Title
Do Men Advance Faster Than Women? Debunking the Gender Performance Gap in Two Massively Multiplayer Online Games

Permalink
https://escholarship.org/uc/item/4qh1m371

Journal
Journal of Computer-Mediated Communication, 21(4)

Authors
Shen, C
Ratan, R
Cai, YD
et al.

Publication Date
2016-07-01

DOI
10.1111/jcc4.12159

Peer reviewed
Do Men Advance Faster Than Women? Debunking the Gender Performance Gap in Two Massively Multiplayer Online Games

Cuihua Shen
Department of Communication, University of California, One Shields Ave, Davis, CA 95616

Rabindra Ratan
Department of Media and Information, Michigan State University, 404 Wilson Road, Room 428, Communication Arts and Sciences Building, East Lansing, MI 48824

Y. Dora Cai
National Center for Supercomputing Applications, University of Illinois, 1205 W Clark Street, Urbana, IL 61801

Alex Leavitt
Annenberg School for Communication & Journalism, University of Southern California, 3502 Watt Way, Los Angeles, CA 90089

Prior research on digital games illustrates a perceived gender gap in participation and performance, suggesting men as playing more and better than women. This article challenges the gender gap using longitudinal behavioral data of men and women in 2 MMOs in the United States and China. Results show that women advance at least as fast as men do in both games. Thus, perceived gender-based performance disparities seem to result from factors that are confounded with gender (i.e., amount of play), not player gender itself. We conclude that the stereotype of female players as inferior is not only false, but also a potential cause for unequal participation in digital gaming.

Keywords: Gender, Online Games, MMO, Gender Gap, Performance.

doi:10.1111/jcc.12159

Digital games are one of the most popular and fast-growing entertainment media in today's media landscape, with an estimated global market size of $93 billion (Gartner, 2013). The broad appeal of games has brought about a gradual demographic shift. According to the Entertainment Software Association,
women now constitute nearly half (41%) of the gamer population in the United States (2016), although this shift is larger in “casual” rather than highly competitive genres, e.g., first-person shooters or sports games (Trepte, Reinecke, & Behr, 2009).

Women's increased presence, however, has not been enthusiastically embraced in gaming culture. They are still perceived as minorities, with different or inferior gaming-related motivation, expertise, level of participation, preferences, and genre choice than male players (Hartmann & Klimmt, 2006; Kafai, Heeter, Denner, & Sun, 2008; Lucas & Sherry, 2004; Ratan, Taylor, Hogan, Kennedy, & Williams, 2015; Williams, Consalvo, Caplan, & Yee, 2009). Anecdotal and research accounts abound that gaming communities and the game industry are often hostile to women, fraught with gender stereotypes, sexism, and harassment (Fox & Tang, 2014; Fullerton, Fron, Pearce, & Morie, 2008; Gray, 2012; Ivory, Fox, Waddell, & Ivory, 2014).

Overt gender disparity in gaming has motivated numerous empirical studies, but few go beyond traditional observational studies to leverage longitudinal behavioral data (for an exception, see Williams, et al., 2009). Investigating the observed gender differences in online gaming—or rather, if such differences exist—is especially relevant for communication technology scholars. Such investigations not only inform how in-game behaviors vary by gender, but also highlight the many gender-based nuances associated with technology uses and practices. This is especially important given that general gendered performance disparities (e.g., on quantitative standardized tests) are biased against women and do not represent true differences in ability (Walton & Spencer, 2009). Further, it is argued that participation in digital games could promote a variety of cognitive, motivational, and social outcomes (Granic, Lobel, & Engels, 2014), from interest in STEM fields (Turner, 2014) to spatial cognition skills (Subrahmanyam & Greenfield, 1994). It could also provide an entry point for girls and women into computing-related fields (Cassell & Jenkins, 1998). Thus, gender gaps in games may perpetuate gender disparity in distributing the many benefits associated with gaming.

This article presents two studies that address one question: Do men advance faster than women in massively multiplayer online (MMO) games? Common stereotypes, as well as previous academic research, suggest that men outperform women, but few empirical studies specifically test this presumed gender performance gap. Using gender role theory as a framework to contextualize the analysis, this article presents the first granular examination of player performance by gender based on longitudinal behavioral data from game servers. Men and women's character advancement speed was examined using mixed-effects models. Our tests in two MMOs in two distinct cultures, EverQuest II in North America and Chevaliers’ Romance III in China, generated the same finding: women advance at least as fast as men do.

Theory and Prior Research

Gender Role Theory
Gender disparity is widely observed in many meaningful realms of society. For example, in the workplace, women earn less income and accrue less wealth than men in nearly every occupation (AAUW, 2014), due at least partially to bias in performance assessments (Reskin & McBrier, 2000). Gender role theory explains the cause of this disparity between women and men and could provide a general framework for the performance gap in games. Gender roles refer to “shared expectations that apply to individuals solely on the basis of their socially identified sex” (Eagly & Karau, 1991, p. 686). Gender roles consist of both descriptive norms, which are shared expectations about ways in which men and women actually behave (i.e., gender stereotypes), and injunctive norms, which are shared expectations about what men and women ideally should do (Eagly & Karau, 2002). The injunctive aspect of gender
roles thus exerts influence through 1) internalization of their ideal selves, and 2) external (social) reinforcement of role-conforming behaviors and disapproval of role-violating behaviors. Gender disparity thus propagates through socialization processes, starting as early as parental encouragement of children's gender-typical behaviors (Lytton & Romney, 1991) and continuing through the reinforcement of gender stereotypes from peers (Carter & McCloskey, 1984), teachers (Hilliard & Liben, 2010), and media (Dill & Thill, 2007; Gerding & Signorielli, 2014).

A key element of gender role theory is that most beliefs about women and men's differences focus on communal and agentic attributes (Eagly & Karau, 1991, 2002). Women are often associated with communal characteristics, such as being socially oriented, assistive, nurturing, kind, and sympathetic. Men, on the other hand, are often associated with agentic characteristics, such as being task-oriented, competitive, ambitious, and independent (Eagly & Karau, 1991, 2002; Kidder, 2002). Placing women and men at different ends of the communal-agentic spectrum has pervasive effects. For example, women are considered more suitable for “pink-collar” (i.e., feminine) jobs, such as nurse, teacher, secretary, etc., which are devalued compared to the work men do (Britton, 2000). Women also face greater hindrances to ascending in their career paths, such as becoming leaders (Eagly & Karau, 2002).

Research on Gender and Games

The communal and agentic characteristics of women and men, respectively, have found extensive support in video game genre preferences, play styles, motivation, and experiences. Boys are found to enjoy competition, while girls prefer cooperation (Cassell & Jenkins, 1998). Accordingly, recent research suggests that casual game players are predominantly women, while players of shooter and role-playing games are predominantly men (Nielsen, 2009; Trepte, et al., 2009). Women are also reluctant to self-identify as “gamers,” due in part to a self-perceived lack of expertise and time commitment, or less interest in “hardcore” games that typically define the gamer identity (Shaw, 2012). Hartmann and Klimmt (2006) found that, compared to men, women were less attracted to games with competitive elements, violent content, or a lack of social interaction opportunities. For MMOs that allow a myriad of activities in the game world, studies have shown that men and women’s differences in motivations also conform to gender stereotypes. Yee’s (2006) survey of MMO players found that men are more likely to play for achievement-oriented reasons, while women are motivated more by social reasons, which was confirmed in numerous studies (Jansz, Avis, & Vosmeer, 2010; Poels, De Cock, & Malliet, 2012; Williams, et al., 2009). Women also are more likely to choose socially oriented, supportive, and assistive character classes in MMOs, such as Priests (Shen, 2014). Finally, compared to women, men have more experience playing games (Williams, et al., 2009) and spend more time playing in general (Lucas & Sherry, 2004).

The gender gap observed in player behavior also extends to other aspects of gaming and games. For example, female characters are overrepresented in casual games (which are found to be more appealing to women), but are underrepresented in other genres, such as MMOs, which are also less liked by women (Wohn, 2011). Kafai (1998) reported persistent gender differences in design choices when boys and girls learned to make their own educational games. For example, compared to girls, boys were more likely to use violent feedback for wrong choices. Scholars believe that the gender gap within games are at least partially attributable to gender disparity in the game industry, which is manifested in misconceptions such as “women don’t play games,” poor working conditions, and sexist workplace culture (Fullerton et al., 2008).

Despite the recognized gender differences in gaming, we are still unsure whether gender is a causal or confounding variable. In researching competitive e-Sports games where female professional gamers participate the least compared to other video game genres, Taylor argues that women’s marginalized status is not a result of skills or commitment but instead is due in part to “an imagined difference between
men and women … [which] remains a persistent myth” (2012, p. 119). An ethnographic study showed that the so-called gender gap in play styles—that women prefer to assume assistive roles while men prefer competition—is in reality a difference between novices and experts: once women secure authentic access to games and familiarize themselves with gameplay, they start to take up “expert” positions and play competitively as much as men do (Jenson, de Castell, & Fisher, 2007). Women who play League of Legends are found less confident in their abilities than men, despite controlling for level of skill, and women who play with a romantic partner (which is quite common) are more likely to choose assistive roles in the game than women who do not (Ratan, Taylor, et al., 2015). Similarly, a study examining the cognitive strategies of men and women in learning to play video games revealed that prior experiences with games, rather than gender, influenced their learning strategies (Blumberg & Sokol, 2004). More experienced players tend to adopt internally oriented strategies (e.g., reading a manual), while less experienced players tend to rely on the help of others. Gender, as these studies show, is a correlate of experience rather than the true cause of gameplay differences.

In summary, research has indicated that games, especially noncasual ones, are still a stereotypical male activity. Consistent with both descriptive and injunctive gender norms (Eagly & Karau, 2002), women—compared to men—are found to play less; seek out more casual, socially-oriented game genres; and prefer social, supportive, and less competitive elements and roles. The gendered behavioral patterns contribute to a self-fulfilling cycle: the less women are attracted or committed to competitive gaming, the less experienced or worse-performing they become compared to men (Brown, Hall, Holtzer, Brown, & Brown, 1997). The gender performance gap then reinforces the “men are better at games” stereotype, which leads to decreased competence, beliefs, motivation, and performance for women (Chan, 2008).

To date, no study has empirically tested the gender gap in game performance in natural environments, due in part to the lack of objective measures. Some prior studies measured performance using simple game score after completing a task (e.g., finishing a car race; Chan, 2008; North & Hargreaves, 1999), while others measured the time taken to score a set amount of game points (Bösche, 2009; Brown, et al., 1997). However, in the above cases the games provided only one-dimensional, unambiguous tasks and were played for a short period in well controlled laboratory settings. This article therefore presents one of the first comparisons of men and women’s performance in MMOs, where a multitude of tasks and character development trajectories are available to players. Appealing to a wide variety of demographic groups globally, MMOs operate from multiple servers which automatically collect massive behavioral data, making them ideal sites to study gender and performance using longitudinal analysis. We operationalize game performance as a function of character advancement (leveling) and voluntary play time—which can be confounded with gender—and tests whether there are true gender differences (RQ). It is also the first study using server-collected longitudinal data from two games and national cultures.

Due to differences in data structure and availability, our research question was operationalized slightly differently in the two studies (see “Measures” section of each study for details). The first study, based on EverQuest II’s (EQ2) combination of temporal event data and aggregate data, operationalizes performance as the overall speed of character advancement. When two characters have the same amount of total play time since joining EQ2, the higher-level character performs better. The second study, taking advantage of finer-grained, temporal event data in Chevaliers’ Romance III (CR3), operationalizes performance as the exact amount of time spent leveling from \( n \) to \( n + 1 \). When two characters both go from \( n \) to \( n + 1 \), the character who spends less time doing so performs better. This methodological difference is akin to the ways in which people judge swimmers’ performance in a race: Spectators derive performance rankings by looking at who is swimming ahead while holding time constant (EQ2). The referees, by contrast, rank swimmers’ performance by the amount of time spent finishing the lap,
holding distance constant (CR3). Despite the difference, both studies consider performance as the speed of character advancement, controlling for other extraneous variables.

**Study I: EverQuest II**

**Data**

Launched in 2004, *EverQuest II* (EQ2) is the sequel of the successful *EverQuest*. Similar to more popular MMOs on the market, such as *World of Warcraft*, EQ2 has various game mechanisms to foster social interactions, including player-to-player trade, chat, collaboration, mentoring, and the guild system. Players have access to 24 different character class choices, which are endowed with different skillsets and abilities. The character class system ensures that individual players must collaborate with others, preferably those belonging to different character classes, in order to build a well-rounded team. This diversity of classes is especially important in later stages of the game, where quests and dungeons become too difficult for individual players to tackle alone. To effectively connect players with others, EQ2 has a “LFG” (looking for group) function where players can specify their needs when searching for teams to join or other team members to recruit. A more persistent form of social connection lies in the guild system, which allows players to create and sustain semipermanent communities. The main function of guilds is to provide a space for socialization and to coordinate various in-game activities among like-minded players (Ducheneaut, Yee, Nickell, & Moore, 2007; Williams et al., 2006).

The developer of EQ2, Sony Online Entertainment, provided the research team access to the behavioral logs of EQ2 collected during an extended period of time in 2006. Player variables, including character class, guild membership, accumulated play time, and character level, were taken as a snapshot in September 2006. Player activities, including chat, trade, and collaboration records, were collected through January to September 2006 (for an in-depth description of EQ2 data, see Shen, 2014; Williams, Yee, & Caplan, 2008). In order to align the snapshot of player variables (cross-sectional) with player activities (longitudinal), we selected the last week of behavioral log data (September 4 – 10, 2006) as the time window for this study. EQ2 operates on numerous servers, i.e., parallel game worlds. Characters are only allowed to interact with others on the same server. This study uses data from the Guk server because it represents the most common server type on EQ2: player-versus-environment (PvE), where players combat nonplayer characters (monsters) instead of other players. Only active players, defined as those who logged into EQ2 at least once during the 7-day time window, were included in our analysis.

**Measures**

**Gender**

The server log contains information on players’ self-identified demographic gender (as opposed to avatar gender). This information was voluntarily reported by the player at the time of profile registration. Gender was coded as a dichotomous variable (male = 1).

**Total play time**

The server log records the total time (seconds, transformed into hours) a character had spent playing the game, from character creation to 10 September 2006.

**Performance**

Although various behavioral metrics may indicate players’ performance in MMOs, speed of character advancement (leveling), was chosen as the main performance metric. This was achieved by including
play time as a control in models predicting player level.\textsuperscript{1} Holding play time constant, characters who reached a higher level are judged to have performed better.

The speed of character advancement (level) was chosen as the performance measure for several reasons. First, character advancement is defined by MMOs themselves and widely acknowledged among gamers as the primary goal for various in-game activities (e.g., combat, questing, and crafting), at least before hitting the level cap (Steinkuehler & Duncan, 2008; Yee, 2001). Second, unlike other game-specific performance metrics, character level is widely used across different MMOs and game genres, making cross-game comparison possible. It is also the metric most visible to players themselves. Lastly, we found character level highly correlated with other alternative performance metrics including the character’s health ($r = .93, p < .001$), power ($r = .93, p < .001$) and total number of nonplayer character (NPC) kills ($r = .69, p < .001$).

In order to represent players who pursue traditional character advancement goals as well as those who focused on secondary character advancement, we included both character level and tradeskill level. For both, data reflected the highest level earned by 10 September 2006, the last day of data collection. Character level is achieved through accumulation of experience points, which are earned by completing quests and killing monsters. Like character level, tradeskill level is also achieved through accumulation of tradeskill experience points, earned by completing specific tradeskill quests. Tradeskill level is often coupled with character level but not necessarily so. Our data showed a strong correlation between the two ($r = .54, p < .001$). The minimum level is 1 and the maximum level at the time of study is 70 for both character and tradeskill levels (the level cap was lifted several times after data collection). Because characters stop progressing after they reach the level cap, all level 70 cases were removed prior to analysis.

As performance (the slope) was determined by both character/tradeskill level and play time, the interaction of play time by gender, if significant, would indicate that men and women differ in character/tradeskill level given the same play time.

Character class

EQ2 has 24 character classes, which have distinct and complementary skillsets. Such differences allow some classes to outperform others under certain circumstances. For example, priests specialize in health restoration but are relatively weak in defense and offense. As a result, priests fare better assisting others in collaborative tasks than fighting on their own. Prior research suggests that character class choice may be correlated with gender, a pattern also found in the EQ2 dataset (see Shen, 2014 for detailed distribution by gender). Therefore, character class was included as a random effect.

Account ID

EQ2’s subscription model allows every paying account to create multiple characters. Most variables (except demographic gender) used in our study are at the character level instead of the account level. Therefore, independent characters from our dataset may come from the same account. In order to address the one-to-many relationship between account and characters, we included account ID as a random effect in all models.

Guild ID

The log data provided information on players’ current guild membership during our time window in the form of a unique Guild ID (unguilded characters have a null value). As guilds vary greatly in their own advancement goals, play styles, resources, and inner dynamics, which guild a player belongs to has a nontrivial influence over individual performance. For example, a critical dimension to categorize
guilds is how much they focus on high-end raiding (organized combat) versus casual gaming (Shen, 2014). Raiding guilds devote most of their resources to the coordination of raiding events among achievement-oriented players, while casual/social guilds focus more on building social bonds among like-minded members through gameplay. A study of *World of Warcraft* guilds showed that many guild-level factors, such as guild size and overall schedule compatibility of its members, contributed significantly to the advancement of guild members (Ducheneaut, et al., 2007). Further, research suggests that there are notable differences in the type of guild which men and women belong to and the ways they behave and are treated within (Williams, et al., 2006). To control for the influence of guild membership on performance, especially as potentially related to player gender, we included guild ID as a random effect for guilded characters.

**Analysis**

We used linear mixed effects models to analyze our data, using the "lme4" package in R (Bates, Maechler, Bolker, & Walker, 2014). Player gender, total playtime, and their interaction term were included as fixed effects, while character class, account ID, and guild ID (if applicable) were included as random effects.

Four models were estimated: two with character level as the dependent variable and two with tradeskill level as the dependent variable. Within each pair, one model focused on characters unaffiliated with any guild, while the other focused on guilded characters only, including guild ID as a random effect. Each model was estimated hierarchically: First, we included only random effects; then, added total playtime as the only fixed effect; then, total playtime and gender in the third model; and finally, the interaction term in the fourth model. The *p* value of all parameter estimates was calculated using the Kenward-Roger method, which was considered conservative among alternatives (Kenward & Roger, 1997).

**Results**

Among all players reporting valid demographic gender information in the server log, 17% were women and 83% were men. In the one-week window we selected for our study, 9483 players were active on the server. Among them, 18% were women and 82% were men, reflecting a sample gender ratio comparable to the complete database. This sharp gender difference is consistent with previous MMO studies (Griffiths, Davies, & Chappell, 2003, 2004; Yee, 2006).

Our sample showed that men (*M* = 33.71, *SD* = 9.90, *t* = 6.35, *p* < .001) were younger than women (*M* = 35.75, *SD* = 11.57), which is also consistent with earlier studies (Williams, et al., 2009; Yee, 2006). Both men and women were devoted players, with men committing an average of 712.55 hours of total playtime (*Median* = 243.16, *SD* = 1093.74), and women 681.74 hours (*Median* = 209.63, *SD* = 1047.26), but the difference was not statistically significant. Within our sample, men had slightly higher average character level (*M* = 40.96, *Median* = 40.00, *SD* = 21.15, *t* = 5.52, *p* < .001) than women (*M* = 37.62, *Median* = 34.00, *SD* = 21.18), but the difference in tradeskill levels was not significant. About 70% of both men and women were affiliated with a guild during our time window (see Table 1).

To answer our research question on men and women’s performance in EQ2, we estimated four mixed effects models predicting either character level or tradeskill level. As shown in Tables 2 and 3, total playtime had a positive effect on both character level and tradeskill level. There was a significant gender main effect on character level but not tradeskill level, showing that men on average reach higher character level than women. The interaction term of total playtime and being male (female as the comparison group) had a negative and significant impact on character level for nonguilded players (Coefficient = −0.008, *p* < .001, see Table 2) as well as guilded players, although the impact only approached significance (Coefficient = −0.001, *p* = .059). The interaction term of total playtime and being male had no effect on
Table 1  Descriptive Statistics for EQ2 Player Sample

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Player age</td>
<td>6448</td>
<td>33.71</td>
<td>33.00</td>
</tr>
<tr>
<td>Character level</td>
<td>6448</td>
<td>40.96</td>
<td>40.00</td>
</tr>
<tr>
<td>Tradeskill level</td>
<td>6448</td>
<td>23.89</td>
<td>20.00</td>
</tr>
<tr>
<td>Total play time</td>
<td>6448</td>
<td>712.25</td>
<td>243.16</td>
</tr>
<tr>
<td>Guild status (yes = 1)</td>
<td>6998</td>
<td>0.71</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note. p < .05. **p < .01. ***p < .001.

Table 2  Mixed Effects Models Predicting EQ2 Character Level

<table>
<thead>
<tr>
<th></th>
<th>Guided players (N = 4850)</th>
<th>Non-guilded players (N = 1759)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>27.573</td>
<td>16.226</td>
</tr>
<tr>
<td>Coef.</td>
<td>S.E.</td>
<td></td>
</tr>
<tr>
<td>Total play time</td>
<td>0.017</td>
<td>0.021</td>
</tr>
<tr>
<td>Male</td>
<td>3.688</td>
<td>2.900</td>
</tr>
<tr>
<td>Total play time X Male</td>
<td>–0.001</td>
<td>–0.008</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>–19266.800</td>
<td>–6997.700</td>
</tr>
</tbody>
</table>

Note. Random effects include character class, player account ID, and guild ID (for guilded players only). Female is the comparison group. All top level (level 70) characters were removed from analysis. Kenward-Roger method was used to calculate p values for all estimates (Kenward & Roger, 1997).

trade skill level for either guilded players (Coefficient = –0.001, p = .135, see Table 3) or nonguilded players (Coefficient = 0.0002, p = .871). Taken together, these models show that women in our EQ2 sample advanced in the game at least as fast as men, and — in some cases — exceeded them.

Study II: Chevaliers’ Romance III

Data

Chevaliers’ Romance III (CR3) is a fantasy-based MMO designed and operated by KingSoft, one of the largest indigenous MMO game companies in China. Although the official user count of CR3 is not available, it has been rated as one of the top three most popular game titles in China by an independent rating agency (CGWR, 2014). Like EQ2, CR3 falls into the traditional MMO genre, with Chinese Kung Fu traditions as its narrative backdrop. Players follow a linear trajectory of progression, from level 1 to level 80 (at the time of the study). Various social activities are tightly integrated into game play, including player-to-player trading, grouping, mentoring, friending, and guild systems. Unlike EQ2, where the majority of the servers are player-versus-environment, all CR3 servers allow both player-versus-environment and player-versus-player (PvP) modes. The default mode is PvE, but in a few designated arenas, players could directly engage each other in combat. This feature arguably makes CR3 a slightly more competitive game space than EQ2.
Table 3 Mixed Effects Models Predicting EQ2 Tradeshift Level

<table>
<thead>
<tr>
<th></th>
<th>Guilded players (N = 5789)</th>
<th></th>
<th>Non-guilded players (N = 1772)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Coef.</strong></td>
<td><strong>S.E.</strong></td>
<td><strong>p</strong></td>
<td><strong>Coef.</strong></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>18.285</td>
<td>0.806</td>
<td>&lt;.001</td>
<td>9.407</td>
</tr>
<tr>
<td>Total play time</td>
<td>0.009</td>
<td>0.000</td>
<td>&lt;.001</td>
<td>0.009</td>
</tr>
<tr>
<td>Male</td>
<td>−0.342</td>
<td>0.808</td>
<td>0.672</td>
<td>0.316</td>
</tr>
<tr>
<td>Total play time X Male</td>
<td>−0.001</td>
<td>0.001</td>
<td>0.135</td>
<td>0.000</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−26576.600</td>
<td></td>
<td></td>
<td>−7123.500</td>
</tr>
</tbody>
</table>

Note. Random effects include character class, player account ID, and guild ID (for guilded players only). Female is the comparison group. All top level (level 70) characters were removed from analysis. Kenward-Roger method (1997) was used to calculate all p values.

The CR3 data used in this study consists of two parts: survey and behavioral data. In late 2011, a large-scale survey was announced on the official CR3 website, containing a URL to an external survey site. Survey participants were offered a virtual weapon desirable to all players, regardless of their character class and experience. The survey remained open for 5 weeks, resulting in a little over 20,000 responses, in which about 18,000 were valid and mostly complete (for an in-depth description of the CR3 survey, see Xiong, 2012).

In addition to the survey, KingSoft also provided partial behavioral log data collected during the periods of May 2010 – September 2010 and December 2011 – June 2012, from selected CR3 servers. Because the time period and server range of the behavioral data only partially overlapped with the survey data, we were able to connect the survey and behavioral data for a subset of around 2000 players, which were used as the main sample for this study. The log data consists of longitudinal records of player events, including leveling up, logging in/out, and in-game economic transactions, among others (it did not include aggregate data such as total play time). The behavioral log did not contain information on character attributes, such as character class or guild membership. Therefore, we were unable to include them as random effects.

Measures

Gender
The survey asked a question on players’ self-identified, demographic gender (as opposed to avatar gender). Gender was coded as a dichotomous variable (male = 1). It was the only variable from the player survey that was included in the models.

New character level
This variable captures the new character level after each level up event. When a character leveled up from \( n \) to \( n + 1 \), new character level has a value of \( n + 1 \). As studies have shown a positive correlation between new character level and time required to level up (Ducheneaut, Yee, Nickell, & Moore, 2006; Shen, 2014), we took new character level into account when comparing men and women’s time to level up.

Time to level up
The behavioral log provided time-stamped records when: 1) a character logged in, 2) a character logged out, and 3) a character leveled up to a new level (i.e., from level \( n \) to level \( n + 1 \)). We developed an
algorithm to extract such records and calculated the exact time (in seconds) a character took to level up from \( n \) to \( n + 1 \). Depending on how active a character was during the time window, a character may have experienced multiple events of leveling up. Therefore, this variable captures the marginal rate of advancement for each new character level \( n + 1 \). A significant gender effect on time to level up would indicate that the gender gap exists.

**Character ID**
The log data provided information on players’ unique character ID. Because one character could have multiple events of leveling up during our observation window, character ID was included as a random effect.

**Player Account ID**
The log data also provided information on player’s unique account ID. Like EQ2, each paying account in CR3 can create and maintain multiple characters. We included account ID as a random effect, within which character ID was nested.

**Analysis**
Similar to Study I, we used linear mixed effects models to analyze our data, using the “lme4” package in R (Bates, et al., 2014). The dependent variable was time used to level up from \( n \) to \( n + 1 \). There were 90495 observations in the dataset and each character were associated with multiple records of leveling up. Player gender and new character level were included as fixed effects, while account ID and character ID (nested within account ID) were included as random effects. The \( p \) value of all parameter estimates was calculated using the Kenward-Roger method, which was considered conservative among alternatives (Kenward & Roger, 1997).

**Results**
The survey collected around 18,000 valid responses, within which 25.51% were women. Around 2000 survey respondents also had behavioral log data for at least one time period, which then formed the basis of our mixed effects model. The gender ratio of this smaller sample is comparable to the total survey sample, with 23.34% women. Table 4 presents descriptive statistics of the CR3 player sample. Overall, men and women had quite similar profiles. Men (\( M = 23.70, t = 6.53, p < .001 \)) were slightly older than women (\( M = 23.24 \)). Women (\( M = 3.70, t = 6.55, p = .006 \)) also self-reported more hours playing the game per week than men did (\( M = 2.75 \)). Within the two periods of behavioral logs we had access to, men and women did not differ significantly with regard to their minimum character level, observed at the beginning of the time window, or their maximum character level, recorded at the end.

To answer our research question on men and women’s performance in CR3, we estimated mixed effects models predicting the time (in seconds) it took to reach each new level (Model 1 in Table 5). New level had a large and positive effect on time to level (Coefficient =103.05, \( p < .001 \)). In other words, the more advanced the character, the more time it took for this character to climb another level, which is consistent with prior MMO research. But there was no gender effect: compared to the female group, male players did not have a significant advantage in leveling speed (Coefficient =−57.419, \( p = .484 \)). To further explore whether a gender gap (or lack thereof) changed across levels, we included an interaction term of new level and gender (Model 2 in Table 5), which was significant (Coefficient =−8.289, \( p < .001 \)). The tipping point came around level 15, after which women’s leveling advantage over men reversed.
Table 4  Descriptive Statistics for CR3 Player Sample

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th></th>
<th>Female</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Median</td>
<td>SD</td>
</tr>
<tr>
<td>Player age</td>
<td>14014</td>
<td>23.70</td>
<td>23.00</td>
<td>4.34</td>
</tr>
<tr>
<td>Play time per week</td>
<td>1279</td>
<td>2.75</td>
<td>0.27</td>
<td>5.05</td>
</tr>
<tr>
<td>Minimum level</td>
<td>1639</td>
<td>3.42</td>
<td>1.00</td>
<td>12.10</td>
</tr>
<tr>
<td>Maximum level</td>
<td>1639</td>
<td>34.35</td>
<td>18.00</td>
<td>31.25</td>
</tr>
</tbody>
</table>

Note. Gender, player age, and play time per week were self-reported from player survey, while the rest were from behavioral logs. *p < .05. **p < .01. ***p < .001.

Table 5  Mixed Effects Model Predicting Time to Level Up in CR3 (N = 90495)

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>S.E.</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>−15.179</td>
<td>73.618</td>
<td>−0.206</td>
<td>0.837</td>
</tr>
<tr>
<td>New level</td>
<td>103.055</td>
<td>0.932</td>
<td>110.620</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>−57.419</td>
<td>81.929</td>
<td>−0.701</td>
<td>0.484</td>
</tr>
<tr>
<td>New level X Male</td>
<td>−8.289</td>
<td>2.107</td>
<td>−3.934</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Random effects include character ID nested in player account ID. Female is the comparison group. Kenward-Roger method (1997) was used to calculate all p values.

Discussion and Conclusion

Gender Performance Gap Debunked

Consistent with descriptive and injunctive gender norms, research on gender and games has established many gender differences in play preferences, motivation, play time, and performance. Compared to men, women are less attracted to competitive game genres and elements, have less prior experience, and spend less time playing. These differences could then contribute to women’s propensity for auxiliary game roles, weak performance, dampened interest, and further conviction that men are better suited for competitive play. However, it remains uncertain whether the performance gap stems from inherent gender differences or some other factor(s) of which gender is a correlate. This article performs a rigorous longitudinal performance analysis of men and women in two large MMOs in an attempt to offer an explanation for the consistent and pervasive gender differences.

The answer to the research question is indisputably negative: in both studies, women performed at least as well as men did. The EQ2 study used the character and tradeskill levels of each player as dependent variables, controlling for total play time and guild status. The CR3 study took advantage of finer-grained longitudinal data: The exact time a character took from level n to level n + 1 was collected and compared between men and women. Using the swimmers analogy, the EQ2 study examined who was swimming ahead (i.e., character or tradeskill level) given the same duration, while the CR3 study examined who had the shortest duration to finish a certain lap (level n to n + 1). We also controlled for random effects including character class, guild membership, and owning multiple characters from the
same account, using mixed-effects models. The rigor of our method allows us to debunk the gender gap with confidence: Based on our data from two MMOs, there is no significant gender-based performance disparity.

Our methodological advantage is especially salient in EQ2: Note that if total play time and random factors such as guild ID and character class were not controlled, it would have appeared that, consistent with gender stereotypes and prior studies, men are overall better players than women, because men achieved a higher character level ($M_{\text{men}} = 40.96$, $M_{\text{women}} = 37.62$, $t = 5.52$, $p < .001$) with approximately the same average play time ($M_{\text{men}} = 712.25$, $M_{\text{women}} = 681.74$, $t = 0.98$, n.s.). Instead, our analysis revealed a significant gender main effect, indicating that on average women may not advance their characters as far as men do. However, women’s speed of advancement, the real performance measure, is at least as fast as men’s (and in some cases, exceeds men’s).

Could our finding be partially attributed to the fact that men and women were pursuing different goals in MMOs? In both studies, game performance was measured by the speed of character advancement, but men and women may not consider character advancement an equally motivating goal. Multiple studies found that male players are more achievement-motivated and female players more socially motivated (Williams, et al., 2009; Yee, 2006), which suggests that women would lag behind men in character advancement. However, both our studies showed that women leveled up at least as fast as men did. Therefore, the motivational difference between genders only makes our finding more robust.

Implications for Future Work

Our finding has broad implications for future (and past) work. Do women play less, perform worse, and opt for less competitive genres and elements because they simply cannot play as well as men due to innate gender differences, or is it because other factors, such as prior experience, cultural expectations, and other circumstances inhibit or outright prevent them from further engagement? Our finding indicates the former is false and the latter is likely true. Controlling for all factors, women who play as much as men reach similar levels. However, for various reasons, women appear to play less or stop early, contributing to the perception that they in general do not perform as well as men. It should be noted that our results do not refute gender differences in playing particular games. Rather, it shows that the differences in men and women’s play styles and behavioral patterns found in many studies likely stem from something other than a (false) performance gap.

This finding also highlights a potential vicious cycle: Although the stereotype of female players as inferior in performance is false, the perpetuation of this stereotype likely discourages female players from playing (Brown, et al., 1997), thereby hindering their performance improvements (Chan, 2008) and thus reinforcing the confounded relationship between gender and skill. This line of reasoning is supported by Stereotype Threat Theory (Steele & Aronson, 1995), which posits that reminders of a negative stereotype that applies to a person’s demographic make that person more likely to conform to the negative stereotype. In one of the few studies on stereotype threat in gaming, Vermeulen, Núñez Castellar, and Van Looy (2014) found that women reported feeling more stress and perceived their skills as lower when they thought they were playing a game against a man (as opposed to a woman). A similarly designed study also found that gender-based stereotype threat induced in a gaming context led women to perceive some STEM fields as better suited for men than women (Ratan, Fordham, Huang, & Strayer, 2015). Together, these studies suggest that the gender stereotypes in gaming are not only false, but potentially cause unequal participation in digital gaming, thereby perpetuating gender disparity in the distribution of the many known benefits associated with gaming, such as interest in STEM fields (Granic, et al., 2014; Turner, 2014). In other words, gender disparity in gaming may occur as part of a self-propagating, vicious cycle that has negative repercussions in unquestionably meaningful contexts that exist outside of gaming.
This understanding is critical in designing a sustained effort to close the gender gap in gaming as well as technology in general. For researchers, we believe it is unproductive to document gender differences and simply attribute them to “play style” or “preference” without probing the source of such differences, as doing so may lead to hardened gender stereotypes (Jenson & de Castell, 2010). Instead, research should place emphasis onto the contexts and conditions under which these differences (or lack thereof) may occur. Such contexts may include the gendered culture of digital games/technology, entrenched expectations of how men and women should behave, peer community pressure and support, genuine access and prior knowledge, and so on. For practitioners and designers, a starting point lies in methodically surveying—and then balancing—the contexts of play for men and women. To identify, make visible, and eventually dissolve the many cultural and contextual obstacles of women’s engagement remains an important future task.

Methodologically, this article illuminates the many advantages of using behavioral data to study gender. Most past studies rely on self-reports, which are prone to random and systematic inaccuracies. Some inaccuracies may originate from gender stereotypes. For example, players tend to underestimate their play time, and women do so more than men (Kahn, Ratan, & Williams, 2014; Williams, et al., 2009). Using behavioral data could mitigate this problem in reported research. Further, behavioral data provide much-needed information on the uneven conditions of the playing field, such as acceptance into guilds and character class choices, which cast significant influence on play outcomes but were seldom acknowledged in prior studies. Longitudinal data also foreground a dynamic trend. As shown here, an effective performance measure lies in the slope (speed of progression) rather than the intercept. Even though there are many reasons why women may lag behind at the outset, a longitudinal analysis could reveal their true trajectory of advancement (Jenson, et al., 2007). For these reasons, we encourage a careful re-evaluation of past findings based solely on self-reports. Future studies should take advantage of behavioral and longitudinal data wherever possible.

Limitations
Despite the rigor of our method, it still has several limitations. First, the gender information in both studies was self-reported by players, thus its accuracy cannot be ascertained. However, as gender data was reported voluntarily and had no bearing on game play, there is no particular reason to believe players would misrepresent their gender. Second, the ecological validity is limited by our focus on MMOs. Other game genres, ranging from casual games to more challenging e-Sports games, may or may not follow the performance patterns observed here. A similar limitation lies in our focus on the U.S. and Chinese samples, which may not represent gamer populations in other cultures. However, because MMOs are highly versatile spaces where a myriad of activities are possible, and also because our tests of both traditional and alternative character advancement trajectories in two different games and national cultures yielded similar results, some generalizability can still be expected. Third, all gamers in our sample were self-selected, and they may not be representative of men and women gamers in general. Specifically, both games examined have severe gender disparity, so that female players who chose to play EQ2 and CR3 during our time window may be relatively more experienced and better-performing than the female gamer population. Future studies in various game genres and diverse cultures are needed to confirm the generalizability of findings reported here.

Conclusion
Do men advance faster than women in MMOs? Prior research found a perceived gender gap in participation and performance, suggesting men as playing more and better than women. This article challenges this gender gap through a longitudinal performance analysis of men and women in two MMOs in the
United States and China, *EverQuest II* and *Chevaliers’ Romance III*, respectively. Controlling for extraneous factors such as play time and guild membership, our results showed that women perform at least as well as men do. Perceived gender-based performance disparities seem to result from factors that are confounded with gender (i.e., amount of play), not player gender itself. The stereotype of female players as inferior is not only false, but it is also a potential cause for unequal participation in digital gaming.

**Notes**

1. An alternative was to divide level by total play time and take the quotient as a singular performance measure. However, studies have found that the marginal speed of advancement in MMOs declines with level: The higher a character has climbed in level, the more challenging it is to advance further (Ducheneaut, et al., 2006; Shen, 2014). The quotient approach, therefore, discriminates against high-level characters, because they tend to have the lowest marginal rate of advancement, while our approach avoids this problem. To check whether results were affected by the deceleration of leveling speed, we also divided our sample into seven subsets based on player level (level 1–10, 11–20, 21–30, 31–40, 41–50, 51–60, and 61–69) and conducted the same analysis on each subset. The results were similar to those based on the whole sample.

2. Note that the current study found no significant difference of total play time between male and female players, yet a previous study based on the same *EQ2* dataset found that women spent more hours per week than men (Williams, et al., 2009). These two findings do not contradict each other because the former concerns the absolute length of game play since character creation, while the latter concerns the weekly intensity. Additionally, the Williams et al. study only considered a subset of the player population, namely, those who participated in a player survey and self-reported their weekly play time, while the current study examined all active players during the observation window from server logs.

**References**


**About the Authors**

**Cuihua (“Cindy”) Shen** (http://www.shencuihua.com) is an Assistant Professor at the Department of Communication, University of California, Davis. She codirects the Computational Communication Research Lab. Her research focuses on the structure and impact of social networks in massively multiplayer online games, online communities, and social media platforms using computational social science methods.

Address: Department of Communication, One Shields Ave, Davis, CA, 95616.

**Email:** shencuihua@gmail.com

**Rabindra (“Robby”) Ratan** (http://robbyratan.com) is an Assistant Professor and AT&T Scholar at Michigan State University’s Department of Media and Information. He is also an affiliated faculty
member of the College of Education's program in Educational Psychology and Educational Technology. His research focuses primarily on the psychological experience and effects of media use, with an emphasis on video games and other interactive environments (e.g., virtual worlds, the road) that include mediated self-representations (e.g., avatars, automobiles).

Y. Dora Cai is the Senior Database Architect at National Center for Supercomputing Applications in the University of Illinois at Urbana-Champaign. She has authored or coauthored over 20 technical papers on Database Management, Data Mining, High Performance Computing, and Scientific Data Analysis. She has been involved in several national and international big data projects.

Alex Leavitt (http://alexleavitt.com) is a PhD candidate in the Annenberg School for Communication & Journalism at the University of Southern California. His research focuses on social interaction and collaboration in – and computational research methods for studying – social media platforms and multiplayer online games.