Mastering the Art of War: How Patterns of Gameplay Influence Skill in Halo

Jeff Huang
Information School
University of Washington
chi2013@jeffhuang.com

Thomas Zimmermann
Nachiappan Nagappan
Microsoft Research

Charles Harrison
Bruce Phillips
Microsoft Games Studios

{tzimmer,nachin,charlesh,bphil}@microsoft.com

ABSTRACT
How do video game skills develop, and what sets the top players apart? We study this question of skill through a rating generated from repeated multiplayer matches called TrueSkill. Using these ratings from 7 months of games from over 3 million players, we look at how play intensity, breaks in play, skill change over time, and other games affect skill. These analyzed factors are then combined to model future skill and games played; the results show that skill change in early matches is a useful metric for modeling future skill, while play intensity explains eventual games played. The best players in the 7-month period, who we call "Master Blasters", have varied skill patterns that often run counter to the trends we see for typical players. The data analysis is supplemented with a 70 person survey to explore how players' self-perceptions compare to the gameplay data; most survey responses align well with the data and provide insight into player beliefs and motivation. Finally, we wrap up with a discussion about hiding skill information from players, and implications for game designers.

Author Keywords
video games; learning; gaming skill; online multiplayer.

ACM Classification Keywords
K.8.0 Personal Computing: General—*Games*

INTRODUCTION
Competitive games are a thriving industry, played by millions, and watched by millions. Tournaments and televised game matches are popular among those skilled enough, and spectators enjoy watching the best players compete in high-stakes tournaments [3]. But the skilled were once unskilled; the best players get there with dedicated practice and determination. So we ask, how do skills in a game develop, and what differentiates the top players? This leads to interesting questions about patterns of play and its relationship with skill. To explore these questions, we use a well established metric of skill from a current popular video game, Halo Reach.

CHI 2013, April 27–May 2, 2013, Paris, France.

Millions of people play Halo Reach, a first-person shooter on the Xbox game console. From records of their online matches, we study their TrueSkill rating [8], a Bayesian scoring system similar to the Elo rating in chess. This rating provides us with a proven metric for representing skill in scenarios where people are actively trying to improve and outmatch their opponents. Players play online matches over and over, providing a controlled set of repeated trials that can give us insight into playing as a form of practice. Aggregated, this amounts to a large set of objective skill measurements over a 7 month period.

Using this data, we look at what factors affect players' skills and their learning ability, a topic studied across many domains. Having a large dataset of complete game histories from over 3 million players allows us to measure and analyze skill from many dimensions and compare skill gain in Halo with experiments from sports training and learning sciences. For example, people who lift weights at the gym are keenly aware that inactivity (i.e., detraining) reverses muscle gains, but restarting a training regime after inactivity replenishes strength quicker than previously untrained people [16]. We see similar effects of playing breaks occurring in Halo Reach.

To study these factors from multiple perspectives, we take a mixed methods approach, like Halo researchers before us [14, 19]. A quantitative data analysis finds robust trends in a large-scale 3 million player dataset, and allows us to also examine the best players in the cohort we study; surveys explain the observed behavior we see in the quantitative analysis as well as provide insight into players' self-perceptions in learning. Put together, we go beyond identifying *how* and *what* players are doing in the game, to understand *why* players act in a particular way.

We contribute an exploration of skill progression and factors affecting skill. The factors we study are diverse, including: play patterns, demographics, skill change in early matches, and other game titles played. Our secondary contribution is providing implications for the game designers and game system on dealing with these factors' effect on skill, and a discussion of whether players should be provided with skill information.

RELATED WORK

Data Analysis in Games
Our study uses a similar methodology as two other studies of multiplayer Halo, both of which use gameplay records and either player surveys or interviews. Mason

and Clauset [14] use the same data source as us—the Halo Reach multiplayer records, and supplement them with a survey. They find that players with more friends on their team perform better individually, while also performing better as a team. Xu et al. [19] take a different approach on studying social motivations, where they aim to understand the social relationships between Halo 3 players through the gameplay records and interviews. They found that players were well aware of who they played with, and rather than only playing to win, also sought to enjoy the social experience of the game. While we employ similar methods and data as these two studies, our focus is on player skill and the change in expertise rather than social relationships.

Game data can also be used for visualizations, such as Wallner and Kriglstein's implementation that overlays gameplay data from replay files over the in-game map [17]. These visualizations allow players to view a previous match at a glance, but there are also applications that benefit the game designers. By constructing player transition models, Harrison and Roberts could predict a players' next action before it happened. Their model of World of Warcraft players' achievement data used a correlation network to predict which achievement they are most likely to earn next [7].

Prior work on characterizing Project Gotham Racing 4 [10] explained the diverse and extensive amount of data that is collected due to the constantly connected nature of the game consoles. The results of this analysis helped provide a better understanding of the differences between long-term and short-term players, the choices they make, their retention and the extent to which various options in the game are utilized (in this case for example, the type of track, vehicle class, or weather conditions). This led to recommendations for ways to reduce development costs by eliminating unused or unpopular options and to help keep new players engaged.

Users' Learning and Skill from Repetition

People acquire skill from frequent use outside of games. An early and well-known study is that of American telegraphers by Bryan and Harter [1]. The sending and receiving rates, measured in characters per minute, are plotted over time to produce figures showing different rates of acquiring expertise among the two tasks, and particularly a plateau in the middle of the receiving plot. This plateau has incited discussion in follow-up work, where the original authors believed that multiple practice curves existed [2], characterized by the two separate skills of mapping Morse code into letters and predicting the word from the initial letters. However, Keller counters in a later study that there is no plateau effect [11], citing unpublished studies by Tulloss, where "there is no sign of a plateau in any of the Tulloss curves".

Other studies in software have looked at motivators for skill acquisition and differences between experts and non-experts in searching the Web. In a study of non-programmers playing a game that teaches programming, Lee and Ko found that participants completed more levels of the game, and thus acquired additional skill in programming, if the goal was framed in terms of helping a personable robot rather than an inanimate terminal [12]. In another study, White et al. examined experts and non-experts behavior over a 3 month period of search logs. They found that expert searchers differed in terms of query vocabulary, sites they visited, and patterns of search behavior [18]. The authors were also able to predict the expertise of a user with modest success; computer science experts were found to be easier to predict than medicine, finance, or legal experts. In our work, we focus less on predicting skill, and more on explaining factors that affect skill.

Case Studies of Video Game Expertise

Case studies situate the researcher inside the gaming experience, either as observers or as players themselves. Reeves et al. take an ethnomethodological approach to analyzing expertise in the first-person shooter, Counter-Strike, by watching an expert *in situ* [15]. They find that expert play constitutes an understanding of the terrain and a sense of where other players are in the environment. Reeves et al. also suggest regarding gameplay holistically, as it does not make sense when taken in pieces. Another researcher, Hock-Koon, becomes an expert himself in the game Alien vs. Predator [9]. He rigorously kept a journal of his training and lessons learned, and developed a theory of elliptical learning. Hock-Koon argues that learning encompasses multiple levels of understanding for a single mechanism in the game. In contast to these case studies that put the researcher into the game, we step back and look at aggregate data from millions of players to seek generalizable patterns.

TEAM SLAYER IN HALO REACH

Halo Reach is the latest version of the popular Halo franchise on the Xbox console, with over 9 million copies sold[1]. It is a first-person shooter, where players battle with rifles, grenades, plasma weapons, and swords. The games start with the player spawning with initial weapons somewhere on a map; additional weapons, health, and other power-ups are available elsewhere. There are both single-player and multi-player components, where the multi-player games are played on Xbox Live, on a local network, or single Xbox with split-screen.

In Team Slayer, by far the most popular multiplayer playlist (a set of game types with similar rules), teams gain a point whenever a member of their team kills an enemy. When a player dies, they respawn at a random spawn location on the map. The team with the most points at 15 minutes or the first team to reach 50 kills wins the match. Thus, each match typically takes 12–15 minutes, with about 5 minutes following the match to view post-game statistics, to assign the next teams and map, and load the next game. In this paper, we focus on studying skill in Team Slayer because of the simplicity

[1] http://www.vgchartz.com/game/35024/halo-reach/

of the game, its popularity, and its consistency of play from match to match. While half the players only play 40 or fewer matches of Team Slayer, the vast majority of the matches are from the minority of players who play hundreds of matches (Figure 1).

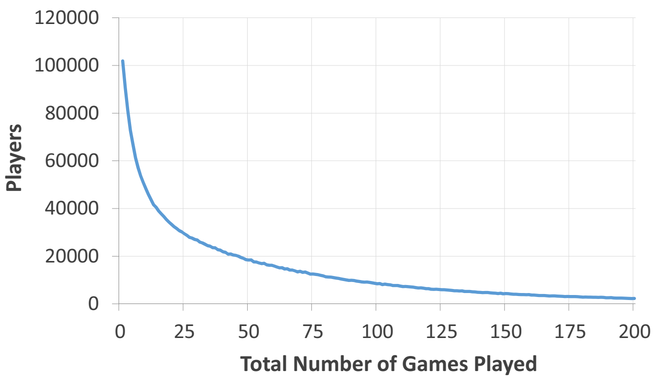


Figure 1. In Team Slayer, half the players played at least 40 matches, and a quarter played 95 matches or more. The chart extends beyond 200 matches as some people played over 1,000 matches during the 7 month period.

TrueSkill

Halo Reach uses a skill rating system called TrueSkill [8], a generalization of the Elo chess rating [6]. TrueSkill is currently used for matchmaking across numerous Xbox titles. The matchmaking system attempts to maximize the probability a match will end in a draw, which generally makes for an exciting match; of course, this is subject to practical constraints such as which players are currently looking for new matches. Halo Reach does not show players their current TrueSkill rating—a design decision we discuss later—so there is little incentive for players to manipulate this rating.

TrueSkill represents a player's skill as a Gaussian distribution, parameterized with a mean μ and standard deviation σ; μ represents the best guess of that player's skill, and σ represents the variation in that guess. The σ generally decreases over time as the player plays more matches since there is more information about their skill. μ starts with an initial value (a prior of 3), that adjusts to a player's "true" value for each multiplayer playlist. The matchmaking system attempts to pair up teams with equal skill (using a conservative estimate of skill computed by $\mu - C\sigma$), striving for balanced matches.

Player performance has been studied retroactively using TrueSkill for games of chess, showing that it can accurately predict the outcome of matches better than other rating systems [4], and in StarCraft, where it agrees with public opinion about the top players in history [5].

METHOD

Data Analysis

We use the TrueSkill "best guess" rating μ as the estimate of a user's skill. The ratings were retrieved from the official Halo servers that compute them for matchmaking. Our dataset consists of the complete first 7 months of games from the 3.2 million Halo Reach players on Xbox Live in its first week of release (September 13–20, 2012). We selected this cohort of players to control for the time when a person starts playing Halo Reach, and the remainder of this paper uses this cohort's historical game records for the 7 month period. Note that we are not sampling—this is the *complete population of players in this cohort*, and our dataset comprises every match played by that population. However, from this data we still know little about the mechanisms through which players improve their skills, which may require alternate methodological approaches (e.g., [15, 9].

Our analysis can be reproduced by other researchers who download game histories from the Halo Reach API such as Mason and Clauset [14], and the TrueSkill ratings can be recomputed using the published equations [8]. When plotting the players' skill in the charts, the median skill at every point along the x-axis was taken for each group. The median reduces the bias that occurs when plotting μ, a skewed variable that makes taking the mean exaggerate the effect of each factor.

Survey

We supplement the large-scale data analysis results with a survey of a random sample of Halo Reach players who opted-in to be asked about their player experience on the Xbox. These participants were from the United States and were rewarded with a chance to win Microsoft software. We asked a dozen open-ended and multiple-choice questions to 300 players and received 70 responses relating to players' perceptions of skill. The multiple choice questions were often followed up with an open-ended question to allow the participant to elaborate on their response, particularly why they chose an answer.

From this survey, we sought to discover how participants' self-perceptions matched the data analysis results. We wanted to know which behaviors they reported confirmed what we found in the gameplay data, and which ones disagreed with the analysis. We believe that such a mixed-method approach could compensate for some of the weaknesses inherent in both methods. For example, the survey responses could shed light on why players behaved a certain way while the gameplay data could not, but the gameplay data could show patterns for different segments of people that would be difficult to identify in a small-scale survey. However, we did not cross-reference the players in the survey with their data for their privacy.

PATTERNS OF PLAY AND LEARNING

An overarching research question we explored is how players' patterns of play affected their skill. Improving one's skill is tantamount to learning, and we wanted to look at specifically how *play intensity*, *breaks between games*, *playing other games*, and *initial skill change* related to a player's skill in Halo Reach Team Slayer. These factors were chosen during discussion among the authors as common beliefs among gamers as chief determinants of skill.

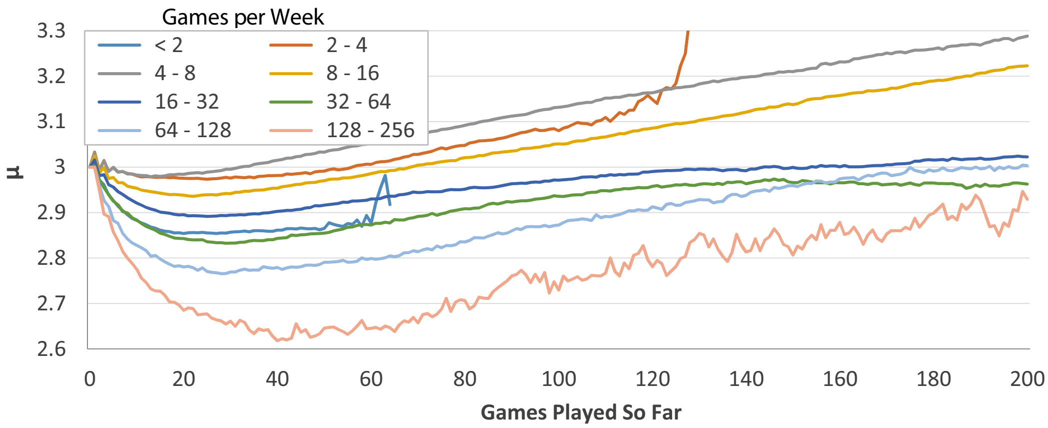


Figure 2. Players who play different numbers of matches per week gain expertise at different rates, generally trending towards higher skill for each match played.

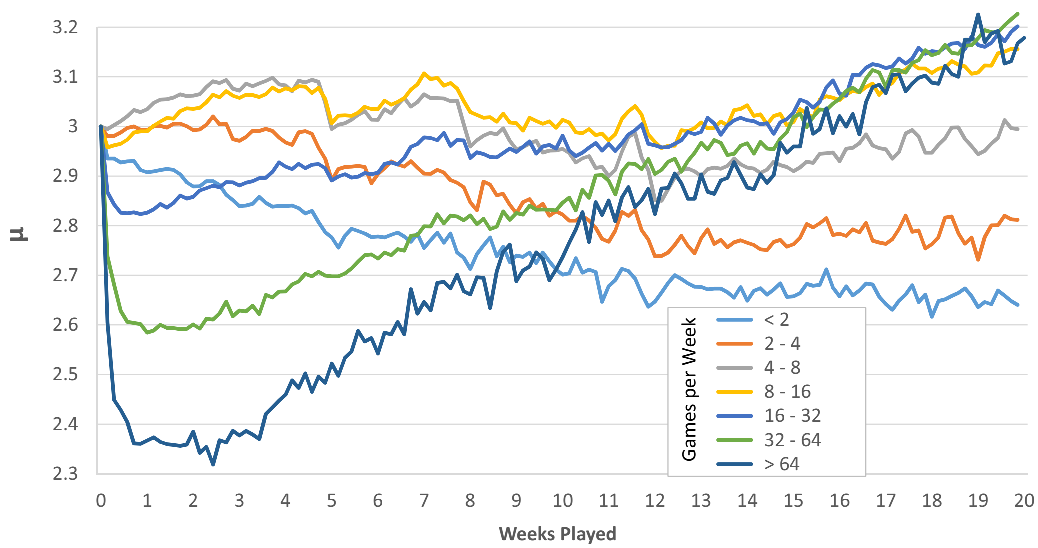


Figure 3. When looking at the skill over the number of weeks played, the more frequent players gain skill faster, even though it takes more matches to reach that level.

Play Intensity

We first investigated how skill is affected by a player's play intensity. Do players improve more if they play the same number of games spread out over more weeks or played more compactly in fewer weeks? Do those who play more games per week improve faster than those who play fewer, and is there a plateau of improvement? To explore these questions, we plot TrueSkill for different players, grouped according to their play intensity measured by games per week.

From looking at the players as a whole, two perspectives are needed. One perspective is at what rate of play intensity do players improve the quickest *per game*. Figure 2 presents in this information by plotting skill over games for players grouped according to games per week. The figure shows that those who play 4–8 games per week seem to do best compared to other groups. However, from a different perspective of which players improve quickest *over time*, Figure 3 reveals that players who play more than 8 games per week can surpass the less frequent players. Despite learning at a lower rate per game, the additional games they played more than compensated for their slower skill gains. Interestingly, those who play more frequently per week tend to start as less skilled players, but improve more rapidly, as shown by the 32–64 and >64 games per week groups (i.e., players who logged over 8 hours a week of multiplayer Team Slayer).

Breaks in Play

Another aspect to the play patterns of players is breaks. Players commonly took breaks of days, weeks, or months due to vacation, to play other games, real life distractions, or just temporary boredom with one game. Here we look at skill change preceding the break to understand why a player took a break, and skill changes in games when the player returns to understand how much skill is lost during a break and how long it takes to recover.

Skill Change Preceding Breaks

We hypothesized from some preliminary discussion with players that they are more likely to continue playing if they are winning, and stop playing after bad losses.

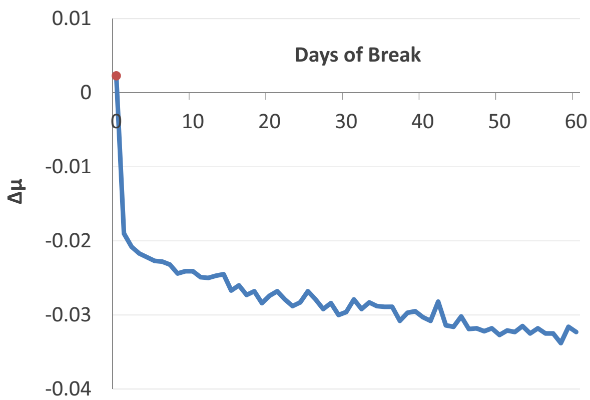


Figure 4. Skill change in the game preceding a break. When there is no break, skill change is barely positive, while breaks are correlated with a drop in skill (caused by a loss in the prior game).

Figure 4 shows evidence of a few behaviors that agree with our hypothesis. Skill gains correlate with no breaks (i.e., 0 day breaks), but only slightly; breaks of a day or more correlate with a decline in skill in the past game, which are caused by losses and longer breaks correlate more with larger skill drops in the past game (possibly losses against weaker opponents). This means players take breaks more often when they lose, which may be instances of frustration.

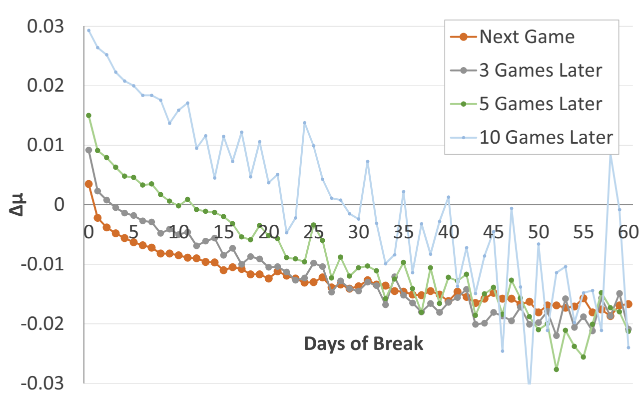


Figure 5. Skill change after different lengths of breaks for the next match, 3 matches after, 5 matches after, and 10 matches after. Larger drops in skill typically follow longer breaks, but players can catch back up quickly.

Skill Change Following Breaks

Figure 5 shows a few behaviors that players exhibit after breaks. The change in skill from before the break to after the break is illustrated by the 4 lines representing the next 1, 3, 5, and 10 matches after the break. When players are not taking breaks (breaks of 0 days), skill generally increases, evidenced by the climbing intercepts on the y-axis. Breaks of 1–2 days correlate with a small drop in skill in the next match played after the break, but has little long-term effect. Longer breaks correlate with larger skill decreases, but the relationship is not linear (60 day breaks do not seem to reduce skill twice as much as a 30 day break). More concretely, a 30 day break correlates with a skill drop of 10 matches of play (10 matches later, the skill returns to the value before the break ($\Delta\mu = 0$)); this is shown by the intersection of the 10 game later line with the x-axis. Thus, the amount of time required to regain skill following a 30 day break is only about 3 hours of gameplay.

Compared to retraining in sports, this catch-up time is short; this may be because there is little physical catch-up required. The player only has to refamiliarize themselves with the controls, and regain the mindset of their previous play.

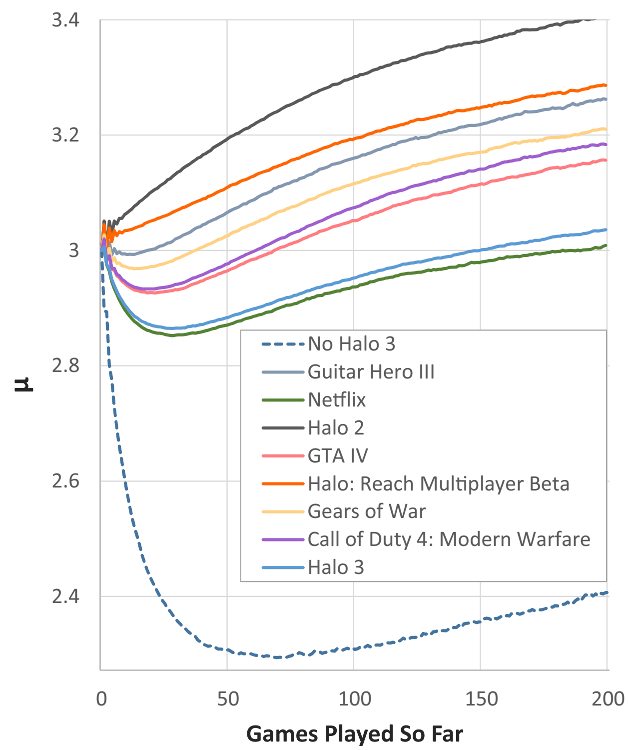


Figure 6. The performance of players who played other games *before* playing Halo Reach; these groups are not mutually exclusive so GTA IV players could also have played Halo 3.

Other Games Played Before Halo Reach

Figure 6 depicts the skill trajectories of players who played game titles before Halo Reach. The figure shows games that were popular among Halo Reach players on Xbox Live[2]. Players who did not play Halo 3 previously are less skilled but gain skill at about the same rate as everyone else. The title that correlates highest with skill is Halo 2 (even more so than Halo 3), but perhaps this is because those players have been engaged with the Halo

[2]The games did not include the original Halo (Halo 1) which was played on an older generation of Xbox Live servers.

franchise longer. Players who participated in the multiplayer beta of Halo Reach also performed well. Surprisingly, other first-person shooters and even Guitar Hero III correlated with higher skill, and Netflix and Halo 3 had similar effects on skill. This seems to be due to a large proportion of players in our study cohort (the players who started playing in the first week of Halo Reach) being Halo 3 players. These observations suggest that the number of game titles played before has an effect on skill development as well and that effect may even be stronger than the skill transfer from individual titles.

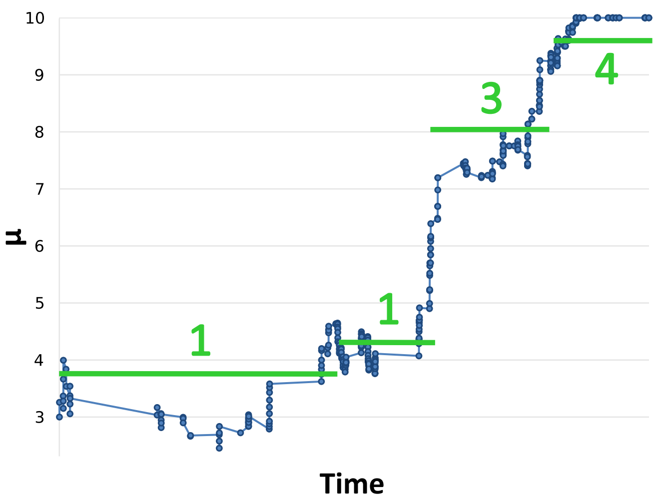


Figure 7. The skill plotted over time for an example individual player. The SAX representation is overlaid, creating the sequence “1134” indicating the player improved drastically in the second half of their matches, eventually becoming one of the best players.

Skill Change

While the median player’s skill increases over time, this is not true for every player. We can classify different players’ skill change over time and look into each group more closely. We converted the skill time-series into a symbolic representation of 4 levels and 4 time segments (4×4) using SAX (Symbolic Aggregate approXimation) [13]. SAX is a popular algorithm for discretizing time-series data. A player’s skill over time is normalized, and divided into equal segments; each segment is then converted into a symbol depending on how much it deviates from the expected mean (Figure 7 shows an example). The segments used in our study were the four periods between the first game and the 100th game (to control for the same number of games per player). Applying SAX to the skill over gameplay data allowed us to aggregate the different patterns of skill change from multiple players.

Table 1 shows that the most common pattern in skill change was a slow steady increase in skill. The second most common pattern showed the opposite trend—a slow decline in skill in the first 100 games. Additionally, numerous other patterns were common, including sharp rises and drops in skills, and improvement followed by decline and vice versa. The most surprising finding is

Pattern	# Players	Total Games
	61791	217
	45814	252
	36320	257
	27290	219
	22759	216
	22452	253
	20659	260
	20633	222
	19858	247
	19292	216
	17573	219
	17454	245
	17389	260
	15670	215
	13692	236
	12516	239

Table 1. Skill change patterns (top 15 most common patterns) for players that played at least 100 games. The total games is the mean number of games played in the entire 7 month period that we have data.

that players who improved in the first 100 games actually ended up playing fewer total games in the entire 7 month period than players with declining skill. We believe two factors play a role in this effect: 1) players that improve are more aggressive and hardcore gamers, who switch to other games earlier; 2) a skill improvement is not obvious to the player, but they do notice themselves performing worse against (unknown to them) stronger opponents, and the additional challenge may cause additional stress and frustration, provoking them to quit.

MODELING SKILL AND TOTAL GAMES

The previous analyses provided descriptive insight into how particular factors in patterns of play affected skill. We wanted to consider the factors holistically, and together, see how well they could predict their 1) *final skill rating* μ and 2) *number of games played* at the end up the 7-month period. This would provide information about how strong each factor comparatively correlates with skill, and may tell us what players are likely to play more games later. This investigation is relevant to game developers who have business needs such as determining the longevity of a game title, what load a server should expect, or whether particular patterns of play give certain players advantages in learning the game. For example, a player who is likely to play fewer games in total and end up with a low TrueSkill rating may have become frustrated with the game at a particular point because they were not improving.

Model Factors and Evaluation

To build the model comprising multiple factors, we looked at the players’ game histories. We used the explanatory power of different types of factors to identify which factors best explain the skill and total games of a player. A multiple linear regression computes an R^2 value measuring how much of the variance was explained by a particular combination of factors. An adjusted R^2 is reported in Table 2, which compensates for the number of terms in the model; the adjusted R^2 is a modified R^2 in which the value represents only the factors' improvement over the model that are better than chance.

We looked at four types of features: play intensity in the first 100 games, player demographics, skill change in the first 100 games, and other games played. For *player demographics*, there were four variables: country, age on the Halo Reach release date, how long they had been on Xbox Live, and the number of friends they had on Xbox Live. For *play intensity*, we looked at the number of days it took to reach 100 games, the standard deviation of the timestamps of the games played, and the number of games played per week. For *skill change*, we the player's first 100 matches, and extracted its 4×4 SAX representation as well as a 3×3 SAX representation of its last 30 matches. Finally, for *other games* played, we looked at whether they had played a prior version of Halo before, and whether they had played other first-person shooter titles before; however, most players did indeed have experience with both Halo and other first-person shooters which may have rendered this factor less effective.

Results

We noticed that the *final skill* for players was relatively predictable given knowledge of a player's discretized skill changes in their first 100 games. This leads us to believe that the most common patterns of steady improvement or decline in skill continue in a player's future games. However, demographics had little correlation with skill—it is difficult to predict how well a player performs from their age, country, number of friends, and join date on Xbox Live. These demographics did slightly improve the prediction over using just the SAX skill change factor, when combined with the skill change feature.

For explaining the *total number of games played*, the most important factor was the number of games played per week during the first 100 games. This seems obvious in retrospect since a player playing at a particular rate during the first 100 games is likely to continue at that rate and so it is possible to extrapolate how many games they will eventually play. Other factors of play intensity did little to improve the model once the games per week factor was incorporated, and other types of factors also had little effect. However, there was still a slight improvement to the model when all factors were combined, suggesting that the types of factors are relatively independent from one another.

MASTER BLASTERS

While we identified factors that correlated with performance, it is still unknown what leads to becoming highly skilled. A secondary objective we had is focusing on the subgroup of players who ended the 7-month period with the highest TrueSkill μ ratings. From the over 3 million players who began playing Halo Reach in its first week of release (a cohort which is a fairly large portion of overall players), we selected the top 100 players for individual analysis. We call these players "Master Blasters" for their mastery of the game.

	Adjusted R^2	
	Final Skill	Total Matches
Demographics	0.050	0.008
Skill Change	0.470	0.005
Other Games	0.010	0.006
Play Intensity	0.009	0.453
Combined Model	0.490	0.471

Table 2. Explanatory power of skill for different groups of factors from our analysis. The Adjusted R^2 value measures the variability in the data that can be accounted for by the model.

Master Blasters possessed some unique characteristics:

- 85 Master Blasters used the DMR weapon to make the most kills. In fact, 40 of the 100 masters used the DMR, sniper rifle, and melee as the top 3 weapons.
- Many Master Blasters had kill/death ratios[3] of 1.6, representing 5 kills for every 3 deaths on average, and nearly all Master Blasters played other multiplayer game types.
- The acquisition of skill for each Master Blaster varied substantially. Several improved drastically after short breaks (Figure 7), unlike the typical player, and others immediately jumped to a high skill rating within a few weeks of play.

However, looking at the individual player records did not show direct evidence of why they became the best. They certainly played more games than average, performed better in the games, and used particular weapons, but this could be said for a hundred thousand other players. They did not differ from the general population in demographics either (in terms of country and friends), although Master Blasters were more likely to be 18–25 years old. The answer to what makes the best players in Halo Reach seems to be absent from the factors we studied; this question continues to elude us, and further investigation is needed.

SURVEY RESULTS

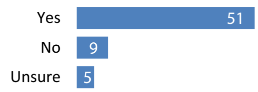
We wanted to present a comparative view to complement our data analysis with opinions from actual Halo Reach players. Thus, we contacted 300 participants who had opted-in to be contacted about their player experience by Microsoft via a survey. The participants had played at least one match of Team Slayer in Halo Reach. Having quantitative data and qualitative opinions via the survey cross-validates the perceptions of the actual players with the in-game data that is collected, providing valuable

[3] Kill/death ratios are a common way to compare players because the TrueSkill rating is hidden. A player must die for every kill, the average kill/death ratio is 1.

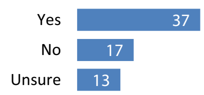
empirical strength of our data analysis results. Of the 300 people contacted, 70 people responded to the survey and answered at least one question; 65 people answered questions through to the end.

More than half the participants played the majority of their multiplayer games on the Team Slayer playlist (the game type used in this study). Nearly all participants also played other types of multiplayer game types. Half the participants reported playing fewer than 12 matches each week, while the other half played between 13 and 100 per week. 71% of the participants reported noticing a change in their skill during the period they played Halo Reach. Only 13% did not think their skill changed, while the rest were unsure.

Participants had diverse opinions about what contributed to improving their skill in Halo. 71% of participants believed they would improve faster by playing more games each week. They noted that familiarity with maps and weapons contributed substantially to their improvement. Additionally, many players noted that slowing down their game in terms of pacing their shots and thinking more during the match was an intentional act they did to improve their skill. Almost all participants thought they improved when playing with better players than themselves, and most (but fewer) thought they improved when playing against better players than themselves.



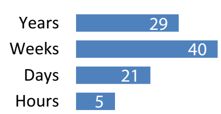
Do you improve by playing with better players?



Do you improve by playing against better players?

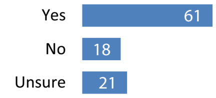
Emotional effects of playing Halo were frequently mentioned in the responses. Many participants reported feeling frustrated when they knew they could do better, particularly when their perceived skill declined from before. One participant remarked, “when I know I was supposed to be able to make that kill, or use that vehicle effectively, that generates a lot of disappointment.”

Participants felt taking breaks hurt their skill once they returned to the game. One participant attributing this phenomenon to reaction time and loss of familiarity with the controls, “When I return after a prolonged absence my aim is less sharp and I play rubbish for a while which is obviously less fun. I sometimes get the added bonus of my creaky brain forgetting the buttons which is never fun either!” The most common response to how long of a break would be needed to impact their skill was several weeks of break. This matches our results from the data analysis where a several week break produced a noticeable decline in skill when the participants returned, although in the analyses we noticed a decline of skill in even a single day of absence. In terms of causes of breaks, one participant described a “string of bad games” would cause them to take breaks, agreeing with our analysis of the pre-break match.



How long of a break will impact your skill?

Participant opinion was split on whether they thought other games influenced their skill in Halo. Some participants thought they became worse in Halo by playing other first-person shooters because those other first-person shooters required different techniques, but an almost equal number thought other FPSes helped them improve hand-eye coordination (and other “FPS skills”). However, most participants reported that they would consider buying other game titles if they thought playing those games would improve their skill in Halo.



Would you consider buying other games that may improve your skill?

The most common reasons participants stopped playing was due to boredom and other life issues (family, etc.). Rarely did it have to do with skill; in fact, most participants said they would not stop playing even if they were not improving, which agreed with our data (declining skill did not negatively influence the total number of games the players played).

Overall, the survey responses agreed with the in-game data analysis. Participants felt they were keenly aware of their skill, and it was an important determinant of how much they played, their enjoyment, and their ingame behavior.

DISCUSSION

Learning and Expertise

After analyses of the different factors and looking individually at the Master Blasters, what have we learned about skill? Certainly, the concept of skill varies for every task, as telegraphy or sports training is a very different activity than blasting enemies on a screen. Some factors we examined span multiple activities: frequency of performing the activity correlate with higher performance up to a point and a catch-up period follows long breaks. But among the top 100 players, skill is acquired differently: some are long-time players of the Halo franchise while others started with Halo Reach; some players gain skill rapidly the moment they start playing, and others reach a plateau and only become the best after a break. Our model shows that skill can be measured and quantified, but even when putting together numerous factors, our explanations (represented by adjusted R^2) can only go half-way.

Handling the Effect of Time on Skill

What might game designers do to deal with the effects of play patterns on skill? One direct approach is to incorporate patterns of play directly into the TrueSkill rating system. For example, some rating systems reduce a player’s skill rating during periods of inactivity, which also eliminates players who “camp” high ratings they may be undeserved. One participant acknowledged the decline in skill following a break, and the frustration of being expected to perform at their pre-break level, “Because the matchmaking system still thinks I am at level xx, so I should be matched against level xx players - even though I am rusty and need a few games to get back into the groove. This results in team mates just being overall d-bags and talking crap because you have a bad game since you havent played in a while." By changing a player's TrueSkill depending on when their most recent activity took place, the rating system can bias TrueSkill towards players who have played recently. Prior game data can be used to deduce the amount of reduction in TrueSkill that would accurately reflect their probability of winning after breaks.

The game can also take a proactive approach to informing the player. During periods of inactivity, a game system can remind the player that their skill may be declining and by returning to the game sooner rather than later, the player will stem the loss of skill. Another proactive task the system can perform is looking at players' skill change over time, represented by the SAX patterns previously. For players who are not predicted to improve on their own, the system may gently provide tips; most participants (78%) responded positively to receiving practical tips to improve their play.

Hidden Truths about Skill

One design consideration in multiplayer games is whether to show the user their skill rating. The designers of Halo switched between showing the TrueSkill in Halo 3, to replacing it with an experience score that only goes up in Halo Reach. Many survey responses noted the displeasure in losing skill, and the frustration caused by seeing a lowered rating may be a bad experience for competitive players. However, players remarked enjoying seeing their skill increase, which gave them a feeling of satisfaction. The lack of a public skill rating prevents the game from providing positive feedback to the player, who may then become frustrated when they feel like they're not improving when in fact they are simply being matched against stronger opponents.

In cases where the game designers decide not to reveal a player rating, perhaps there may be some benefit to unevenly balanced matches. When a player has improved their skill, rather than pitting them against tougher opponents who may make them feel more frustrated, the players could be matched against the same caliber of opponents before their increase in skill, i.e., slightly weaker opponents. This allows the improving player to enjoy the spoils of expertise for a bit. Another manipulation is when the game notices players declining in skill, the game may want to stop them before they reach a "pain point", where they may become so frustrated with the game they may give it up permanently. This is more useful when the skill ratings are public, as the large drop in skill may leave the player feeling helpless and unable to recover. The hiding of information from the players for their own benefit relates to Cheung and Huang's suggestion that the decision about what information to reveal to stakeholders in a game can have substantial implications for how much the stakeholders enjoy the game (in their case, revealing game information and events to spectators of the game StarCraft) [3]. These questions of hiding and showing skill information and their affect on player emotions are promising areas of further investigation.

Limitations

One common limitation of post-hoc data analysis is the inability to claim causation effects since the game variables unavailable for manipulation to create a controlled experiment. When two variables such as gameplay intensity and skill are correlated, it cannot be determined whether increased gameplay intensity caused an increase in skill or whether some external factor caused both factors to increase; for example, this external factor may be that those players who play more frequently are naturally capable of gaining expertise quicker. The third possibility, that an increase in skill causes more gameplay intensity is also possible, but is unlikely in our study since the player's TrueSkill is not shown to the player. Regardless, we caution claiming that particular factors will cause an increase in skill, but rather our findings describe the nature of players who have higher skill.

TrueSkill starts with an initial prior, and it also takes some time to adjust to a new skill rating. The lag may result in an inaccurate rating in the beginning (due to the prior) or if a player's skill changes substantially. We observed the median TrueSkill decreases from the starting value $\mu = 3$ initially probably because a player's actually skill is lower than that, and it's not until about 30–35 games later that the TrueSkill μ rises again.

User-reported data from 300 players when they signed up for the opt-in player experience panel showed that 18 of them (6%) reported sharing their Xbox live account with other people. When those sharing an account play the same game, and particularly the same playlist in Halo Reach, their different skills will confuse the TrueSkill rating system. The better player may raise the skill rating when they are playing, while the worse player will tend to lower the skill rating, causing it to be highly variable. Thus, during matchmaking, the TrueSkill μ may not accurately reflect the skill of the current player. Additionally, Xbox Live accounts can be handed off to another person, resulting in a similar inaccurate reflection of skill in the TrueSkill rating; it would take a number of games for the rating to readjust to the new player's skill.

Our study did not incorporate the TrueSkill rating from playlists other than Team Slayer. While Team Slayer is the most popular multiplayer variant, players may be learning relevant skills for Team Slayer from similar variants or playlists (for example, by gaining familiarity with maps or weapons in both playlists). Because the TrueSkill ratings are independent from playlist to playlist, TrueSkill does not capture this learning effect either. Players who are playing a variety of multiplayer game types to be technically improving their skill faster than what the TrueSkill system knows from their matches on Team Slayer.

Players may drop out of games due to connection issues, to avoid losing, or because they have to do something else. Halo penalizes those players by marking a drop-out as a loss, even when the team the player was on wins the game. Additionally, players who drop out frequently may be matched up with others who do the same, and eventually the players will be banned from playing online if they drop out too frequently. These drop-outs may affect the assigned TrueSkill ratings for a player, whether deservedly or not.

CONCLUSION

Our study reported numerous findings relating gamers' patterns of play with their skill as computed by the TrueSkill rating system. Through the large data available, we were able to conduct detailed controlled analyses of how particular factors correlated with skill in Halo Reach's Team Slayer playlist. Our primary findings show that play intensity, breaks, and skill changes in early matches may affect the skill level of the player. The factors put together have a moderate ability to explain the variation in skill between players. We supplemented the data analysis with a survey that included open-ended questions; this provided insight into what players themselves felt were determinants of skill.

Overall, we believe our contribution informs the exploration of skill and expertise in a novel domain—video games. Large scale studies where people are highly motivated and meticulously measured are difficult to conduct in real world settings. Our findings participate in the conversation of expertise and learning behavior common in domains such as sports training and education. Additionally, leveraging in-game data can help game designers create better experiences for players, increasing their enjoyment and eliminating frustrating scenarios.

ACKNOWLEDGMENTS

The authors thank Thore Graepel and Tim Menzies for advice during early stages of the work, Coco Inoue for assistance in providing data and continual feedback, and Tom Ball, Venkatesh-Prasad Ranganath, Ramarathnam Venkatesan, and Ben Zorn for feedback during the study. We also thank the Halo Reach players who participated in our survey. Jeff Huang performed this work during a summer internship at Microsoft Research.

REFERENCES

1. Bryan, W. L., and Harter, N. Studies in the physiology and psychology of the telegraphic language. *Psychological Review 4*, 1 (1897), 27–53.
2. Bryan, W. L., and Harter, N. Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review 6*, 4 (1899), 345–375.
3. Cheung, G., and Huang, J. Starcraft from the stands: understanding the game spectator. In *Proc. of CHI* (2011), 763–772.
4. Dangauthier, P., Herbrich, R., Minka, T., and Graepel, T. Trueskill through time: Revisiting the history of chess. In *Proc. of NIPS* (2007), 337–344.
5. d_ijk_stra. Statistical bonjwa theory: the history of bw. Retrieved August 1, 2012 from `http://www.teamliquid.net/forum/viewmessage.php?topic_id=357124` .
6. Elo, A. E. *The rating of chessplayers, past and present.* Batsford, 1978.
7. Harrison, B., and Roberts, D. L. Using sequential observations to model and predict player behavior. In *Proc. of FDG* (2011), 91–98.
8. Herbrich, R., Minka, T., and Graepel, T. Trueskill™: A bayesian skill rating system. In *Proc. of NIPS* (2006), 569–576.
9. Hock-koon, S. Affordances of elliptical learning in arcade video games. In *Proc. of DiGRA Nordic* (2012).
10. Hullett, K., Nagappan, N., Schuh, E., and Hopson, J. Empirical analysis of user data in game software development. In *Proc. of ESEM* (2012), 89–98.
11. Keller, F. The phantom plateau. *Journal of the Experimental Analysis of Behavior 1*, 1 (1958), 1–13.
12. Lee, M. J., and Ko, A. J. Personifying programming tool feedback improves novice programmers' learning. In *Proc. of ICER* (2011), 109–116.
13. Lin, J., Keogh, E., Lonardi, S., and Chiu, B. A symbolic representation of time series, with implications for streaming algorithms. In *Proc. of SIGMOD Workshop on Research issues in data mining and knowledge* (2003), 2–11.
14. Mason, W., and Clauset, A. Friends ftw! friendship and competition in halo: Reach. In *Proc. of CSCW* (2013).
15. Reeves, S., Brown, B., and Laurier, E. Experts at play understanding skilled expertise. *Games and Culture 4*, 3 (2009), 205–227.
16. Staron, R. S., Leonardi, M. J., Karapondo, D. L., Malicky, E. S., Falkel, J. E., Hagerman, F. C., and Hikida, R. S. Strength and skeletal muscle adaptations in heavy-resistance-trained women after detraining and retraining. *Journal of Applied Physiology 70*, 2 (1991), 631–640.
17. Wallner, G., and Kriglstein, S. A spatiotemporal visualization approach for the analysis of gameplay data. In *Proc. of CHI* (2012), 1115–1124.
18. White, R. W., Dumais, S. T., and Teevan, J. Characterizing the influence of domain expertise on web search behavior. In *Proc. of WSDM* (2009), 132–141.
19. Xu, Y., Cao, X., Sellen, A., Herbrich, R., and Graepel, T. Sociable killers: understanding social relationships in an online first-person shooter game. In *Proc. of CSCW* (2011), 197–206.