2022 14th International Conference on Quality of Multimedia Experience (QoMEX)

Data-Driven Study of Long-Term Gaming Experience

Jens Helge Reelfs
Brandenburg University of Technology

Oliver Hohlfeld
Brandenburg University of Technology

***Abstract*—A subset of massively multiplayer online games (MMOG) feature long-term game rounds in which players interact for months or even years. The player experience of such long-term games cannot be entirely captured by current study methods, in particular not at scale assessing large player populations. To address this challenge, we posit that long-term, round based games such as Tribal Wars (browser-based) enable a data-driven perspective on long-term game dynamics and experience. In a preliminary study, we monitor and characterize the entire longitudinal game state of a Tribal Wars round that was played by 16k players for 1.5 years, enabling us to investigate behavioral patterns of all active players. We identify features that capture in-game success and relate to the player experience; showing only successful players keeping up playing. We open source our dataset [1] enabling reproducibility & future research.**

***Index Terms*—Long Term Gaming QoE, MMOG, Tribal Wars**

I. Introduction

Current Gaming Quality of Experience evaluations assume short test stimuli for interactive tests, e. g. in the order of 90-120 seconds for short or 10-15 minutes for long stimuli as per P.809 [2]. Stimuli of this length suffice assessing player experience for *most* games. The implicit assumption is that the stimuli duration is short enough to avoid fatigue and render interactive tests feasible, yet long enough to represent a typical game situation that enables assessing quality features (e. g. evaluating interaction delays usually requires short tasks only).

Yet, for some games single rounds span multiple months or years, e. g. MMOGs such as World of Warcraft (WoW) or Tribal Wars (TW). While e. g. WoW implements an endless game world without a defined start or end, others such as TW establish game rounds having a clearly defined start date and round-ending goal or maximum time. In TW, multiple rounds co-exist featuring various game settings and goals (e. g. casual, regular). Finishing a round requires players' long-term commitment at cumulative playtime costs to quit. We posit that a well-defined round structure and the surprising commitment of many players render games such as TW prime candidates to study long-term dynamics and user experience.

While many quality features for player experience can be assessed in short interactive tests, they cannot capture the player experience over an entire round that lasts for years. Hence, it remains unclear how the make longitudinal player experience tangible *at scale*. Interactive tests at this widened timeframe are infeasible and questionnaire-based surveys do not scale to large populations of unknown users. Consequently, many related aspects such as defining and long-term integration of single or cumulative (bad) usage experiences remain unknown.

978-1-6654-8794-8/22/$31.00 ©2022 IEEE

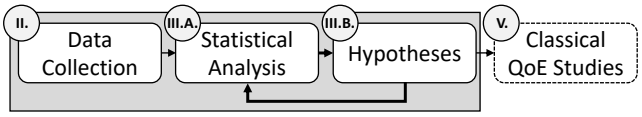


Fig. 1: Data-Driven Large Scale QoE Study Pipeline

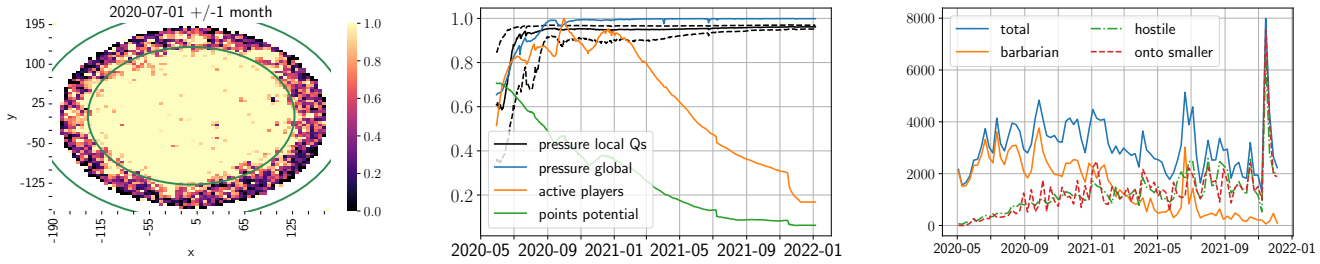
Contribution. As shown in Figure 1, we make the case for a longitudinal data-driven perspective to gain insights into long-term game dynamics. We assess the entire game state evaluating behavioral patterns of *all active players*—not only few selected players under study. Our presented first step analyzing domain-specific game metrics is as crucial as attractive as a key enabler in the realm of widened timeframes, such as studying games with single round durations lasting months or even years. We argue that this approach can identify behavioral patterns and helps to hypothesize about circumstances impacting player experience. In a second step (beyond this paper), identified hypotheses can then be examined in dedicated tests following traditional study designs, e. g. P.809.

Approach. *Data Collection:* We enable this study by continuously monitoring the game state of all 16k players of a single Tribal Wars game round for more than 1.5 years. *Statistical Analysis:* With data collected from the game's public API we take an empirical perspective on the dynamics of a long-term game requiring consistent player commitment over years. *Hypotheses:* By leveraging churn as an experience signal, we monitor the game progression over time and identify patterns influencing player experience (e. g. causing players to quit).

II. Long-Term Gaming & Data Collection

Tribal Wars Game. We base our study on the popular massively multiplayer online game (MMOG) *Tribal Wars*, that was first released in 2003 [3]. In Tribal Wars, players start off controlling a medieval village that needs to be expanded and protected. Players can team-up in tribes and conquer other players' villages expanding their empire. Any player may join various isolated and dedicated game *rounds*. Such game rounds have specific goals to be reached typically played for multiple years. Successfully playing and finishing a round thus requires a long-term and frequent time commitment by a player.

Data Collection. The game platform provides a public API to collect current snapshots of the entire game state (see API documentation [4], statistics [5]). We retrieve the data for a single casual round (dep13) every hour starting in May 2020 until December 2021. It captures the state of all 16k players, 93k villages, 1.4k tribes, and other game-related events and properties of the game world. We enable future research by open sourcing it [1]. Note that *casual* rounds enforce heavy restrictions on possible hostile attacks and conquers.



(a) 2d Game World: Village Pressure (b) Village Pressure, Players & Potential (c) Village Conquers

Fig. 2: **(a)** *Left:* Heatmap of local pressure (ratio: available to total villages) on July 1 2020. With new players, the game concentrically spawns new villages (circles indicate ± 1 month). Due to player expansion, unoccupied villages decrease over time, resulting in a declining village pressure gradient towards the younger outer regime. **(b)** *Center:* Village pressure changes drastically within early phases, albeit joining players counteract on short term. With 2021, players lose interest while potential for further construction decreases. **(c)** *Right:* Weekly conquers over time, loosely correlating to active players. Due to increasing pressure, conquering barbarian villages becomes unlikely, hostility in takeovers escalates—mostly onto smaller players.

Ethics. The collected game state data is publicly provided by the publisher to enable statistical analysis of the game state. It does not contain personal or otherwise sensitive information.

III. Long-Term Game Dynamics Analysis

We begin by studying game dynamics via engineered domain-specific game & player features (III-A). This is fundamental to further derive player experience indicators (III-B).

A. Game Dynamics: Inevitable Tension in Expansion

The need for continuous growth. The monitored round is a *casual* round, which lets most players focus on only constructing and expanding their villages rather than on defending against or attacking others. Thus, continuous growth & expansion becomes the ultimate goal. Yet, the room for expansion gets smaller over time, creating tension. Consequently, the player experience largely depends on the remaining space.

Acquiring new villages. We begin by illustrating this resource concentration within the game world. For each new joining player, the game spawns a new player village and additional unoccupied villages, *barbarians*, concentrically on a 2d world map, i. e. the oldest villages are within the center. In our case we find only ≈ 5.7 villages per unique player, which heavily dissatisfies demands limiting room for expansion thereby creating pressure. To showcase this pressure for expansion, we compute the density of non-barbarian villages for each occupied village individually in a radius of up to 15. We aggregate these values for July 1, 2020, as a heatmap in Figure 2a in which darker colors represent low pressure areas that still have barbarian villages left to satisfy expansion demands. Naturally, we identify concentric outer regions with available barbarian villages due to having spawned only recently. However, over time the used game world expands with more players and outer low-pressure regions transform rapidly as shown by the circles indicating the expanse one month earlier and later (darkgreen).

We confirm this trend in Figure 2b, where we plot pressure-related measures and the amount of active players over time. In the beginning of our observation period, global village pressure (blue) is only at a ratio of 0.65, i. e. 65% of all available villages are already occupied. This ratio increases rapidly over 3.5 months almost reaching full village occupation (≈ 1.0). Next, we look at the local pressure defined as the ratio of unoccupied villages in the local neighborhood of all occupied villages. The quantiles (0.25, 0.5, 0.75; black) of local village pressure vary accounting for developments at the outer regimes of the gaming world, but follow the same trend.

Expanding existing villages. Besides acquisition of new villages, construction in possessed villages is equally important; thus, we also show the construction potential proxied by remaining optimal village points (green) across time, which naturally decreases due to limited village resources.

Number of players. The number of players controls the game resources and the player experience controls the number of active players. We thus study the number players (normed by maximum; orange). Within early stages, amounts increase alleviating expansion pressure (players and local pressure Q25 counteract), while players start dropping out in 2021. The number of players further controls game resources. Flattening figures in new players (now shown) results in ever slowing down game world expansion. The vanishing potential to fortify existing villages (green line) correlates with declining numbers of active players. Further ever higher pressure correlates with shorter player lifetimes at later game stages (not shown).

The only *competitive* solution for continuing to play is further expansion by either scanning for nearby abandoned villages, targeting newly spawned villages at possibly very long runtimes, or engaging in hostile actions against other players within allowed casual game restrictions. Either way, obtaining new villages becomes highly competitive the more the game progresses and the more villages are already occupied by other players. We spotlight in-game village conquers over time in Figure 2c. Focusing the overall conquers (total; blue), we observe a correlation to active player numbers, primarily driven by conquers on barbarian villages (orange)

until end of 2020. In later game stages, abandoned villages from leaving players get re-occupied within only few days (not shown). In line with observed expansion pressure, players experience an increasingly aggressive environment at striking figures of hostile conquers (green), which largely happens disproportionate in player points (red), i. e. the rich get richer. An increasing trajectory in per player kill counts over time confirms this increasingly hostile environment (not shown).

In summary, continuing to play requires ever harder expansion due to dwindling potentials & increasing competitiveness.

B. Player Experience Indicators

Informed by game dynamics shown before, we set out to identify features that relate to player experience. Since we cannot directly measure the player experience, we leverage user churn as a negative experience signal. We align players in-game time enabling a systematic comparison of player metric populations over time. Thus, we first interpolate starting dates of players that joined before our observation period from the conquers log by distance to the game map center. Next, we slice weekly buckets through the relative in-game time assigning each player to a non-churn and churn class depending on whether the user will quit within the next two weeks. We study per-player metrics derived from amounts of villages, point developments, kills, (hostile) conquers, and the village pressure. This allows for identifying significant differences in metric populations according to a MannWhitneyU-test.

In Table I, we present (positive) indicator significance responses (90% percentiles of the weekly buckets) on players' first four weeks (timeframe for most players), and overall in comparison. We further provide insights about the in-indicator relation between the non-churner and churner group by comparing averages across buckets.

The relation between both groups' average indicator values immediately point into the success direction to varying degrees ("="; "$>$": single digit relative difference; "$\gg$": an order of magnitude; "$\ggg$": two orders of magnitude). While most identified player experience indicators mostly reflect in-game activity and success, global village pressure and amounts of acquired barbarian villages are significant within a player's first 4 weeks. Heavy user influx and low pressure regions within early game stages explain this finding providing a hypothesis that apparent rising expansion pressure might be a driving factor for a bad user experience resulting in churn; both measures are indistinguishable at later game stages creating heavy tensions in continuing to expand.

metric	p-values < 0.05 first 4 weeks	overall	avg relation non/churn	
player rank	$\mathbf{1.626 \cdot 10^{-7}}$	**0.004248**	$<$	0.762
player villages	$\mathbf{2.104 \cdot 10^{-7}}$	**0.000003**	$>$	3.568
player points	$\mathbf{2.866 \cdot 10^{-2}}$	**0.000018**	$>$	3.629
has ally fraction	$\mathbf{4.209 \cdot 10^{-117}}$	**0.007655**	$>$	1.574
village pressure global	$\mathbf{1.443 \cdot 10^{-11}}$	0.230208	$=$	0.996
village pressure local	$\mathbf{1.783 \cdot 10^{-4}}$	0.414349	$=$	1.004
villages gain	$\mathbf{1.569 \cdot 10^{-8}}$	0.067301	$\gg$	16.82
villages gain barbarian	$\mathbf{5.366 \cdot 10^{-8}}$	0.219401	$\gg$	16.92
kill all	$\mathbf{2.372 \cdot 10^{-12}}$	0.073920	$>$	2.697
kill other	$4.408 \cdot 10^{-1}$	**0.067183**	$\ggg$	343.6

TABLE I: Experience indicators excerpt. We find significant differences in feature distributions of soon churning players compared to others. Most features and relations reflect measures of in-game success. Pressure-related metrics are only significant in the beginning of a player's career.

IV. Related Work

Gaming QoE. The QoE community has focused on studying factors influencing gaming QoE within recent years. Efforts focused on certain game architectures—e. g. cloud [6] & mobile games [7]—or game genres: e. g. MMORPGs [8] & FP Shooters [9]—or VR gaming and related cybersickness [10]—to name a few. These studies provided an understanding of gaming QoE factors [11], which resulted in standardized study methods for gaming quality (P.809 [2]) or the opinion model predicting gaming QoE for cloud gaming services (G.1072 [12]). All of these studies have so far focused on short study durations. Contrasting short durations, we complement empirical long-term gaming insights from meta-data [13], showcasing how to leverage long-term in-game dynamics for assessing actual player quality of experience.

Long-Term QoE Integration. Research on the perceived quality over multiple usage episodes (multi-episodic QoE) emerged in 2011 with multi-day experiments [14], [15]. Later, *individual sessions* were studied (i. e., continuous use of the same service with multiple usage episodes) [16], [17]. Despite first findings, the formation process of multi-episodic perceived quality remains far from being understood. We posit that a data-driven perspective on the complete game state complement complex field or lab tests and can thus provide an interesting perspective on the long-term usage of a service.

V. Conclusion & Future Directions

No gaming QoE study is based on monitoring the in-game state of *all players* in a dedicated game round *for years*. We shed light on this still unexplored area of gaming experience. While selecting an appropriate game is challenging, we posit that long-term, round-based games such as Tribal Wars are a prime candidate to study player experience. We reason that well-defined rounds spanning years enable controlled statistical evaluations of player dynamics; creating the foundation of future work in the direction of multi-episodic gaming QoE.

We identify features capturing in-game success and hypotheses relating to the player experience. While e. g. questionnaires about specific game events or sequences may provide sentiments for better modeling experience, we argue that such a data-driven perspective is a necessary first step identifying possible influence factors for subsequent other methods that determine causality (e. g. classical interactive tests). That is, our presented data-driven perspective results only in significant correlations, which cannot replace classical studies that identify causality, yet it can enable them by identifying test cases. By open sourcing our dataset & evaluations [1], we enable reproducibility and pave the way for exploring gaming QoE in a new field to the community: long-term gaming experience.

REFERENCES

[1] J. H. Reelfs and O. Hohlfeld, "Data-Driven Study of Long-Term Gaming Experience (Tribalwars)," Jul. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6862106

[2] ITU-T Recommendation, "P.809: Subjective evaluation methods for gaming quality," 2018.

[3] InnoGames, "Tribal wars," https://www.innogames.com/games/tribal-wars/.

[4] "Tribal wars world state api," https://forum.die-staemme.de/index.php?threads/weltdaten-und-configs.183996/.

[5] "Tribal wars stats," https://www.twstats.com.

[6] S. Schmidt, "Assessing the quality of experience of cloud gaming services," Ph.D. dissertation, TU Berlin, 2021.

[7] I. Slivar, Z. Dragozet, and L. Skorin-Kapov, "User customization of the gaminganywhere android mobile client interface," in *NetGames*, 2015.

[8] M. Suznjevic, L. Skorin-Kapov, and M. Matijasevic, "The impact of user, system, and context factors on gaming qoe: A case study involving mmorpgs," in *QoMEX*, 2013.

[9] S. Vlahovic, M. Suznjevic, and L. Skorin-Kapov, "The impact of network latency on gaming qoe for an fps vr game," in *QoMEX*, 2019.

[10] S. Vlahovic, M. Suznjevic, N. Pavlin-Bernardic, and L. Skorin-Kapov, "The effect of vr gaming on discomfort, cybersickness, and reaction time," in *QoMEX*, 2021.

[11] S. Möller, S. Schmidt, and J. Beyer, "Gaming taxonomy: An overview of concepts and evaluation methods for computer gaming QoE," in *QoMEX*, 2013.

[12] ITU-T Recommendation, "G.1072: Opinion model predicting gaming quality of experience for cloud gaming services," 2020.

[13] W.-c. Feng, D. Brandt, and D. Saha, "A long-term study of a popular mmorpg," in *ACM SIGCOMM Workshop on Network and System Support for Games*, ser. NetGames'07, 2007.

[14] S. Möller, C. Bang, T. Tamme, M. Vaalgamaa, and B. Weiss, "From single-call to multi-call quality: a study on long-term quality integration in audio-visual speech communication," in *INTERSPEECH*, 2011.

[15] D. Guse and S. Möller, "Macro-temporal Development of QoE: Impact of Varying Performance on QoE over Multiple Interactions," in *Proceedings of AIA-DAGA Conference on Acoustics*, 2013.

[16] D. Guse, B. Weiss, F. Haase, A. Wunderlich, and S. Möller, "Multi-episodic perceived quality for one session of consecutive usage episodes with a speech telephony service," *Quality and User Experience*, vol. 2, no. 1, Aug. 2017.

[17] D. Guse, O. Hohlfeld, A. Wunderlich, B. Weiss, and S. Möller, "Multi-episodic perceived quality of an audio-on-demand service," in *QoMEX*, 2020.