eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Predicting Skill Learning in a Large, Longitudinal MOBA Dataset

M. Aung
Department of Computer Science
Department of Psychology
University of York
York, United Kingdom YO10 5DD
Email: mta510@york.ac.uk

V. Bonometti
Department of Computer Science
University of York
York, United Kingdom YO10 5DD
Email: vb690@york.ac.uk

A. Drachen
Department of Computer Science
University of York
York, United Kingdom YO10 5DD
Email: anders.drachen@york.ac.uk

P. Cowling
Department of Computer Science
University of York
York, United Kingdom YO10 5DD
Email: peter.cowling@york.ac.uk

A.V. Kokkinakis
Department of Computer Science
Department of Psychology
University of York
York, United Kingdom YO10 5DD
Email: avk504@york.ac.uk

C. Yoder
Riot Games
Los Angeles, California 90064
Email: cyoder@riotgames.com

A. Wade
Department of Psychology
University of York
York, United Kingdom YO10 5DD
Email: wade@wadelab.net

Abstract—The exploration of the relationships between behavior and cognitive psychology of game players has gained impetus in recent years because such links provides an opportunity for improving user experiences and optimizing products in the games industry. At the same time, the volume and global scope of digital game telemetry data has opened up new experimental opportunities for studying human behavior at large scales. Prior research has demonstrated that a relation exists between learning rates and performance. Although many factors might contribute to this correlation at least one may be the presence of innate cognitive resources, as demonstrated in recent work relating IQ and performance in a Multi-player Online Battle Arena game. Here, we extend this work by examining the relationship between early learning rate and long term performance using a 400,000 player longitudinal dataset generated by new players of the widely-played MOBA *League of Legends*. We observed that the learning rate of new players in a competitive season explains a significant amount of variance in the performance at the end of the year. This analysis was then extended by training two multivariate classifiers (Logistic Regression, Random Forest) for predicting players who by the end of the season would be considered masters (top 0.05%), based on their performance in the first 10 matches of the same season. Both classifiers performed similarly (ROC AUC 0.888 for Logistic Regression, 0.878 for Random Forest), extending the time frame for skill prediction in games based on a relatively sparse sample of early data. We discuss the implications for these findings based on preexisting psychological studies of learning and intelligence, and close with challenges and direction for future research.

Index Terms—skill learning, MOBA, prediction

I. INTRODUCTION

Digital games generate considerable amounts of behavioral telemetry data [1]. However, despite the ability to track player behavior in detail within games, build detailed behavioral profiles [2] and even predict player behavior [3], behavioral analytics continue to struggle with explaining observed behaviors [4], [5]. Similarly, while work focusing on player psychology has existed for some time [6], purely psychological studies based on large-scale telemetry data are rare, in part due to difficulties in acquiring and parsing high quality, well-controlled data. In the domain of games research, there has been work exploring the correlations between player behavioral data and motivations for play [7], [8], albeit at smaller scales. Our work primarily follows recent research on skill learning and cognition in games, [9], [10], [11], [12], [13] using larger-scale data (thousands of players and upwards).

The relationship between player behavior and psychology is an ongoing research topic, with uses including designing games that are adaptive to player responses and better information modelling for AI agents [14]. Psychological research also benefits from the large datasets provided by game telemetry which increase statistical power and provide the ability to follow skill learning in individual subjects over long periods of time [9].

A growing body of evidence exists for common cognitive factors underlying early skill learning and late-stage performance, and significant achievements have been made with

prediction modelling based on smaller or larger scale video game data [9], [10], [11], [12]. Here the focus is on the application of this knowledge to inform classification models predicting future performance, based on data of the span of a whole season.

II. CONTRIBUTION

The work presented here contributes to the understanding of the relationship between player performance and skill learning, extending previous research on this topic. Previous research has established a correlation between skill learning and game-play behavior[10] as well as various cognitive and motivational factors [15], [12]. However, these results have typically been obtained for games designed specifically for the purpose of education or research. In this paper we explore the relationship between early learning rates and player performance after one year by analyzing datasets from more than 400,000 players of the popular commercial game *League of Legends* players during the 2016 season. All player-registered-accounts were new to the game, and were sampled randomly from a total player base of more than 100 million monthly active players as of 2016 [16]. Using this dataset, we find a strong relationship between early skill learning rates and final performance in a large-scale commercial online multiplayer game and we also present results of preliminary prediction model building. Based on previous work, we consider the possibility that this relationship is mediated by common cognitive factors [13], and propose future work to test this theory.

III. RELATED WORK

The relationship between cognitive skills and digital game-play has been partially explored in previous work, but most research has focused on snapshot data, examining correlations between psychological factors and performance at a single point in time [17]. In case of longitudinal studies, the focus is often on retrieving behavior in purpose-built non-commercial games over a limited period of time. An interesting question is therefore whether cognitive resources influence skill learning and gaming performance in commercial games where the player has full control on the frequency and duration of gameplay. Recent work [15], [10], has circumvented this problem making use of behavioral telemetries directly obtained from game servers, thus overcoming the common issue of needing to reconstruct the acquisition process of early skill learning.

Some psychological research in skill learning has provided preliminary evidence of a correlation between player cognition and game performance, similar to other domains [13], [10], [15] (see [15] for a recent review).

In general, digital games might be an exceptional tool for studying skill learning because players can be followed and assessed from their first contact with the game. A notable example is provided by Stafford *et al.* [9], who analyzed data from 854,064 players, from an online game and established a relationship between practice volume, spacing, variability and outcome performance. However, this work was not designed

as a long-term longitudinal study and made use of a relatively simple game designed specifically for the controlled purposes of that experiment. Following up on this work, a second study examined the time-series data of 20,000 players from the commercial online game *Destiny*, investigating factors that contribute to skill acquisition and learning rate [10]. Games have also been used for similar purposes by Thompson *et al.* [11], [12]. In related work, Kokkinakis *et al.* [13] provided evidence for a correlation between player performance and IQ. This work was snapshot based, i.e. based on a specific instant in time. Given the assumption that these correlations operate across any point in the learning curve of a player, the work of Kokkinakis *et al.* [13] and others, e.g. Bonny *et al.*[15] are the basis for investigating cognitive factors underlying skill learning at large scales in online games.

Related to the work investigating skill learning in games is the attempt in esports analytics focusing on predicting the outcome of player performance, either within or between matches [18], [19], [20], [21], [22]. The majority of this work is focused on match prediction, i.e. predicting the outcome of specific matches. An example is provided by [22], who developed match win prediction models for professional-level matches in the Multi-Player Online Battle Arena (MOBA) game DOTA 2, comparing mixed-rank and professional-only rank data in terms of their applicability to a professional-level real-time prediction system. The classifier used was a hyper-parameters-tuned Random Forest model which employed a variety of in-game behavioral features as well as higher level metadata such as hero character combinations. This type of telemetry has also been used to investigate patterns of fights that occur across professional DotA 2 games [23]. Random Forest is a commonly applied model in this body of work, similar to prediction modeling work in general game analytics e.g. [24], [25].

The only current work focusing on longer-term player performance prediction in esports is an unpublished report [26] focusing on *League Of Legends*, presenting a skill prediction model for the game. However, the work is preliminary, based on 500 matches which limits the generalizability of the results. Furthermore, the work does not address the challenge of predicting the peak skill of players based on very early performance. Work such as Bonny *et al.* [15] and Kokkinakis *et al.* [13] explore longitudinal relationships between skill and behavior (or training), but do not attempt to provide prediction models.

In summary, previous work has established the foundation for behavioral prediction in games across a growing number of types and genres (e.g. [1], [21], [27], [23]), including in the MOBA genre of *League of Legends*, which is the case study used here, and similar esports titles. Work such as Stafford *et al.* [9], [10] and Thompson *et al.* [11], [12] has provided a tentative basis for exploring skill development in digital games, establishing correlations between skill learning and behavior, as well as cognition (e.g. IQ) and skill. Here we expand on this foundation.

IV. League of Legends

League of Legends is a Multi-player Online Battle Arena game (MOBA) developed by Riot Games, published in 2009. It is the most popular esports game in the world [16], [28]. The game is supported financially by microtransactions [3], [25], [24].

The game is set in an arena environment where ten players ('summoners') control 'champions', or characters forming two teams of five players. These teams compete against one another to eliminate the opposing team's home base in the arena. Each match lasts approximately half an hour - although much shorter and longer matches are possible. Champions can gain more abilities during the game - primarily by accumulating 'experience points' (XP) or 'Gold' which can be used to buy performance-enhancing items. The change in Gold and XP as a function of time are two of the most commonly used performance metrics in *League of Legends*. Other important metrics in include the number of opponents that a player has killed, the number of deaths that a player has experienced (champions are 're-animated' after a time-out increasing in accordance to the champion's XP) and the number of 'assists' that one player has provided to another shortly before an enemy's death. Jointly, these metrics vary not only as a function of the player's skill and the overall skill of the two teams, but also depend on the specific champion played and the combat strategy employed. In general, *League of Legends*, similar to other MOBAs such as *DOTA 2* or *Heroes of the Storm* is conceptually simple but hard to master due to the complexity of the underlying gameplay [28].

A. *Skill and ranking in* League of Legends

Performance in *League of Legends* is calculated using an ELO-based relative skill rating system originally devised for chess. It is similar to other multi-player online games such as *Destiny* [10], using a generalization of ELO called TrueSkill, a Bayesian skill rating system developed by Microsoft [29]. The system is based on wins and losses and serves the function of matchmaking, provides information to players about their rank compared to others, and can be used as a qualification for tournaments. The ELO system used by Riot is specifically adapted for the 5v5 format used in League of Legends.

Players are divided into different ranks depending on their overall skill in *League of Legends*. There are seven tiers, with the top tier being limited to 200 players. The Master rank is limited to about 0.05% of the population.

Performance is formally recorded as a hidden value Match Making Rating or Ratio (MMR). The MMR of a player is not the same thing as the rank of the player which is determined by a bin and can be influenced by additional external factors such as long periods of inactivity.

V. Data Set

A. *Sample*

The dataset was provided by Riot Games, the developer and publisher of *League of Legends*. The dataset contains behavioral telemetry data derived from the 2016 season of the game (ranked play in *League of Legends* is organized in game seasons of roughly one year). From the global player base, a random sample was drawn, covering 413,341 users (players) and approximately 140 million rows.

Each row in the dataset contained records for a single match in relation to a particular player account. The earliest match entry was recorded on 21 January 2016 and the final match played in the data was logged on 6 November 2016. This period of time falls under the 2016 Competitive Season of the game. All accounts had been created at the start of the season and played a minimum of 150 competitive ladder games during the season. 150 matches was set as a lower bounds to remove any largely inactive players. All matches in the data were restricted to the default 5 versus 5 ranked "Solo/Duo Queue" ranked mode. All player MMRs were initialized to the same starting value by default. After every match, this rating was then updated based on a system that takes into account the average rating of a player's team, an average rating of the enemy team, whether the player's team won or lost. Winning a match resulted in an increased rating, and a loss results in a decreased rating.

B. *Telemetry*

Data from League of Legends are publicly available via a data API service provided by Riot. The dataset provided here was similar in to the data that might be acquired from these public sources but had the advantage of being randomly drawn from the population, and, critically, contained Rating (MMR) scores for each player, which are not publicly available. 16 features were provided for this analysis (Table I).

TABLE I
RAW IN-GAME DATA, ON A PER MATCH BASIS

Field	Description
Account ID	Unique Identifier of a player account
Platform ID	The server the game was played on
Game ID	Unique identifier of the match
Neutral Creep	Number of neutral AI enemies killed
Enemy Creep	Number of AI enemies killed
Win	Boolean indicating a win or loss
Timestamp	When the match was logged
Date	Date of match played
Hour	Hour of match played
Gold Earned	Total gold earned in the match
Damage Dealt	Total dealt to other players
Time Dead	Total seconds spent dead
Time Played	Total seconds played in the match
Kills	Total Kills
Deaths	Total Deaths
Assists	Total Assists
Rating	The rating of the player before the match
Position	The role the player was assigned

The raw data forms the basis for feature engineering (see also Table II).

VI. Methods

Data were preprocessed and analyzed in a Python 3.6 environment using Pandas [30], Numpy and SciPy [31] for data handling and statistical analysis. Scikit-learn [32] was the reference framework for machine learning.
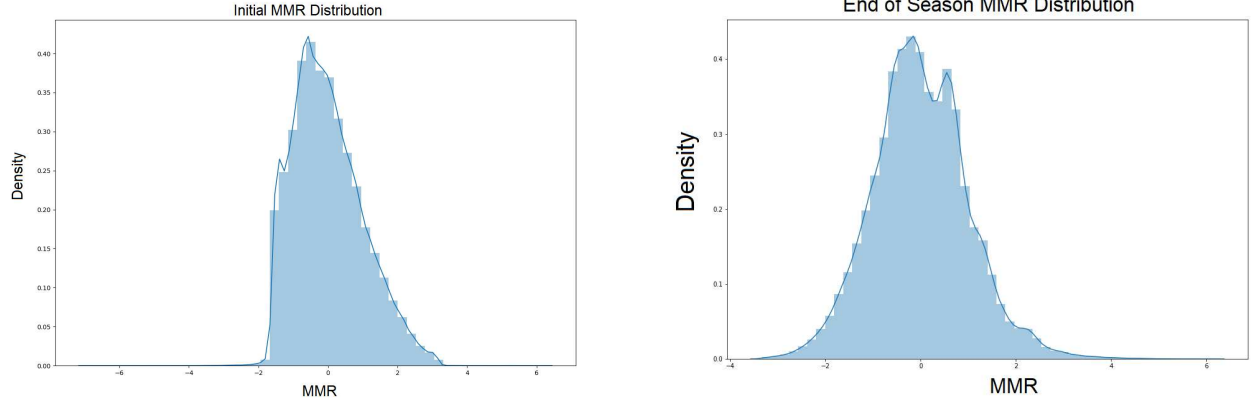
Fig. 1. Distributions of players by MMR during calibration games (left) and after the season end (right). Due to the confidential nature of the MMR values, the axes have been standardized.
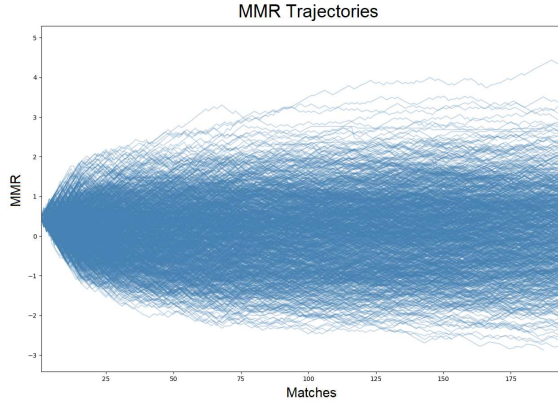

Fig. 2. Trajectories of MMRs over matches from a subsample of players. All the trajectories stem from the same starting point and spread in the initial stages mirroring a power law curve. Due to the confidential nature of the MMR values, the axes have been standardized.

A. Data Preprocessing

The data provided by Riot were drawn directly from the telemetry servers of League of Legends. In addition to imposing a 150 game minimum requirement, and season time bin we performed some additional pre-processing steps to ensure data quality.

We first eliminated players playing more than 3,000 matches in total, as we were cautious of excessive playtimes indicating possible contamination from shared accounts or automated systems. Following this, we also filtered players whose first MMR entry differed from the pre-defined starting value. This was determined to be an artifact caused by players migrating between different servers during the season, displaying only records of their play on their latest server in our data. We then eliminated player who abandoned (went 'away from keyboard' early in the game) during the first 10 games they played because scores during this period were a critical component

of our analysis. To eliminate these players we filtered users that recorded Time Played durations of less than 900 seconds. Despite a legitimate match in LoL might be shorter, these extremely rare cases hold very little information about the player performance and are indistinguishable from those where the player decided to leave the game. We also discarded users who recorded simultaneous Kills, Deaths and Creep Kill scores of 0 for the same reason. Finally we also excluded users having their nominally unique id duplicated on multiple servers.

From the original data of 413,341 players, 313,184 were retained after preprocessing. Standardized distributions of MMR for this sample can be observed in figure 1, and the trajectories of MMRs over the season can be observed in figure 2.

B. Regression Analysis

For each region we evaluated whether the rate of change in the MMR of the first 10 matches predicted the mean MMR of the last 10 matches. For comparison, we generated a synthetic null data set by computing 100,000 random walks with length and MMR transition probabilities drawn at random from distributions matching the existing data.

C. Feature Engineering

Since the aim of this work was to evaluate the impact of early season performance on final season outcome we computed a set of features based on the original Key Performance Indices (KPIs) over the first 10 matches of each user (again see Table 1 for further details). Two approaches were adopted: a brute force one where we retrieved various statistical descriptors of the original KPIs and an informed one where we used knowledge derived from our regression analysis and previous work [10], [27], [24], [25], [12] for retrieving possible useful features.

In first instance a series of temporal KPIs were created based on the in-game time alive (i.e. Time Played - Time Dead): Neutral Creep per Minute, Enemy Creep per Minute, Gold per Minute, Damage per Minute, Kills per Minute, Deaths per Minute and Assists per Minute. For each of the

original and temporal KPIs we retrieved mean, median and standard deviation over the first 10 matches in accordance to the methodology found in [33]. Following the intuition of [10] we computed a series of progression metrics retrieving the first derivative obtained by regressing a particular KPI over the ordered number of matches (i.e. range from 1 to 10). We calculated the first derivative for: Gold, Damage Dealt, Time Alive, Time Dead, Kills, Deaths, Assists, Gold per Minute, Deaths per Minute, Assists per Minute and MMR over the first 10 games.

To avoid problems of instability when calculating ratios with a denominator close to zero we computed the percentage for the following sets: {Time Alive, Time Dead}, {Deaths, Kills, Assists}, {Win, Loss}, {Morning Session, Afternoon Sessions, Evening Sessions, Night Sessions}, and {Position Utility, Position Middle, Position Bottom, Position Top, Position Jungle}.

We also calculated a series of miscellaneous features like Mean Temporal Distance between matches, number of Consecutive Wins, number of Consecutive Losses and variability in the role assumed by the player as measured by the Gini Index. These were added because previous work has utilized these metrics e.g. for prediction modeling in games or to explore skill learning (e.g. [10], [27], [24], [25], [12]

As target variables for our regression analysis and multivariate classification task we retrieved for each user the mean MMR over the last 10 matches and the difference between the mean MMR of first and last 10 matches. A summary of features generated can be observed in Table II.

TABLE II
SUMMARY OF THE FEATURES ENGINEERED FROM ORIGINAL RAW DATA.

Feature Type	Description	Example
Statistics	Common statistical descriptors	Mean kills Median kills Standard deviation Kills
Progression	Rate of change over matches	First derivative of kills over first 10 matches
Percentages	Percentages over particular sets of raw data	Percentage of Kills Percentage of Deaths Percentage of Assists
Miscellaneous	Features covering specific aspects of the game	Role variability Mean temporal distance between matches
Targets	Metric employed for regression and classification	Mean MMR over the last 10 matches (Mean MMR last 10 matches) - (Mean MMR first 10 matches)

D. Predictive Skill Modeling

Despite an interesting goal would have been forecasting the players' performance in continuous fashion (i.e. regression), we decided to focus on an early detection of extremely proficient players (i.e. classification)[34]. This solution allowed us both to maximize the prediction power and to provide a starting point for addressing issues relevant for the competitive games industry (i.e. player scouting). For our purpose we used two common machine learning algorithms,

Logistic Regression (LogReg) and Random Forest (RanFor) [35], able to capture both linear and non linear interactions between the features. We chose these algorithms because despite their simplicity, they can often achieve good result while still providing useful insights (i.e. visualization of features importance). Furthermore, these are models heavily used in game analytics research for prediction tasks (see e.g. [27], [18], [24]).

In first instance we created a labeling system for differentiating the best performing users from the rest of our sample. The metric employed for this labeling system was the difference between the mean MMR of the first 10 matches and the mean MMR of the last 10 matches, this has been done for avoiding that informations contained in the input features (derived from the first 10 matches' KPIs) leaked in the metric employed for creating the labels consequently biasing the classification model. Nevertheless, for transparency reasons, we also conducted the same classification task employing labels derived from the mean MMR of the last 10 matches but due to space constrains the relative results are reported exclusively in table IV. The labeling system employed a percentile based encoding where all the players below the 99.95 percentile were encoded as negative samples while all the others as positive. We then divided the original data-frame in validation (n = 209,834) and test set (n = 103,350) via Stratified Shuffle Split [32]. This was essential given the extreme imbalance in the label distribution. We used the validation set for searching for optimal hyperparameters and the test set for performing the final prediction. For each model the best combination of hyper-parameters was found by using a Grid Search 10 Fold Stratified Shuffle Cross Validation and selecting the best model based on the average ROC AUC score. Since the labels distribution was extremely imbalanced for avoiding under or over-sampling our dataset we applied a weight to each label inversely proportional to its frequency in the input data [32].

To improve the performance of the Logistic Regression and allowing the interpretation of the coefficients associated to each feature, when using this model we rescaled the features using a method that is robust to outliers (i.e. removing the median and rescaling the data accordingly to the quantile range). After tuning the hyper-parameters to discover the best model we retrieved the top 20 features contributing the most to the classification performance, although this can provide insights, given the high inter-correlation between our features caution has to be posed in the interpretation of their importance.

VII. RESULTS

A. Regression Analysis

The learning rate computed from the first 10 games was correlated significantly with the final average performance level (fig. 3). This correlation achieved significance across all servers with p values less than .0001 in all case. Effect sizes (r2) ranged from .25 to .37. Performance improved with the number of initial games chosen with an approximately linear

dependence up to 40 games. As expected, our randomized control dataset using a large set of simulated players (n=100,000) also exhibited a statistically significant relationship between slope and final score (p<.0001) reflecting the fact that a slight positive slope in the initial stages of a random walk will tend, on average, to result in a slightly positive final value. However, in this case the effect size was very small ($r2 = .008$). Similar results were observed in each server independently (Table III).
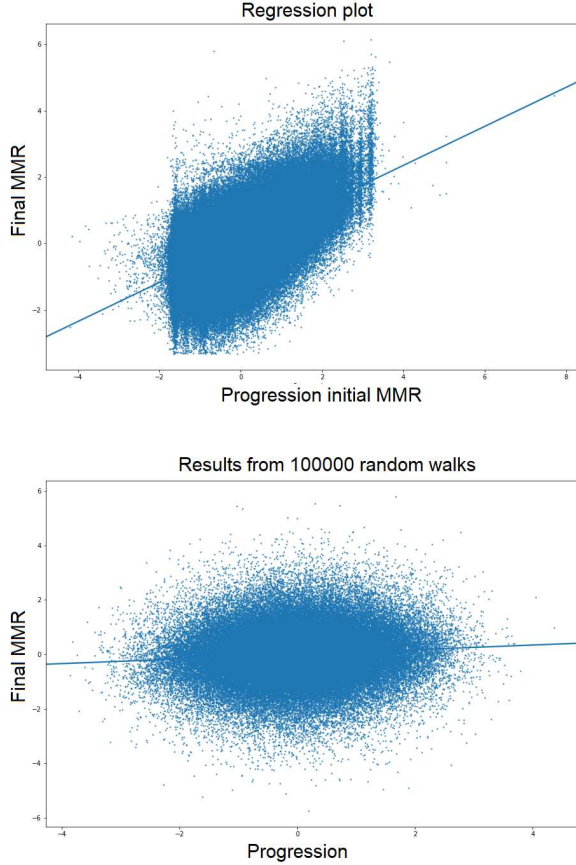

Fig. 3. Regression plots for total sample learning rate and final MMR (top), and random walk synthetic set (bottom). Due to the confidential nature of the MMR values, the axes have been standardized.

TABLE III
REGRESSION RESULTS

Set	r2	p value
North Europe	0.308	<.0001
West Europe	0.378	<.0001
Brazil	0.297	<.0001
Latin America 1	0.257	<.0001
Latin America 2	0.307	<.0001
Oceania	0.353	<.0001
North America	0.374	<.0001
Japan	0.315	<.0001
Total sample	**0.345**	**<.0001**
Random Walk	**0.008**	**<.0001**

B. Predictive Skill Modelling

The best hyper-parameters found by the grid search for the Logistic Regression were L1 penalty with inverse regularization equal to 0.01 while those for the Random Forest included entropy as a split evaluation metric, maximum depth of the tree equal to 10, maximum number of features employed by each tree equal to the square root of the total number of features, maximum number of leaf nodes equal to 15 and number of trees populating the forest equal to 60. As mentioned before, we only took into account the results derived from the adoption of the difference based labeling system, however, for visibility purposes, in table IV we also reported the results from the alternative labeling system. The fields in Table IV specify the model employed, the metric on which the labeling system is based, the weighted f1 score (i.e. accounting for imbalance in the labels distribution), the ROC AUC score, the number of true positive, true negative, false positives and false negatives. For a better overview of the models' performances we computed and plotted normalized confusion matrices showing the percentages of correct and incorrect classifications (fig. 4) as well as bar charts showing the top 20 features contributing the most in the classification task (fig. 5).

TABLE IV
RESULTS OF PREDICTIVE ANALYSES

Model	Metric	f1	AUC	TN	FN	TP	FP
LogReg	**Diff**	**0.940**	**0.888**	**92,077**	**61**	**456**	**10,757**
RanFor	**Diff**	**0.938**	**0.878**	**91,780**	**70**	**447**	**11,054**
LogReg	Final	0.953	0.923	94,534	37	480	8,300
RanFor	Final	0.952	0.923	94,309	36	481	8,525

Results of Logistic Regression and Random Forest for prediction using either end of season MMR (Final) or MMR change (Diff). AUC: area under ROC, TN: true negative, FN: false negative, TP: true positive, FP: false positive

VIII. DISCUSSION

We find that the initial rate of MMR change is strongly related to the final end-of-season MMR in League of Legends. This suggests that a common factor which we identify as cognitive performance underlies learning and performance in this game - and almost certainly in other similar MOBAs. Our results build on the finding of by Dewar and Stafford [9], extending them in several ways. Dewar and Stafford's data were obtained from users playing a non-commercial on-line game specifically designed for educational and research purposes. In this respect their findings mirror those of Quiroga *et al.* [36] who used custom-made game-like tests to probe IQ. In comparison, the game we analyze here is a commercial product with an active user-base that numbers in the hundreds of millions. The statistical findings we present are therefore extremely robust due to the sample size used, and of general interest because of their ecological relevance. We also presented promising results from a multivariate classification task showing that end of season exceptional performance can be identified employing metrics derived from the early matches.
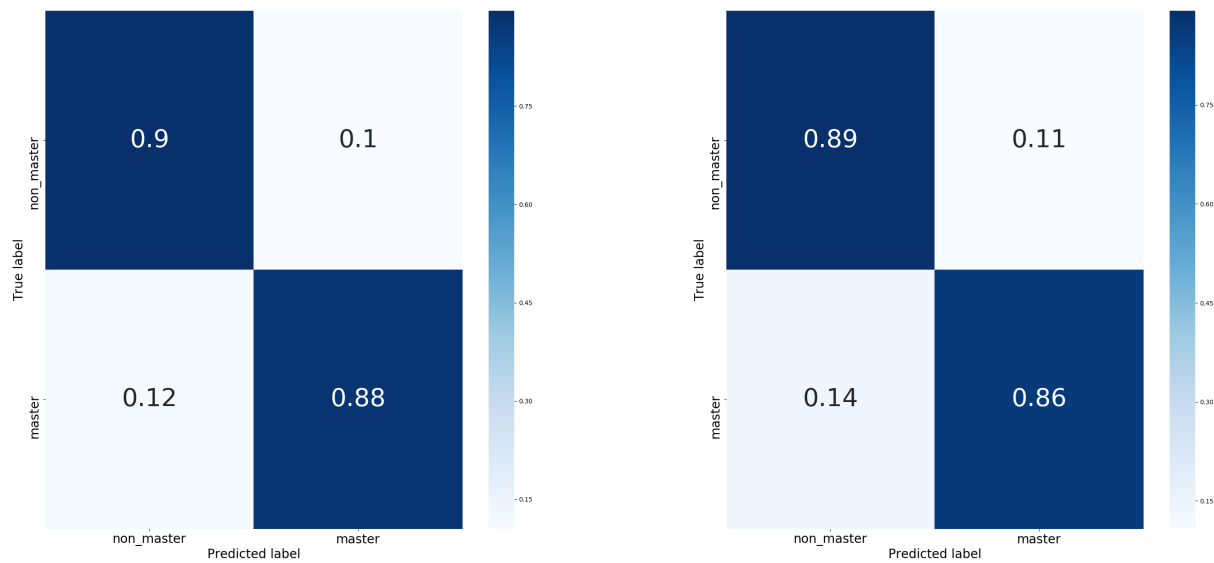
Fig. 4. Normalized Confusion Matrices for logistic regression difference based labeling system (top) and random forest difference based labeling system (bottom).
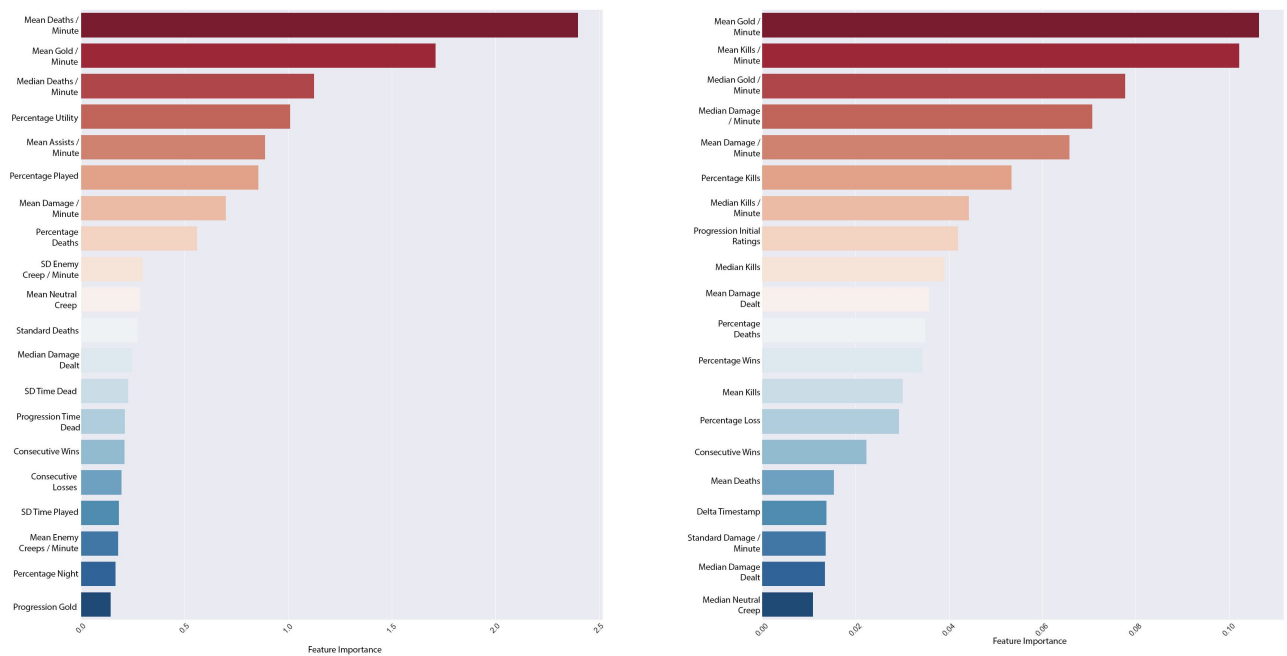

Fig. 5. Feature importance: 20 most important features for logistic regression (left) and random forest classifier (right). The two models identify different sets of features as the most important predictors. However, the top five features for both models all deal with player deaths, gold gain and player kills and damage dealt. Logistic regression adds in the percentage of time spent playing utility roles also. Looking at the top 20 predictors, there is some difference between the two models but both include similar feature sets: kills, deaths, damage and gold. Notably, win and loss conditions feature relatively low on the features ranking ("consecutive wins" placed 13th for both models). This indicates that the win/loss features are perhaps too aggregate (i.e. encapsulating performance of both teams in the game) to be highly significant predictors of individual skill/performance.

Our results support a growing body of work indicating that cognitive performance ([13], [10], [15] and possibly other psychological factors [7], [8]) are exposed by on-line game telemetry. This observation can be used in at least two ways: 1) it has significant interest to the psychology community because it provides a way of evaluating cognition at the population level in real-time and at a global scale. In previous work [13], we have raised the possibility of large-scale video game data being used to perform 'cognitive epidemiology' - a population-level assessment of cognitive health which might provide an early indicator of environmental changes (for example, disease, pollution or social factors) that affect cognition. 2) It is of interest to the e-sports analytics community because it provides a theoretical basis for performing longitudinal game analytics - allowing analysts to predict, for instance, churn rate, future performance levels and, potentially, complex player-player interactions.

Our future work is focused on exploring the link between video game data and psychological factors still further. While cognitive performance is important, it is just one of a wide range of psychological factors that can be extracted from these rich datasets. We expect that these factors will, like the one studied here, provide important insights into psychology at a global level while also providing the games industry with theoretically validated tools to improve their products and user experience.

IX. LIMITATIONS AND FUTURE WORKS

We acknowledge that focusing our work on a single MOBA title might pose limitations to the generalizability of the results. These types of analysis are complicated by the requirement to gain access to raw ELO scores for these game. This often requires licencing agreements with the companies that are not straightforward to obtain. Nevertheless, one possible direction for future work would be to attempt to replicate the results presented here employing data from different games.

X. DECLARATION OF CONFLICTING INTERESTS

The authors declare that they have no conflicting interests.

XI. ACKNOWLEDGEMENTS

REFERENCES

[1] M. Seif El-Nasr, A. Drachen, and A. Canossa, *Game Analytics - Maximizing the Value of Player Data*. Springer, 2013.

[2] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau, "Guns, swords and data: Clustering of player behavior in computer games in the wild," in *IEEE Conference on Computational Intelligence and Games*. IEEE, 2012, pp. 163–170.

[3] M. Milosevic, N. Zivic, and I. Andjelkovic, "Early churn prediction with personalized targeting in mobile social games," *Expert Systems with Applications*, 2017.

[4] A. Drachen and S. Connor, "Game analytics for games user research," in *Games User Research*. Oxford University Press, 2018, pp. 333–354.

[5] A. Drachen, J. Green, C. Gray, E. Harik, P. Lu, R. Sifa, and D. Klabjan, "Guns and guardians: Comparative cluster analysis and behavioral profiling in Destiny," in *IEEE Conference on Computational Intelligence and Games*. IEEE, 2016, pp. 1–8.

[6] N. Yee, "The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments," *Presence: Teleoper. Virtual Environ.*, vol. 15, no. 3, pp. 309–329, 2006.

[7] A. Canossa, J. B. Martinez, and J. Togelius, "Give me a reason to dig minecraft and psychology of motivation," in *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*. IEEE, 2013, pp. 1–8.

[8] G. Van Lankveld, P. Spronck, J. Van den Herik, and A. Arntz, "Games as personality profiling tools," in *Computational Intelligence and Games (CIG), 2011 IEEE Conference on*. IEEE, 2011, pp. 197–202.

[9] T. Stafford and M. Dewar, "Tracing the Trajectory of Skill Learning With a Very Large Sample of Online Game Players," *Psychological Science*, vol. 25, no. 2, pp. 511–518, Feb. 2014. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0956797613511466

[10] T. Stafford, S. Devlin, R. Sifa, and A. Drachen, "Exploration and Skill Acquisition in a Major Online Game," in *The 39th Annual Meeting of the Cognitive Science Society (CogSci)*. York, 2017.

[11] J. J. Thompson, M. R. Blair, L. Chen, and A. J. Henry, "Video game telemetry as a critical tool in the study of complex skill learning," *PloS one*, vol. 8, no. 9, 2013.

[12] J. J. Thompson, C. M. McColeman, E. R. Stepanova, and M. R. Blair, "Using video game telemetry data to research motor chunking, action latencies, and complex cognitive-motor skill learning," *Topics in Cognitive Science.*, vol. 9, no. 2, pp. 467–484, 2017.

[13] A. V. Kokkinakis, P. I. Cowling, A. Drachen, and A. R. Wade, "Exploring the relationship between video game expertise and fluid intelligence," *PLOS ONE*, vol. 12, no. 11, p. e0186621, Nov. 2017. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0186621

[14] G. N. Yannakakis, "Game ai revisited," in *Proceedings of the 9th Conference on Computing Frontiers*. IEEE, 2012, pp. 285–292.

[15] J. W. Bonny and L. M. Castaneda, "Number processing ability is connected to longitudinal changes in multiplayer online battle arena skill," *Computers in Human Behavior*, vol. 66, pp. 377–387, 2017.

[16] "Riot games reveals league of legends has 100 million monthly players," *Forbes*, 2016. [Online]. Available: https://www.forbes.com

[17] A. D. Castel, J. Pratt, and E. Drummond, "The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search," *Acta Psychologica*, vol. 119, pp. 217–230, 200.

[18] J. Kawale, A. Pal, and J. Srivastava, "Churn prediction in MMORPGs: A social influence based approach," in *International Conference on Computational Science and Engineering*, vol. 4. IEEE, 2009, pp. 423–428.

[19] H. Xie, S. Devlin, D. Kudenko, and P. Cowling, "Predicting Player Disengagement and First Purchase with Event-frequency Based Data Representation," in *Proc. of CIG*, 2015.

[20] M. Schubert, A. Drachen, and T. Mahlmann, "Esports analytics through encounter detection," in *Proceedings of the MIT Sloan Sports Analytics Conference*, 2016.

[21] F. Rioult, J.-P. Metivier, B. Helleu, N. Scelles, and C. Durand, "Mining tracks of competitive video games," *AASRI Procedia*, vol. 8, pp. 82–87, 2014.

[22] V. Hodge, S. Devlin, N. Sephton, F. Block, A. Drachen, and P. Cowling, "Win Prediction in Esports: Mixed-Rank Match Prediction in Multiplayer Online Battle Arena Games," *arXiv preprint arXiv:1711.06498*, 2017.

[23] P. Yang, B. E. Harrison, and D. L. Roberts, "Identifying patterns in combat that are predictive of success in MOBA games." in *FDG*, 2014.

[24] J. Runge, P. Gao, F. Garcin, and B. Faltings, "Churn Prediction for High-value Players in Casual Social Games," in *Proc. of IEEE CIG*, 2014.

[25] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauck-hage, "Predicting player churn in the wild," in *Computational intelligence and games (CIG), 2014 IEEE conference on*. IEEE, 2014, pp. 1–8.

[26] J. Min, S. W. Jang, and J. Ha, "Predicting Matchmaking Rating (MMR) Change in League of Legends."

[27] R. Sifa, F. Hadiji, J. Runge, A. Drachen, K. Kersting, and C. Bauckhage, "Predicting purchase decisions in mobile free-to-play games," in *Proc. Artificial Intelligence and Interactive Digital Entertainment International Conference*, 2015, pp. 79–85.

[28] Y. J. Kim, D. Engel, A. W. Woolley, J. Y.-T. Lin, N. McArthur, and T. W. Malone, "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." ACM Press, 2017, pp. 2316–2329. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2998181.2998185

[29] "Trueskill: A bayesian skill rating system."

[30] W. McKinney, "pandas: a foundational python library for data analysis and statistics," *Python for High Performance and Scientific Computing*, pp. 1–9, 2011.

[31] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–, [Online; accessed ¡today¿]. [Online]. Available: http://www.scipy.org/

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and . Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011. [Online]. Available: http://www.jmlr.org/papers/v12/pedregosa11a.html

[33] 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.

[34] "Rank distribution: League of legends," *League of Graphs*, 2008. [Online]. Available: https://www.leagueofgraphs.com/rankings/rank-distribution

[35] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[36] M. A. Quiroga, S. Escorial, F. J. Romn, D. Morillo, A. Jarabo, J. Privado, M. Hernndez, B. Gallego, and R. Colom, "Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can!" *Intelligence*, vol. 53, pp. 1–7, Nov. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016028961500104X