What Makes a Good Team? A Large-scale Study on the Effect of Team Composition in Honor of Kings

Ziqiang Cheng
Zhejiang University

Yang Yang*
Zhejiang University

Chenhao Tan
University of Colorado Boulder

Denny Cheng
Tencent Corporation

Alex Cheng
Tencent Corporation

Yueting Zhuang
Zhejiang University

ABSTRACT
Team composition is a central factor in determining the effectiveness of a team. In this paper, we present a large-scale study on the effect of team composition on multiple measures of team effectiveness. We use a dataset from the largest multiplayer online battle arena (MOBA) game, Honor of Kings, with 96 million matches involving 100 million players. We measure team effectiveness based on team performance (whether a team is going to win), team tenacity (whether a team is going to surrender), and team rapport (whether a team uses abusive language). Our results confirm the importance of team diversity and show that diversity has varying effects on team effectiveness: although diverse teams perform well and show tenacity in adversity, they are more likely to abuse when losing than less diverse teams. Our study also contributes to the situation vs. personality debate and show that abusive players tend to choose the leading role and players do not become more abusive when taking such roles. A detailed version with additional analysis and prediction experiments can be found on arXiv: https://arxiv.org/abs/1902.06432.

CCS CONCEPTS
• Applied computing → Law, social and behavioral sciences.

KEYWORDS
team composition, team performance, tenacity, toxic behavior, MOBA

ACM Reference Format:

1 INTRODUCTION
The increasing complexity and scale of tasks in modern society require individuals to work together as a team [13, 41]. Crucial to the effectiveness of a team are the individuals in the team, i.e., team composition. Extensive prior research has studied team roles to analyze team composition [6, 7, 31, 32, 37, 39]. The most influential work is Belbin’s team role framework, also known as Belbin Team Inventory. A central hypothesis in Belbin’s framework is that balance in team roles is associated with team performance. Belbin [7] defines nine roles, and finds that teams with certain role combinations result in poor performance, even if they are formed by members with the sharpest mind and the most experience.

Another challenge in studying the effect of team composition on team effectiveness arises from the definition of effectiveness. Team performance is the most straightforward definition and has been studied in a battery of studies [10, 13, 16, 36]. Example measures of team performance include the impact of published papers from a team [41], winning a sports game [4, 25], etc. However, the effectiveness of a team can also be reflected by the tenacity in face of adversity, and the rapport between team members [9]. The effect of team composition likely depends on the definition of team effectiveness. Therefore, it is important to understand the effect of team composition on the effectiveness of teams in multiple measures.

To address these two challenges, we identify multiplayer online battle arena (MOBA) as an ideal testbed and provide a large-scale study on the effect of team compositions on multiple measures of team effectiveness. We use a dataset from Honor of Kings, the largest MOBA game in the world. This platform is ideal for studying the effect of team composition for three reasons. First, there are defined roles in the game and players choose their roles to form a team. Given that there are five players in each team, we are able to enumerate all possible team compositions. We also have information about the characteristics of each role. Second, the popularity of Honor of Kings leads to digital traces of hundreds of millions of players in hundreds of millions of games. These detailed game records allow us to explore the notion of team “effectiveness” beyond the simple measure of team performance, winning or losing. For instance, we study the effect of team composition on abusive language use, which reflects the rapport in a team. This research question naturally connects to the literature on toxic behavior in online communities and gaming [11, 12, 27].

Finally, we have access to the past history of players for capturing their background and experience. Historical information enables us to investigate questions in the “situation vs. personality” debate [17, 23]: how much of our behavior is determined by fixed personality traits or by the specific situation at hand? We will explore this question in the context of toxic behavior in online gaming.

∗Corresponding author: Yang Yang, yangy@zju.edu.cn.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW ’19, May 13–17, 2019, San Francisco, CA, USA
© 2019 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.
ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3312504

1See https://en.wikipedia.org/wiki/Team_Role_Inventories for a quick explanation.
In this paper, we conduct a large-scale study on the effect of team composition in the context of online gaming, particularly, in Honor of Kings. Our dataset comes from this popular MOBA game in China, and includes 96 million games and 100 million players in total. We provide details of the dataset and character roles in §2.

We quantitatively study the effect of team composition on three distinct definitions of team effectiveness: (§3) team performance — whether a team is going to win — is directly based on the result/goal in a game and is the most common measure in prior studies; (§4) team tenacity — whether a team is going to surrender — indicates a team’s resilience to adversity and is understood in existing literature as most studies only look at individual tenacity or perseverance [5, 20]; (§5) toxic behavior, i.e., abusive language use, reflects the rapport in a team and is extracted from the activities during a game.

Our results on team effectiveness confirm the theory that balance in team roles is important for effective teams. We find that diverse teams are more likely to win. Furthermore, diverse teams are less likely to surrender, indicating a higher level of tenacity. Note that the likelihood to surrender does not correlate with winning rate for commonly used team compositions. However, we observe diverging effects on abusive language use: diverse teams are more likely to abuse when losing, but are less likely to abuse when winning.

In addition to exploring the effect of team composition we investigate individual-level abusive language use to understand how teamwork influences individuals and shed light on the “situation vs. personality” debate [17, 23]. We find that players who choose assassins (the leading role) are more likely to abuse. We further demonstrate that the reason is that abusive players are more likely to choose assassins, instead of the alternative explanation that a player becomes more abusive when playing assassins. Finally, we present related work in §6 and offer our concluding thoughts in §7.

### Dataset description

Our dataset is derived from daily logs spanning three weeks in the August of 2017. Specifically, we randomly sample 20K games in each day of the third week, thereby obtaining 140K games and 1.3M players participating in these games. Among them, there are 78K ranked games. Unranked games are only used to analyze players’ role preferences.

### Team roles

There are five roles in Honor of Kings: warrior ( ), mage ( ), marksman ( ), assassin ( ), and support ( ). A

---

**Organization and paper highlights.** In this paper, we conduct a large-scale study on the effect of team composition in the context of online gaming, particularly, in Honor of Kings. Our dataset comes from the popular MOBA game in China, and includes 96 million games and 100 million players in total. We provide details of the dataset and character roles in §2.

We quantitatively study the effect of team composition on three distinct definitions of team effectiveness: (§3) team performance — whether a team is going to win — is directly based on the result/goal in a game and is the most common measure in prior studies; (§4) team tenacity — whether a team is going to surrender — indicates a team’s resilience to adversity and is understood in existing literature as most studies only look at individual tenacity or perseverance [5, 20]; (§5) toxic behavior, i.e., abusive language use, reflects the rapport in a team and is extracted from the activities during a game.

Our results on team effectiveness confirm the theory that balance in team roles is important for effective teams. We find that diverse teams are more likely to win. Furthermore, diverse teams are less likely to surrender, indicating a higher level of tenacity. Note that the likelihood to surrender does not correlate with winning rate for commonly used team compositions. However, we observe diverging effects on abusive language use: diverse teams are more likely to abuse when losing, but are less likely to abuse when winning.

In addition to exploring the effect of team composition we investigate individual-level abusive language use to understand how teamwork influences individuals and shed light on the “situation vs. personality” debate [17, 23]. We find that players who choose assassins (the leading role) are more likely to abuse. We further demonstrate that the reason is that abusive players are more likely to choose assassins, instead of the alternative explanation that a player becomes more abusive when playing assassins. Finally, we present related work in §6 and offer our concluding thoughts in §7.

### Dataset description

Our dataset is derived from daily logs spanning three weeks in the August of 2017. Specifically, we randomly sample 20K games in each day of the third week, thereby obtaining 140K games and 1.3M players participating in these games. Among them, there are 78K ranked games. Unranked games are only used to analyze players’ role preferences.

### Team roles

There are five roles in Honor of Kings: warrior ( ), mage ( ), marksman ( ), assassin ( ), and support ( ).

---

**Table 1: Statistics of the dataset.**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall statistics</td>
<td></td>
</tr>
<tr>
<td>#ranked games</td>
<td>54,434,817</td>
</tr>
<tr>
<td>#team compositions</td>
<td>126</td>
</tr>
<tr>
<td>#surrender games</td>
<td>5,807,005 (10.7%)</td>
</tr>
<tr>
<td>#abusing games</td>
<td>30,668,577 (56.3%)</td>
</tr>
<tr>
<td>#early games(t ≤ 11 min)</td>
<td>5,049,934 (9.3%)</td>
</tr>
<tr>
<td>#middle games(11 &lt; t ≤ 16 min)</td>
<td>23,089,993 (42.4%)</td>
</tr>
<tr>
<td>#late games(t &gt; 16 min)</td>
<td>26,294,890 (48.3%)</td>
</tr>
<tr>
<td>Top three frequent combinations</td>
<td></td>
</tr>
<tr>
<td>assassin</td>
<td>16,186,003 (14.9%)</td>
</tr>
<tr>
<td>mage</td>
<td>14,516,443 (13.3%)</td>
</tr>
<tr>
<td>support</td>
<td>12,630,805 (11.6%)</td>
</tr>
<tr>
<td>Personal statistics</td>
<td></td>
</tr>
<tr>
<td>#players</td>
<td>1,306,754</td>
</tr>
<tr>
<td>#players who played ranked games</td>
<td>1,173,372</td>
</tr>
<tr>
<td>#ranked games each player has</td>
<td>48.4 (0.05)</td>
</tr>
<tr>
<td>average winrate</td>
<td>52.4% (1.0 × 10⁻⁴)</td>
</tr>
<tr>
<td>average abusing probability</td>
<td>9.3% (1.0 × 10⁻⁴)</td>
</tr>
<tr>
<td>average surrender probability</td>
<td>5.0% (6.9 × 10⁻⁵)</td>
</tr>
<tr>
<td>#target players who have played at least 20 ranked games</td>
<td>758,494 (58.0%)</td>
</tr>
</tbody>
</table>

**Table 2: Description and frequency of each role.**

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
<th>Freq(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>warrior</td>
<td>Warriors have a large pool of health and considerable damage; they often undertake the stress from enemies in side lanes.</td>
<td>27.1</td>
</tr>
<tr>
<td>mage</td>
<td>Mages have heavy spell damage, but a small pool of health; they often take the middle lane and prepare to assist teammates.</td>
<td>25.8</td>
</tr>
<tr>
<td>marksman</td>
<td>Marksman have a small pool of health but can cause heavy damage; they need protection and may need time to grow.</td>
<td>25.7</td>
</tr>
<tr>
<td>assassin</td>
<td>Assassins are explosive and control the pace, but the health pool is often small; they are usually the leading role in a team.</td>
<td>11.5</td>
</tr>
<tr>
<td>support</td>
<td>Supports always aid teammates; they absorb damage while disrupting opponents by stunning and displacing them.</td>
<td>9.9</td>
</tr>
</tbody>
</table>

---

3Honour of Kings is developed and published by Tencent, which is one of the largest Internet companies and also the largest game service provider in China.
We first investigate the effect of team composition on whether a team wins or not, a direct measure of team performance. Although the matching system in Honor of Kings is designed to balance the ability of two teams, we observe that the winning rate of most team compositions is significantly lower than 50%, indicating that some role combinations cannot effectively work with each other.

Although two teams are with similar “skills”, there exist winning and losing team compositions (Fig. 1a). Since we have 126 different team compositions in total, it is straightforward to estimate the winning rate of each team composition using the fraction of winning games. Fig. 1a shows that the winning rate of different team compositions spans a wide range, from 8.3% to 53.6%. It seems that some team compositions are doomed to lose. In addition, the figure also shows the cumulative distribution of #teams for each team composition, suggesting that most losing teams do not occur frequently. Notice that a similar plot will also be made for surrendering and abusing and we will see that the CDFs present very different shapes.

The fluctuation of winning rate between different team compositions among all games may not be directly due to the effect of team compositions; another possible explanation is the ability of individual players: only relatively bad players choose certain bad team compositions. Although a matching system is designed to balance the ability of the players in two teams, it is difficult to always make sure that two teams exactly have the same ability. Fig. 1b shows the distribution of the rank level gap (δ) between two teams. To calculate δ, we sum up the individual rank level in a team, and subtract the result of losing team from that of winning team.

<table>
<thead>
<tr>
<th>Team compositions</th>
<th>Win rate</th>
<th>Used frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>team compositions with the highest winning rate</td>
<td>- - - - -</td>
<td>53.6%</td>
</tr>
<tr>
<td>- - - - -</td>
<td>53.2%</td>
<td>2.0</td>
</tr>
<tr>
<td>- - - - -</td>
<td>52.6%</td>
<td>14.9</td>
</tr>
<tr>
<td>team compositions with the lowest winning rate</td>
<td>- - - - -</td>
<td>17.4%</td>
</tr>
<tr>
<td>- - - - -</td>
<td>8.3%</td>
<td>7.0 × 10^{-3}</td>
</tr>
<tr>
<td>team compositions that are most used</td>
<td>- - - - -</td>
<td>52.6%</td>
</tr>
<tr>
<td>- - - - -</td>
<td>52.0%</td>
<td>13.3</td>
</tr>
<tr>
<td>- - - - -</td>
<td>48.5%</td>
<td>11.6</td>
</tr>
</tbody>
</table>

Table 3: Example team compositions and winning rates.
4 TEAM COMPOSITION AND SURRENDER

We examine the mean and the standard deviation of surrender probability across team compositions conditioned on game duration in Fig. 2a. We make two observations. First, the surrender probability monotonically decreases as game duration increases, partly because a team is only allowed to surrender starting from the 6th minute and games end in 6-8 minutes mostly because a team surrenders. Second, the standard deviation across team compositions first increases and then decreases as game duration increases, indicating that in late games every team surrenders with similar probability to signal game over. Based on Fig. 2a, we define three game stages: early games (t ≤ 11), where the surrender probability has great mean and great std; late games (t > 16), where the surrender probability has small mean and small std; middle games (11 < t ≤ 16), the middle part.

Having established the three types of games, we hypothesize that surrender in early games is the most indicative of team tenacity and varies across team compositions, while surrender in middle and late games is more of a formality and should not depend on team compositions. To study that, we sort team compositions by surrender probability in early games in Fig. 2b. According to the same order, we show the surrender probability in middle games and late games, as well as the cumulative distribution function of #teams. Consistent with our hypothesis, the surrender probability of different team compositions span a wide range from 33.6% to 84.6% in early games, but is pretty stable in middle games and late games. The CDF of #teams presents a different shape from that in Fig. 1a: commonly used team compositions are neither the most tenacious nor the least tenacious.

Diverse teams tend to be tenacious (Fig. 2c). Role diversity influences surrender probability in early games, especially when the number of roles is large. It seems that diverse teams with five roles are the most tenacious in early games. In comparison, the influence on middle and late games is minimal.

A weak team can still be tenacious (Fig. 3). We have shown that diverse teams tend to both perform well and show tenacity in adversity. One concern is that these two measures are correlated and tenacity is simply a side effect of team strength (winning). This hypothesis suggests that winning team compositions should be less likely to surrender. To further understand this issue, we study the correlation between surrender probability and winning rate for different team compositions. Fig. 3a shows that there is indeed a negative correlation between winning rate and surrender probability vs. team composition.

Figure 3: Surrender probability vs. team composition. Fig. 3a shows results for all team compositions. Fig. 3b shows that diverse teams are less likely to surrender in early games (error bars represent standard errors).
probability in early games, while little correlation in middle and late games. However, it seems that the correlation between winning rate and surrender probability in early games is dominated by a few outliers with very low winning rates. Given that we know that these teams only take a small fraction of all games in §3, we filter infrequent team compositions that appear less than 10,000 times and find that in the remaining team compositions (98.7% of all teams), surrender probability is not correlated with winning rate in all game stages (Fig. 3b), suggesting that team tenacity is almost independent of team strength/performance.

5 TEAM COMPOSITION AND ABUSIVE LANGUAGE USE

Our final measure of team effectiveness is concerned with the rapport in a team during the game. We use abusive language use to capture the rapport. Toxic behavior in online communities and gaming has received significant interests from our research community recently [11, 12, 27]. Here we provide the first systematic study on the effect of team composition on team-level abusive language use.

5.1 Team-level Abusive Language Use

Team compositions vary in abusing probability (Fig. 4a). In our dataset, we have a label of whether a message uses abusive language for all text messages based on a dictionary-based method officially used by Tencent. A team abuses if any player in that team abuses. For each team composition, we define abusing probability as the fraction of games that this team composition abuses.

We find that similar to winning and surrendering, team compositions vary in abusing probability, ranging from 28.7% to 56.2%. The cumulative distribution function of #teams looks much more similar to surrendering in Fig. 2b than winning in Fig. 1a: most commonly used team compositions are neither the most nor the least abusive. We will investigate further the interaction of individual-level abusing and team-level abusing in §5.2.

Losing teams are more likely to abuse. We hypothesize that losing teams are more likely to abuse because winning usually brings positive team morale, while losing leads to frustration and dissatisfaction [27]. This is indeed the case, as we examine the correlation between abusing probability and winning rate by the slope of abusing probability with respect to winning rate is -0.41 (p-value is 5.2 × 10^{-4}, and we omit the figure due to the space limitation). Therefore, it is important to distinguish abusing probability between winning teams and losing teams.

Diverse teams tend to abuse more when losing and abuse less when winning (Fig. 4b). We further explore the effect of role diversity on team-level abusing. We observe different trends in winning teams and losing teams. When a team wins, team diversity is associated with low abusing probability; but it becomes the other way around if a team loses.

5.2 Individual-level Abusive Language Use

Different from winning and surrendering, abusive language use is an individual behavior. It provides a great opportunity to understand the effect of individual composition on individual behavior and shed light on the “situation vs. personality” debate [17, 23]. Note that we conduct this part of experiments on players who have played at least 20 games to ensure sufficient samples for statistics.

Figure 4: Team composition and abusive language use. Fig. 4a presents sorted abusing probability for different team compositions and its corresponding cumulative distribution function in #teams. In Fig. 4b, x-axis represents #roles in a team, and y-axis represents team-level abusing probability.

A team does not equal the sum of individuals (Fig. 5a). We define individual abusing probability based on the fraction of games that a player abuses in. Because of linearity of expectation, the total number of expected abusing players in a team is simply the sum of each player’s abusing probability:

\[ \mathbb{E}(\text{#abusing players in a team}) = \sum_{v \in \text{Team}} P_{\text{abuse}}(v) \]  

where \( P_{\text{abuse}} \) is estimated separately when a player wins or loses because abuse is associated with losing. This expectation is only valid if all players act independently in a team, but studies on team-work have shown that a team does not equal the sum of individuals [7]. Therefore, we examine the discrepancy between observed and expected values to understand how team composition influences individual players. This analysis is done on games in the third week since it requires historical information of every team member.

Figure 5a shows the difference between observed and expected values for winning and losing teams, where \( \text{win}/\text{lose diff.} \) refers to the disparity between observed and expected number of abusing players(\( \mathbb{E}(\#\text{abusing players} - \mathbb{E}(\#\text{abusing players}) \)). For most team compositions, the difference between observed values and expected values is not zero, indicating that individuals abuse differently depending on team compositions. In particular, in the most commonly used teams (when the CDF grows quickly on the right of the plot), individuals are more likely to abuse than expected when losing.

Players who prefer leading roles are more abusive (Fig. 5b, 5c, 5d). There always exists a leader or a major contributor in a team. In Honor of Kings, assassins usually take this role in a team because of their explosiveness: assassins can carry the team and control game pace. A failed assassin may lead the team to lose.

Fig. 5b shows that assassins are more likely to abuse than other roles. We call this the “abusive assassin” phenomenon. One natural question arises: are assassins more abusive because abusive players tend to choose assassins, or players become more abusive when choosing assassins? To answer this question, we compare the abusing probability of the same roles chosen by different players and the same player choosing different roles. We define the experienced role for a player if that player chooses a role frequently, i.e., playing in more than 50% of games. This procedure identifies experienced roles and sheds light on the “situation vs. personality” debate [17, 23].
assassin players, experienced warrior players, etc. Fig. 5c presents the abusing probability of experienced assassin players when they choose each role (the same player choosing different roles). We observe that these experienced assassin players are much more likely to abuse than other players no matter what role they choose.

Fig. 5d examines the alternative hypothesis: players become more abusive when choosing assassins. We compare the same player’s abusing probability when they choose assassins with choosing other roles, grouped by their experienced roles. We find that players do not become more abusive when choosing assassins. If anything, experienced mages and experienced marksmen are actually more abusive when they choose roles other than assassins.

Overall, our results in Honor of Kings support the hypothesis that abusive players tend to choose assassins, the leading role in a team, and players do not become abusive when choosing assassins.

6 RELATED WORK

Team formation. Researchers from the data mining community have formulated the problem of team formulation as a constrained optimization problem based on each individual’s skills and their social networks [2, 3, 28, 29]. For instance, Li et al. [29] study a family of problems in team enhancement including team member replacement, team expansion, and team shrinkage.

Social/team roles. Digitalization of human traces have increasingly made implicit social/team roles explicit and enabled large-scale study on theories of social/team roles [6–8, 15, 42], e.g., Yang et al. [43] study how social roles influence the process of online information diffusion. In the context of gaming, Stetina et al. [38] examined problematic gaming behavior and depressive tendencies among people who play different types of online-games.

Online gaming. Many researchers have recognized that online gaming can serve as a platform for studying individual/team behavior and communities [19, 25], where players not only develop individual skills but also coordinate and communicate with others [18, 21, 22, 30, 40]. Among our three measures of effectiveness, team performance is the most heavily studied topic, e.g., Pobiedina et al. [34, 35] explore factors that influence player’s performance in Dota 2, and [1, 14] demonstrate that different role combinations may affect team performance. Furthermore, gamification can also potentially play an important role in education and scientific discovery [24, 33]. In addition to these exciting opportunities, online gaming leads to issues such as addiction and depression [26, 38].

7 CONCLUSION

In this paper, we study the effect of team composition in the largest MOBA game: Honor of Kings. We quantitatively show the varying effects of team composition on team performance (winning), tenacity (surrender), and rapport (abusive language use); although diverse teams tend to perform well and show tenacity in adversity, they are more likely to abuse when losing. The double-edged influence of team diversity suggests the importance of balancing team composition. We also examine how team composition influences individual behavior in abusive language use. In addition to showing that a team is not the sum of independent individuals, we contribute to the “situation vs. personality” debate and show that assassins abuse more because abusive players tend to choose assassins instead of players becoming abusive when choosing assassins. Our work suggests that the gaming environment may be improved by adjusting team matching and preventing players from using team compositions that correlate with increased abusive language use.

Limitations. Our study is observational and can be strengthened through experimental studies. Our work is also limited by the data that we have access to. Due to privacy issues, we do not have sensitive individual information like location and consumption record. Moreover, although Honor of Kings is the largest MOBA game, the selection bias in our data may limit the generalizability of our findings.

Future directions. Several promising directions arise from our work. It remains challenging to develop a holistic model that captures the interaction between individual members, learns novel representations for role combinations, and makes accurate predictions for effectiveness metrics to validate and deeply understand our findings in this paper. Furthermore, although MOBA games provide an ideal environment for understanding the effect of team composition, it is important to generalize our findings in other scenarios, e.g., with implicit roles beyond gaming.

Acknowledgments. The work is supported by NSFC (61702447), the Fundamental Research Funds for the Central Universities, and a research funding from HIKVISION Inc.
REFERENCES


