. Ventura, V. Shute, and W. Zhao, "The relationship between video game use and a performance-based measure of persistence," Computers & Education, vol. 60, no. 1, pp. 52-58, 2013, https://doi.org/10.1016/j.chb.2013.06.033
[45] K. E. DiCerbo, "Game-based assessment of persistence," Journal of Educational Technology & Society, vol. 17, no. 1, pp. 17-28, 2014.
[46] D. L. Paulhus, "Measurement and control of response bias," in Measures of personality and social psychological attitudes J. P. Robinson, P. R. Shaver, and L. S. Wrightsman Eds.: Academic Press, 1991, pp. 17-59.
[47] L. M. Collins and S. T. Lanza, Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. John Wiley & Sons, 2009.
[48] H. Akaike, "Likelihood of a model and information criteria," Journal of econometrics, vol. 16, no. 1, pp. 3-14, 1981, https://doi.org/10.1016/0304-4076(81)90071-3
[49] G. Schwarz, "Estimating the dimension of a model," The annals of statistics, no. 6, pp. 461-464, 1978, https://doi.org/10.1214/aos/1176344136
[50] G. Celeux and G. Soromenho, "An entropy criterion for assessing the number of clusters in a mixture model," Journal of classification, vol. 13, pp. 195-212, 1996, https://doi.org/10.1007/BF01246098
[51] Y. Lo, N. R. Mendell, and D. B. Rubin, "Testing the number of components in a normal mixture," Biometrika, vol. 88, no. 3, pp. 767-778, 2001, https://doi.org/10.1093/biomet/88.3.767
[52] K. E. Masyn, "Latent class analysis and finite mixture modeling," in The Oxford handbook of quantitative methods, T. D. Little Ed.: Oxford University Press, 2013, pp. 551-611.
[53] K. Nylund-Gibson, R. P. Grimm, and K. E. Masyn, "Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models," Structural equation modeling: A multidisciplinary Journal, vol. 26, no. 6, pp. 967-985, 2019,

https://doi.org/10.1080/10705511.2019.1590146
[54] L. K. Muthén and B. O. Muthén, Mplus User's Guide. Eighth Edition. . Los Angeles, CA, 2017.
[55] T. Asparouhov and B. Muthén, "Auxiliary variables in mixture modeling: Three-step approaches using M plus," Structural equation modeling: A multidisciplinary Journal, vol. 21, no. 3, pp. 329-341, 2014, https://doi.org/10.1080/10705511.2014.915181.
[56] T. Asparouhov and B. Muthén, "Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model," Mplus web notes, vol. 21, no. 2, pp. 1-22, 2014.
[57] L. K. Muthén and B. O. Muthén, "How to use a Monte Carlo study to decide on sample size and determine power," Structural equation modeling, vol. 9, no. 4, pp. 599-620, 2002, https://doi.org/10.1207/S15328007SEM0904_8
[58] I. C. Wurpts and C. Geiser, "Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study," Frontiers in psychology, vol. 5, p. 920, 2014, https://doi.org/10.3389/fpsyg.2014.00920
[59] A. Ioannou and E. Makridou, "Exploring the potentials of educational robotics in the development of computational thinking: A summary of current research and practical proposal for future work," Education and Information Technologies, vol. 23, no. 6, pp. 2531-2544, 2018, https://doi.org/10.1007/s10639-018-9729-z.
[60] S. Anwar, N. A. Bascou, M. Menekse, and A. Kardgar, "A systematic review of studies on educational robotics," Journal of Pre-College Engineering Education Research (J-PEER), vol. 9, no. 2, p. 2, 2019, https://doi.org/10.7771/2157-9288.1223.
[61] Y. Zhang and Y. Zhu, "Effects of educational robotics on the creativity and problem-solving skills of K-12 students: A meta-analysis," Educational Studies, vol. 50, no. 6, pp. 1539-1557, 2024, https://doi.org/10.1080/03055698.2022.2107873.
[62] I. L. Janis and L. Mann, Decision making: A psychological analysis of conflict, choice, and commitment. Free press, 1977.