Achievement and Friends: Key Factors of Player Retention Vary Across Player Levels in Online Multiplayer Games

Kunwoo Park[*]   Meeyoung Cha[**]   Haewoon Kwak[†]   Kuan-Ta Chen[‡]

[*]Graduate School of Web Science Technology, School of Computing, KAIST, South Korea
[**]Graduate School of Culture Technology, KAIST, South Korea
[†]Qatar Computing Research Institute, Hamad Bin Khalifa University, Qatar
[‡]Academia Sinica, Taiwan

{kw.park,meeyoungcha}@kaist.ac.kr   haewoon@acm.org   ktchen@iis.sinica.edu.tw

ABSTRACT

Retaining players over an extended period of time is a long-standing challenge in game industry. Significant effort has been paid to understanding what motivates players enjoy games. While individuals may have varying reasons to play or abandon a game at different stages within the game, previous studies have looked at the retention problem from a snapshot view. This study, by analyzing in-game logs of 51,104 distinct individuals in an online multiplayer game, uniquely offers a multifaceted view of the retention problem over the players' virtual life phases. We find that key indicators of longevity change with the game level. Achievement features are important for players at the initial to the advanced phases, yet social features become the most predictive of longevity once players reach the highest level offered by the game. These findings have theoretical and practical implications for designing online games that are adaptive to meeting the players' needs.

Keywords

Player retention; virtual life trajectory; player level; online multiplayer games; longevity

1. INTRODUCTION

Player retention is a critical and long-running quest in online game industry. What makes players stay happy in a game and follow through its scenario? What makes them continue the game even after having reached the highest level offered? To answer these questions, researchers have studied the motivations of game players for over a decade [7, 24, 25]. Studies based on theoretical investigations, user surveys, and log data analyses have identified several factors that are critical to retention. For example, players are known to find enjoyment in games from completing missions, empowering through growth and level ups, forming communities, competing against other players, discovering plots and characters, and more.

Previous studies have tried to group these motivating factors and measure their relative strengths in retaining players. Researchers have found that players can be grouped into a few set of clusters based on their game motivations as action-social (i.e., players who enjoy fast-paced scenario with player interaction), mastery-achievement (i.e., players who indicate interests in narrative, expression, and world exploration), and immersion-creativity (i.e., players who appeal to strategic game plays, taking on challenges, and becoming powerful). Game designers carefully implement reward mechanisms of each motivation type throughout game scenarios to meet the needs of different players. Existing work has assumed that the relationship between players and motivations is rigid (e.g., does not change over time) and is irrespective of the players' virtual life phases.

This study brings a multifaceted aspect to this important question by examining retention over various phases of individual lifetime. We assume that one's potential and capacity to enjoy a game changes over time, and hence the need and the ability to achieve higher levels quickly and to socialize within games for cooperative shifts must be different for each individual. By observing in-game behavior logs throughout various phases of real game players, this paper sets out to answer the following research questions:

- For each phase within an online multiplayer game, what are the characteristics of players who achieve the next higher levels and get retained?

- Why do some individuals continue to play even after having reached the max level?

We utilize logs gathered from one of the oldest massively multiplayer online role-playing games (MMORPGs) in the world, Fairyland Online in Taiwan. We gained access to the complete set of actions of 51,104 individuals, describing their achievement logs (quests and level ups), financial logs (gaining wealth), as well as social logs (chats among players). Myriads of action logs on tens of thousands of individuals who ultimately achieved different levels and played the game for different amounts of time allow us to design a natural experiment on the lifetime retention problem. We identify the factors attributing to game longevity through the detailed log analysis and make the following observations:

1. Achievement features are important for players during the initial to advanced phases; players who are achievement-oriented and gather large amounts of rare items and virtual money are more likely to be retained and succeed in achieving the next levels.

2. Achievement-related traits, however, are no longer as important for players who reach the max level. Social features become the most predictive of success and longevity beyond this point.

3. Having strong social relationships (measured by the number of friends) is a good indicator of player retention and their effect continues to show significance through virtual life phases of players.

Our findings bring theoretical and practical implications for studying and designing online games. The finding that a player's needs vary over one's virtual life trajectory needs to be carefully addressed by further research and game designs. In particular, findings on longevity of the max level players is new. This finding is particularly important as their behaviors have not been studied much, even though expert players are valuable to the user ecosystem.

2. RELATED WORKS

Since Bartle [3] defined the four-type player taxonomy based on motivation in text-based games, there have been numerous efforts to understand why people join and continue to play online games. Among them is Yee's findings on three motivational components—achievement, social, and immersion—based on factor analysis of survey results on Bartle's player types. This study also identified that motivations can vary across different demographics. While general MMO players are found to be achievement-oriented [24], females were more likely to play online games to have social relationship with other players. The reason that players' motivation has been studied for decades is partly because of its ultimate connection to player retention.

Among recent findings, Debeauvais et al. [7] asked World of Warcraft players about their motivation for play and game usage patterns through questionnaire surveys and found that socially-motivated players are more likely to discontinue games while achievement-oriented players tend to continue. Borbora et al. [5] built a prediction model of player motivation from log data. From data mining experiments using player activity logs from Everquest II, they found achievement is a dominant motivation for predicting player churn (i.e., opposite of player retention). Above studies consistently report that achievement is a major motivation for retention in online games. On the other hand, some studies found social activity to be more important for retention. Based on the log data of EverQuest II, a study showed social influences from peers help predict player retention better [11].

Most recently, another group of researchers observed that game interactions such as interacting with toxic players can have negative impacts on retention in League of Legends [19]. As cyberbullying has been considered as one of the factors that make players annoyed, feel fatigued, and even leave the game [15], there have been much efforts to define, detect, and prevent toxic playing in online games [4,13,14]. However, in this work, we do not investigate the effect of cyberbullying on player engagement due to the limitation of our dataset.

While many studies have put efforts to contribute to understanding player retention, much of the findings have been drawn from a snapshot view—players aggregated by demographic features yet not considering *how they grow over time* within a game. Like human life itself, players face different challenges and engage in specific actions depending on


Figure 1: A screenshot of Fairyland Online [1]

their levels. For example, Ducheneaut et al. [8] observed that online game players are more likely to play alone in an early stage, but become socially active at higher level. Players need to collaborate with one another to defeat strong monsters or complete difficult quests as their level elevates. Moreover, players enjoy an entirely different in-game experiences once they achieve the maximum level, as they become socially active without consuming much game content [9]. This means that factors leading to higher levels or being retained may be different across the entire player lifetime within online games.

However, little attention has been paid to characteristics of churners over different phases of players. To the best of our knowledge, only one study by Shores et al. [19] investigated how indicators of player retention compare between new-joiners and experts in a MOBA (multiplayer online battle arena) game, and there is room for improvement. First, churn types can be examined for more than two groups. Player behaviors continuously change with level, and hence it is more natural to observe the whole picture of player life trajectories. Second, comprehensive data provide richer views. While the study relied on add-ons to gather data, the kinds of data that could be gathered externally is limited. Utilizing in-game logs provide full picture of player behaviors that might be important for predicting retention. Third, the studied game is a specific type that does not capture the growth of players naturally. MOBA game is repeated matches of the ground which have importance on team formations [12], yet MMORPG allows characters to explore and grow as individuals.

3. DATASET

Fairyland Online is one of the longest serviced MMORPGs, which has been played in Taiwan and other nearby countries since its launch in 2003. As depicted in Figure 1, the game is set on a virtual world that sets on fairy tales. Players can create their own avatars by choosing a race among human, elf, and dwarf and a gender of either female or male. On the virtual world, players explore their kingdoms, complete quests by fighting with monsters, and form social relationships with one another. Every action in the game is recorded in the game servers with accurate timestamps. Thanks to the Larger Network Technologies that serviced Fairyland Online, we gained access to the log data describing all actions that have been performed in the game.
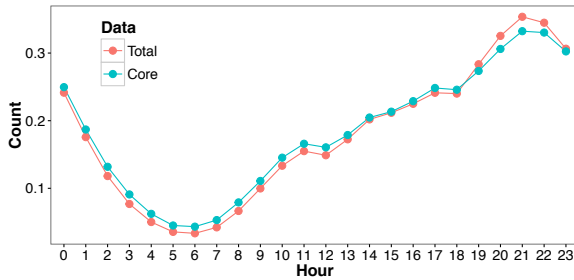
(a) Game access within a day
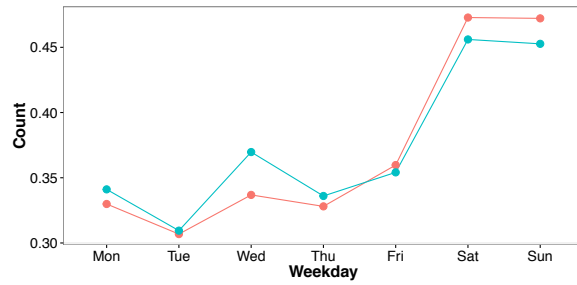
(b) Game access within a week

Figure 2: Temporal aspects from the normalized log counts of the entire and Core datasets

Fairyland Online servers log three different types of datasets. Firstly, there are logs related to achievement experience points (e.g., learn skills, completing quests). When a player gains enough experience points, his or her level will increase. Secondly, there are a group of actions related to gaining or losing wealth on the virtual world (e.g., buy or sell items, earn or use money). Game items can be retrieved by defeating monsters or by purchasing with virtual money. The last logs are about chats among players. There are four different channels to chat: Say, Whisper, Family, and Party. Say is a public channel through which a player can communicate with multiple companions. Messages on Say channel are broadcasted. Thus, if someone writes a message through Say channel, whoever in a virtual proximity can see it. Whisper is a private channel between two players. Because Whisper channel is private, no one except for the speakers and receivers can overhear those messages. Family is a dedicated channel for players who belong to the same 'family', which is equivalent to what is called 'guild' in other MMORPGs [20]. Party is the mode of communication for short-term groups. For privacy concerns, the chat contents themselves were encrypted.

The three kinds of datasets we received covered different time spans. We consolidated them to find a common overlapping period, over which we gain a full view of the achievements, financial, and social activities within the game. The overlapped portion covered nearly 60 million activity instances logged for 51,104 game players. We refer to this final complete dataset as Core in this paper and describe its statistics in Table 1. The table also shows the unique number of game players and record instances logged over the three original datasets.

Table 1: Dataset summary

Type	Period	Players	Records
Core	2003 4 21–2003 7 8	51,104	59,466,664
All	2003 2 21–2004 5 31	157,812	262,711,811

Given the timestamps of actions, we may infer when players accessed the game throughout a day and a week. The 24-hour plot shown in Figure 2(a) depicts that, for both the entire logs and for the Core period, players show strong diurnal patterns. The game was played more during late night with a peak between 8 pm and 10 pm. A disproportionately fewer players logged on during early morning times. In the morning, the highest time is around 11 am (before lunch time), then an increasing number of players join in the af-

ternoon and evening hours. Differences between the entire log and Core is marginal, indicating that the final Core data set we study is representative of the entire log in terms of temporal patterns.

Figure 2(b) presents the normalized daily access pattern across the data, again for the entire log and for Core. We find the game is played 1.4–1.5 times more frequently on weekends than during the weekdays. Later, we will investigate such detailed temporal features of players (e.g., weekend vs weekday oriented, most active time of a day) in predicting player retention. Note that the temporal patterns seen here have also appeared in other game studies [18], which suggests that the studied Fairyland Online shares commonalities with other representative MMORPGs.

4. METHODOLOGY

4.1 Phase Definition

The key objective of this research is to know what keeps a player in a game over various phases (and in particular toward the very end) of his or her virtual life. To answer this question, we start with arbitrary grouping of one's lifetime. In this work, we do not consider the very first phases of a game (i.e., beginners), which is a specific subset of our problem. We focus on players who have spent enough time to be accustomed to the rules of the game and determine what factors positively or negatively contributed to reaching the next level.

Players in Fairyland Online can have a level between the lowest of 1 and the highest of 50. Among such players, we decide target users to represent each phase of the online game by one quantity: the observable level ranges for each player during Core (i.e., level of 10–15, 20–25, 30–35, 40–45, and 45–50). Players belonging to the 10–15 level group must have achieved a level of at least 15 and have their traces since level 10 visible in the Core period.

Table 2 describes the five representative phases of virtual life that we examine in this paper. Where the exact division among groups lies is less of importance in this work. Rather, we are more interested in finding trends that become more prominent among the advanced, long-term game players. For each phase, we define success differently. The first four phases of 1–4 in Table 2 allow us to examine whether each player *successfully achieves* the next five levels. For phase 1 (i.e., level 10–15), we consider players who ultimately reach level 20 as success and otherwise as unsuccessful. For phase 5, the success is defined as whether the game is able to *keep a given player* or not. We decided a player is churned when

he or she becomes inactive for consecutive 90 days over the next 270 days. To test the validity of the length of days to decide user churn, we varied the number of days from 30 to 180 and they showed similar results with Table 7, which will be presented in the next section.

Table 2: Phase information along with grouping criteria, definition of success at a given phase, the number of corresponding individuals, and the probability of success within the group

Phase	Levels	Success defined	Num	Prob
1	10–15	Achieve level 20	3,818	0.6612
2	20–25	Achieve level 30	1,739	0.7483
3	30–35	Achieve level 40	1,370	0.8299
4	40–45	Achieve level 50	674	0.5638
5	45–50	Retained after level 50	221	0.4198

Out of the 51,104 individuals in the Core dataset, we only consider players whose observable levels are within the given ranges as described by Phase 1–5 as target players. We also ensure that each player has at least 31 days observable within in Core after the end of the observation level. For instance, for Phase 1, we ensure that individuals had at least 31 days after they reached level 15 in the log. This gives ample time for them to meet the success criteria (i.e., achieving level 20 for Phase 1). The buffer time of 31 days was determined from the log analysis. We investigated how long it takes to achieve the higher level for each phase listed in the table for a subset of 91 players who joined and achieved the maximum level during the Core data period. These players took 13.92 days on average with the standard deviation of 1.29 and a maximum of 30.97 days to achieve 5 level ups (i.e., from level 45 to level 50). Thus, we set the buffer length as 31 days. For Phase 5 players, however, we did not enforce this buffer length, as they no longer need to achieve a higher level. For these players, churn was measured over the entire log period (beyond the Core period). From Phase 1 to Phase 5, we identified 3818, 1739, 1370, 674, and 221 players meeting the above criteria, respectively. Note that a player can belong to multiple groups so long as he or she meets the criteria.

The last column of Table 2 displays the probability of success for each phase, where the success criterion is also listed in the table itself. The fraction of players who success in a given phase is the highest among Phase 3 players (i.e., individuals who were in level 30–35) and is the lowest for Phase 5 players (i.e., individuals who reach the highest level offered). Nonetheless, the success probability or the retention rate is considerably stable, remaining over 0.4 throughout the five phases.

4.2 Studied Features

We utilize a total of 16 features for predicting user retention. The features are divided into three major categories based on their characteristics: temporal, achievement-related, and social.

4.2.1 Temporal Category

The temporal features describe when individuals played the game. Temporal patterns not only reveal how often a user plays the game (e.g., every day vs. once a week) but also reveal certain demographic traits. For example, play time

can be used to infer which players are likely students (e.g., peak immediately after the school hours) or which players likely work regular hours (e.g., playtime starts only after the typical business hours). We extract two features like below:

- Frequent hours (*morning*, *working_hour*, *evening*, and *night_owl*): To capture playing patterns, we measured how frequently a player accesses the game with the following time blocks. We define 4 variables that represent specific playing patterns: morning from 6 am to 9 am, working hours from 9 am to 6 pm, evening from 6 pm to midnight, and night owl from midnight to 6 am. We conducted vector normalization on those variables to remove the effects of the number of engaged days.

- Weekday vs weekends (*weekends*): the fraction of playtime contributed from weekends.

4.2.2 Achievement-Related Category

Many studies have reported the importance of achievement as a goal in player retention [5, 7]. To measure its effect, we utilize the following features related to in-game achievements.

- Possessed item count (*item*): the total count of owned items measured by subtracting the number of item-losing logs from item-gaining logs. This quantity is a proxy of in-game achievements.

- Rare items on hand (*rare_item*): Owning rare items can be more important achievement than general item counts. To decide which items are rare, we approximated chances of getting an item by counting the frequency over the whole item frequencies measured from Core. Then, we considered items appeared with the probability lower than 0.01 to be rare items. We measured the number of rare items on hand in the same manner as we did to count for items on hand.

- Amount of money on hand (*money*): the amount of virtual money that each player has, calculated by the differences between money-gaining logs and money-losing logs.

- Level of difficulty (*difficulty*): the level of difficulty, measured by a combination of the number of deaths and broken items. The appropriate level of difficulty has been considered as an important element for user engagement in online games [6].

- Performance (*performance*): The performance of achieving level can represent level of motivations on achievement. We measured it by changing the sign of time length of observation period, because each user can take different time to achieve 5 levels of the observation period based on his or her performance. A larger performance value hence indicates that a player leveled up quickly.

4.2.3 Social Category

Social features are another important group of indicators for player engagement [9, 11, 23]. Below we describe the list of social features we tested in this paper.

- Number of social interactions (*num_social*): the number of all messages that a player sent through any channel—a measure of social activeness.

- Response rate (*response_rate*): the probability of giving responses when a player received messages from an unknown player—a measure of social openness.

- Number of friends (*friends*): To figure out effects of social interactions in a more detail, we define friendships. Based on whisper logs, we counted the number of distinct days paired communications take place. If a player has paired communications with another player for at least three different days, we considered the communication partner to be a friend. To measure this variable, we counted the number of friends who communicated at the observation period as a feature for retention.

- Number of non-friends (*nonfriends*): We considered those who have paired communications but are not friends to be non-friends. The number of non-friends were observed from the observation period of each phase.

- Friends' level (*friends_level*): To represent the level of friends, we got the median level of friends who have communicated on the observation period.

- Non-friends' level (*nonfriends_level*): the median level of non-friends, who have communicated with the player during the observation period.

- Number of max-level friends (*friends_maxlevel*): the number of friends who communicated with the player and already achieved the maximum level at the moment of communication on the observation period.

- Number of max-level non-friends (*nonfriends_maxlevel*): the number of non-friends who have paired communication with the player and already achieved the maximum level when the communication happens.

- Is a member of a family (*has_family*): a binary variable whether the player belongs to a family, which is a membership-based group. We inferred it based on whether a user has sent messages through Family channel.

4.3 Player Retention Model

A logistic regression model was used to determine factors that affect player longevity. The regression model helps us investigate how various indicators attribute to player retention across different life phases within the virtual world, while allowing us to control for the effects of other variables. Hence we choose to use interpretable models in this paper rather than implementing other kinds of prediction models that might achieve higher performance.

Prior to analyses, a step was taken to balance the data. Because the success rate at each phase is biased toward one side, we employed an over-sampling technique to prepare an equal number of success and fail cases for each phase. All variables were scaled to have a mean of 0 and a standard deviation of 1. In addition, since variables of regression models can turn out to be significant simply due to a large number of predictors, we performed variable selections using Lasso [22] by choosing the lambda whose cross-validated mean squared

error is within one standard error of the minimum. In the results section, we report findings of the regression fitting after this feature selection step.

5. RESULTS

This research assumes the important indicators of the player retention vary throughout the different phases in Fairyland Online. To test this idea, for each phase of the game, we fit the logistic regression model of the successful cases and unsuccessful cases (as defined in Table 2) across the 16 features from three categories (i.e., temporal, achievement, and social). We compare the relative importance of each category in predicting player retention.

5.1 Low- to Medium-level Patterns

Among the five phases of the game level, here we focus on Phase 1 to Phase 3, which are logs of players who have become accustomed to the game. Our goal is to understand what kinds of players are more likely to be retained and further succeed in achieving the next levels. The three phases were observed among more than a thousand individuals. Below we only list the final set of features deemed as meaningful (out of the 16 features) for each of the three phases, after the Lasso variable selection step.

Table 3: Results for Phase 1 players (level 10–15)

Category	Predictor	Estimate	Significance
	(Intercept)	0.062	
Temporal	*night_owl*	0.424	***
	weekends	-0.244	*
Achievement	*performance*	0.371	***
Social	*friends*	0.539	***
	nonfriends	-0.331	***
	nonfriends_maxlevel	-0.09	**
Model χ^2		435.8	***

*:p<0.05, **:p<0.01, ***:p<0.001

Table 3 presents the fitted results for Phase 1, which shows the estimates and significance of variables. The table also shows the model χ^2 value based on the likelihood ratio test with a null model. We see that at least one feature from the temporal, achievement, and social categories appears as significant. Among the temporal features, *night_owl* is positively associated with the success, while *weekends* is negatively correlated. This means individuals who mainly played after midnight and during weekdays (but not just on weekends) were more likely to reach the next levels— This could indicate that at an early stage, time dedication is an important marker of success. Among the achievement features, *performance* (i.e., negative quantity of the playtime) increases the probability of success in that players with speedy game style are more likely to be retained and achieve the next levels. As many studies found, achievement is one of the main motivations for continuing to play online games [5, 7]. Our analysis also confirms that in an early virtual stage achievement help players reach the next levels without leaving games.

Among the social features, *friends* is positively associated with success, while *nonfriends* is negatively associated with success. This may indicate that players with many friends yet fewer weak social ties are more likely to achieve the next

levels. This trend supports findings from several studies on the importance of social interactions in games [9, 11]. In contrast, the negative effect of *nonfriends* is interesting. It may indicate that communication with too many random users may be harmful for long-term engagement. Furthermore, *nonfriends_maxlevel* shows negative association with the success rate in that players who communicate with many max-level non-friends are less likely to continue with the game. Communicating with too advanced non-friends may be a negative experience on future engagement, because players can feel left behind [21].

Table 4: Results for Phase 2 players (level 20–25)

Category	Predictor	Estimate	Significance
	(intercept)	-0.016	
Temporal	*morning*	-0.096	*
	evening	0.072	
	night_owl	0.184	***
Achievement	*item*	-0.203	**
	money	0.252	***
	performance	0.497	***
Social	*friends*	0.53	***
	nonfriends	-0.302	***
	friends_level	-0.095	*
	friends_maxlevel	-0.177	***
	nonfriends_maxlevel	0.124	*
	has_family	-0.129	**
Model χ^2		281.34	***

Phase 2 players indicated several more number of significant variables related to retention, as shown in Table 4. Among the temporal features, *night_owl* is positively associated with success, while *morning* is not. It seems still important to devote extra time on the game after midnight for achieving higher levels, while playing the game since early morning (i.e., 6–9 am) seem a not effective strategy for further engagement. Among the achievement features, *performance* is again positively associated. We newly found *money* to be positively correlated, yet *item* is negatively associated. Accumulating in-game money increases the probability of the continued usage at Phase 2, because it may become difficult to quit the game after one gathers a large sum of virtual money. In addition, virtual wealth means one's ability to upgrade game avatars, which helps achieve the next levels easier. Those two possible explanations can be linked to the success of achieving more levels. However, simply owning many items decrease the chance of success.

Among social features, *friends* is again positively associated with the success, while *nonfriends* is not. This finding implies that having more friends and fewer weak social relationship is linked to helping players achieve higher levels. Also, *has_family* is newly found to be significant with a negative estimate at this stage. In other words, players are less likely to succeed when they joined a family. This trend also supports the importance of focusing on strong social relationship to the continued usage of online games. Lastly, we observed that *friends_maxlevel*, and *friends_level* are negatively associated with success. This trend can be similarly explained with the negative association of *nonfriends_maxlevel* for prediction of Phase 1.

Players in Phase 3 exhibit similar trends as in Phase 2 (Table 5). Among temporal features, *night_owl* is positively

Table 5: Results for Phase 3 players (level 30–35)

Category	Predictor	Estimate	Significance
	(intercept)	-0.017	
Temporal	*working_hour*	-0.082	
	night_owl	0.251	***
	weekends	-0.199	***
Achievement	*item*	-0.274	**
	rare_item	0.292	***
	money	0.265	**
	difficulty	0.125	**
	performance	0.686	***
Social	*num_social*	-0.234	**
	response_rate	-0.106	
	friends	0.607	***
	nonfriends	-0.28	***
	friends_level	-0.133	**
	nonfriends_level	0.114	*
	nonfriends_maxlevel	-0.077	
Model χ^2		375.67	***

correlated, yet *weekends* is not—This pattern is similar to what we have seen before in earlier phases. Among the achievement features, *performance* is again important for predicting player retention. There are some new trends; while *item* is still negatively associated, players who gather larger amounts of rare items (i.e., *rare_item*) are likely to succeed. With the positive association of *money*, this finding supports the claim that owning virtual wealth is related to the success of achieving higher level. In addition, *difficulty* (i.e., the number of deaths and broken items) was found to be a positive indicator for success. Once the game reaches a certain stage, an appropriate level of difficulty may help players better enjoy games, as reported in a previous study [6].

Next, from social features, *friends* is positively correlated, while *nonfriends* and *num_social* are not. Again this finding demonstrates the importance of having communication with close friends rather than simply being socially active. Lastly, among variables on whom users talked to, *nonfriends_level* was newly found as a positive estimator. We hypothesize that players who ask for more help to other players of higher level are more likely to succeed. As reported in previous works [10, 16], communicating with experts can sometimes be helpful in online games because they share knowledge, useful tactics, and strategies that are critical in proceeding with the next phases. Because this process does not require having any strong relationships with those with high level, *nonfriends_level* is a positive indicator yet *friends_level* may remain to have the opposite effect. As discussed in the results seen in earlier phases, having social relationships with users with max level or higher may give detrimental effects on future engagement.

In summary, we found two consistent trends from the regression analysis of low- to medium-level phases (i.e., Phase 1-3). One is that performance on achieving levels and playing patterns related to devoting more times increase the probability of success for achieving more levels. These findings can be connected to the importance of motivation on achievement for player retention. Another is that players who have more friends yet fewer weak social relationships were more likely to be engaged in Fairyland Online continuously. Playing games with friends may give positive

effects on achieving more levels. These findings are consistent with previous findings on player retention on other games [5, 7, 9, 11].

5.2 High-level Patterns

Players who reached a level 40 or above out of 50 in Fairyland Online may be considered advanced users. What are the factors that lead to successfully reaching the endgame for these advanced players? Table 6 displays the regression result for Phase 4 players. At this very last stage, the only meaningful feature left after the Lasso feature selection is *performance* (i.e., -1×play time). Players who enjoy speedy game and are quick at leveling up are more likely to succeed to reach the max level. It is interesting to see that achievement-related feature alone is a critical factor of success.

Table 6: Results for Phase 4 players (level 40–45)

Category	Predictor	Estimate	Significance
	(intercept)	-0.007	
Achievement	*performance*	0.485	***
Model χ^2		40.505	***

Once players reach the max level, a different story unfolds. In contrast to the Phase 4 players, the only meaningful feature of longevity left at this stage is the social category, where *friends* is the only significant indicator for retention for players who reach the highest level. This finding suggests that having a substantial number of friends is consistently important in determining who will continue to play online games even after completing all missions. Note that this variable was also important during the earlier phases, further suggesting the importance of social interactions for player retention from earlier stages to the endgame. As found in previous studies [9], online games become more of a social space after the max level. To be engaged in such online games in the long run, players must have constructed strong social relationships from early on in their virtual lives.

Table 7: Results for Phase 5 players (level 45–50)

Category	Predictor	Estimate	Significance
	(intercept)	0.035	
Social	*friends*	0.966	***
Model χ^2		44.172	***

5.3 Trajectory Over Lifetime

Having examined the factors of player retention step-wise, we now jointly view trends over the entire life stages in the Fairyland Online game. The set of features examined are from three main categories: temporal, achievement-related, and social. Which of these categories are important for predicting player retention at each phase? To answer this question, we compared the relative importance of the three categories in predicting player retention via training separately on each categorical features. For testing, 5-fold cross validation was used on the unbalanced original dataset with keeping the distribution of target labels. Then, we conducted
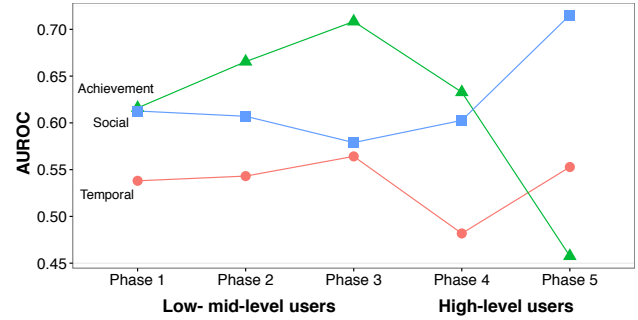

Figure 3: Prediction accuracy across the five phases seen by the area under the ROC curve (AUROC) of logistic regression for each feature category

over-sampling for each split to be balanced. We applied this sampling technique after each split to prevent for same instances to be both in training and test set for each split. We finally measured the area under the ROC curve (AUROC) of logistic regression classifiers using each set of features.

Figure 3 presents the changes of the AUROC values of the three categories over the five phases. The AUROC value is between 0 and 1, where a value of 1 means the prediction model is perfect. A prominent trend we see is the role of achievement features that show the best performance during the early to late phases of the game (i.e., Phase 1 to 4). The social category shows a comparable trend to the achievement category for Phase 1, 2, and 4. This category then becomes the most important in Phase 5 (i.e., the max level players), at which point the other features are no longer important. The temporal features are, for most of the phases, better than random guessing (i.e., AUROC of 0.5), although they do not show a big gain against the baseline. We discuss implications of these findings in the next section.

6. DISCUSSION & CONCLUSION

Maintaining user base of a substantial size is critical for many companies in running their services. Companies across various industries (e.g., telecommunication companies [2], health app providers [17], and so on) have put their efforts to understand characteristics of people who discontinue services and predict them in advance based on data mining approaches. Game industry and researchers also noticed the importance of the player retention problem, and many studies hence tried to understand player motivations [24, 25], behavioral characteristics [8, 9], and to build prediction models based on studied characteristics [5, 11].

However, existing studies did not disentangle user groups and conducted analyses without considering player levels. Because game designs of MMORPGs let players to have a certain amount of activities as level increases [8], game players face different challenges as their level increases and this evolution can affect user retention. As in Figure 4, we also observed that social interactions increase as level increases in our dataset. Thus, to precisely understand indicators for player retention, effect of features should be separately measured across different virtual life phases in online games. Another aspect that has received little attention is retaining individuals who have reached the highest level offered
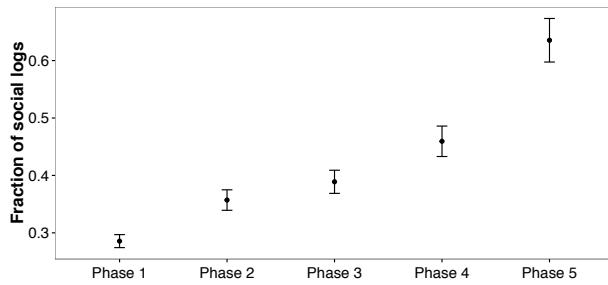
Figure 4: The fraction of logs related to social activities out of the entire logs at different life phases

by the game. These expert players not only help newbies adapt to the game, but also are a major source of profit for game industry. Therefore, retaining the max-level players is a critical problem.

Motivated by these missed opportunities, this research aimed to answer two research questions: (i) what are the indicators for player retention over different phases of players and (ii) how does the relative importance of retention features change over the game phases. Through a series of quantitative analyses using 51,104 individuals based on in-game logs, we have made several key findings for the question. Theses results are important for the following reasons. Firstly, we noted that the key indicators of longevity change with player phases. This finding implies that other studies on user behavior also need to consider phases of gamers. Secondly, our findings have practical implications to on-line game developers, as they need to carefully consider the changing needs of players over various life stages. Game designers may offer fast achievement-oriented scenarios at the beginning, while motivate players to form strong social relationships long before they reach any advanced level. We note that these suggestions are hypothetical, because observations indicate correlation not causality. Future studies can conduct controlled experiments or qualitative studies to further test causal relationship of feature effects. Another implication is that game industry may apply these findings to construct churn prediction models. For example, prediction models could be constructed separately for each phase and hence better capture signals of churning individuals.

On top of the above findings, we also found significant indicators observed for certain phases. Playing after midnight was positively associated with the success of Phase 1 to 3, while playing in the early morning is negatively associated with the continued usage. There may exist certain playing patterns, which can be linked to player retention. Also, obtaining rare items or larger amount of money increased the chance of success in low- to medium-level. Owning virtual wealth may be helpful to achieving more levels, or it could make them feel commitment to keep playing online games by having a large amount of wealth in the virtual world. Lastly, we found significant indicators on whom a player talked to. For example, the level of friends was a significant indicator during the initial phases. This finding implies that social network positively affects retention when individuals form interactions with partners of appropriate levels. Because these findings are newly found in this study, predicting player retention can be better improved with further investigation on those variables.

This paper has several limitations. Among them is the use of a single data source. Every MMORPG has different game elements and player traits, and hence our findings can not be directly generalized to other online games. Nonetheless, we expect Fairyland Online is representative of a typical MMORPG in terms of the temporal trends, which shows similarity to other games [18]. In the future we hope to replicate the study with other online game logs. Another limitation is that, although we tried extensive features across three different categories based on the related literature, there can exist missing features which might be critically linked to player retention. For example, the number of churned friends was found to be an indicator of player churn in one online game [11]. Due to limited data, we could not employ this feature for analysis. In a future work, we hope to look into a longer time period and investigate the effects of other possible indicators including churned friends. Lastly, we did not investigate players who are at their very early stages (i.e., level 1-10). This was because the initial level up in Fairyland Online was fairly easy and there was not much data associated with this time period. However, new-joiners are of great interest to game industry because they are critical to increasing user base and future studies can delve deeper into the behaviors of new-joiners across different games.

Acknowledgement

7. REFERENCES

[1] Fairyland Players. `http://bit.ly/2mb19UO`, 2008. [Online; accessed 19-Feb-2017].

[2] A. Backiel, B. Baesens, and G. Claeskens. Predicting Time-To-Churn of Prepaid Mobile Telephone Customers Using Social Network Analysis. *Journal of the Operational Research Society*, 2016.

[3] R. Bartle. Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. *Journal of MUD research*, 1(1):19, 1996.

[4] J. Blackburn and H. Kwak. STFU NOOB!: Predicting Crowdsourced Decisions on Toxic Behavior in Online Games. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 877–888. ACM, 2014.

[5] Z. Borbora, J. Srivastava, K. Hsu, and D. Williams. Churn Prediction in MMORPGs Using Player Motivation Theories and an Ensemble Approach. In *Proceedings of the International Conference on Privacy, Security, Risk and Trust*, pages 157–164. IEEE, 2011.

[6] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun. Boredom, Engagement and Anxiety as Indicators for Adaptation to Difficulty in Games. In *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*, pages 13–17. ACM, 2008.

[7] T. Debeauvais, B. Nardi, D. Schiano, N. Ducheneaut, and N. Yee. If You Build It They Might Stay:

Retention Mechanisms in World of Warcraft. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 180–187. ACM, 2011.

[8] N. Ducheneaut, N. Yee, E. Nickell, and R. Moore. "Alone Together?" Exploring the Social Dynamics of Massively Multiplayer Online Games. In *Proceedings of the 24th Conference on Human Factors in Computing Systems*, pages 407–416. ACM, 2006.

[9] N. Ducheneaut, N. Yee, E. Nickell, and R. Moore. Building an MMO With Mass Appeal: A Look at Gameplay in World of Warcraft. *Games and Culture*, 1(4):281–317, 2006.

[10] D. Fields and Y. Kafai. A Connective Ethnography of Peer Knowledge Sharing and Diffusion in a Tween Virtual World. *International Journal of Computer-Supported Collaborative Learning*, 4(1):47–68, 2009.

[11] J. Kawale, A. Pal, and J. Srivastava. Churn Prediction in MMORPGs: A Social Influence Based Approach. In *Proceedings of the Computational Science and Engineering*, volume 4, pages 423–428. IEEE, 2009.

[12] J. Kim, B. C. Keegan, S. Park, and A. Oh. The Proficiency-Congruency Dilemma: Virtual Team Design and Performance in Multiplayer Online Games. In *Proceedings of the 34th Conference on Human Factors in Computing Systems*, pages 4351–4365. ACM, 2016.

[13] H. Kwak and J. Blackburn. *Linguistic Analysis of Toxic Behavior in an Online Video Game*, pages 209–217. Springer International Publishing, Cham, 2015.

[14] H. Kwak, J. Blackburn, and S. Han. Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games. In *Proceedings of the 33rd Conference on Human Factors in Computing Systems*. ACM, 2015.

[15] J. Mulligan and B. Patrovsky. *Developing Online Games: An Insider's Guide*. New Riders, 2003.

[16] B. Nardi, S. Ly, and J. Harris. Learning Conversations in World of Warcraft. In *Proceedings of the 40th*

Annual Hawaii International Conference on System Sciences, pages 79–79. IEEE, 2007.

[17] K. Park, I. Weber, M. Cha, and C. Lee. Persistent Sharing of Fitness App Status on Twitter. In *Proceedings of the 19th Conference on Computer Supported Cooperative Work & Social Computing*, pages 184–194. ACM, 2016.

[18] D. Pittman and C. GauthierDickey. A Measurement Study of Virtual Populations in Massively Multiplayer Online Games. In *Proceedings of the 6th SIGCOMM Workshop on Network and System Support for Games*, pages 25–30. ACM, 2007.

[19] K. Shores, Y. He, K. Swanenburg, R. Kraut, and J. Riedl. The Identification of Deviance and Its Impact on Retention in a Multiplayer Game. In *Proceedings of the 17th Conference on Computer Supported Cooperative Work & Social Computing*, pages 1356–1365. ACM, 2014.

[20] C. Steinkuehler and D. Williams. Where Everybody Knows Your (Screen) Name: Online Games as "Third Places". *Journal of Computer-Mediated Communication*, 11(4):885–909, 2006.

[21] E. Tandoc, P. Ferrucci, and M. Duffy. Facebook Use, Envy, and Depression among College Students: Is Facebooking Depressing? *Computers in Human Behavior*, 43:139–146, 2015.

[22] R. Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[23] A. Tyack, P. Wyeth, and D. Johnson. The Appeal of MOBA Games: What Makes People Start, Stay, and Stop. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 313–325. ACM, 2016.

[24] D. Williams, N. Yee, and S. Caplan. Who Plays, How Much, and Why? A Behavioral Player Census of a Virtual World. *Journal of Computer Mediated Communication*, 13(4):993–1018, 2008.

[25] N. Yee. Motivations for Play in Online Games. *CyberPsychology & Behavior*, 9(6):772–775, 2006.