

UNIVERSITY OF LJUBLJANA

FACULTY OF ARTS

DEPARTMENT OF PSYCHOLOGY

TUULI IRINA PÖLLÄNEN

**The influence of cognitive skills and team cohesion on
player performance in Multiplayer Online Battle Arena**

Ljubljana, 2014

Abstract

The influence of cognitive skills and team cohesion on player performance in Multiplayer Online Battle Arena

Video games have a lot to offer to scientific research by enabling us to test hypotheses about expert cognition on a population that is easily accessible for web-based experiments. This study sought to explore several topics with little existing literature, by attempting to identify cognitive skills relevant to gameplay in League of Legends; by looking at how cognitive skills contribute to performance differences between experienced players, instead of focusing on novice-expert comparisons; by assessing the usability of a sports team cohesion inventory (GEQ) and the NASA task load index (NASA TLX) for game studies; and by assessing the usability of JavaScript-based web-experiments in cognitive psychology game studies.

I developed and launched a web experiment consisting of GEQ, NASA TLX, Eriksen Flanker task, mental rotation task, spatial working memory task and the Tower of London. Analysis of the results confirmed a sufficient structure and reliability of a two-factor solution of a slightly modified GEQ, and the subscales of the NASA TLX significantly predicted game performance. Analysis of the cognitive tasks revealed ambiguous results, and it appeared that cognitive skills together with task load are a significant but not an easy-to-interpret factor in player performance. Only the spatial span task performed poorly, and as the results over all are not clear from this sample, I recommend repeating the study with slightly modified instruments. Generally, the web-experiment itself functioned well, and it appears that the approach can indeed be used for cognitive psychology game studies.

Keywords: Multiplayer Online Battle Arena, performance, cognitive skills, team cohesion, web-based experiment

Table of Contents

Introduction.....	4
Why Research Video Games?.....	4
Multiplayer Online Battle Arenas and League of Legends.....	6
Intention Of The Study.....	9
Methodological Challenges.....	11
Method.....	14
Participants.....	14
Instruments.....	14
Inventories.....	15
Cognitive tasks.	19
Procedure.....	24
Results.....	27
Correlational Analysis.....	27
Regression.....	29
Discussion.....	30
References.....	33

Introduction

Why Research Video Games?

Culturally topical phenomenon influencing millions of people. With a change in attitudes and development in game design and technology, video games are now recognized as one of the most popular branch of multimedia and entertainment. In 1999, only 8% of Americans (not adolescents) admitted to playing games online, but by year 2003, the number had grown to 37%, and continued to rise (Fallows, 2004, as cited in Ducheneaut et al., 2006). As a conclusion to their large scale survey in the United States, the Entertainment Software Association (2011) estimated that at least one person in 72% of American households play video games. The average gamer is 37 years old and has been playing for 12 years, and 82% percent of gamers are 18 years of age or older. 42% percent of all players are women and women over 18 years of age are one of the industry's fastest growing demographics. In October 2010, one of the most popular multi-massive online games (MMOs) alone, World of Warcraft, created by Blizzard Entertainment, passed the line of 12 million subscribers worldwide (Blizzard Entertainment, 2010). This was the beginning of the reign of games with massive player bases, and in the recent years, with the emergence of a highly commercialised competitive video game play culture, games have become even more popular and player bases are even larger. Riot Gaming, the company that developed League of Legends – the Multiplayer Online Battle Arena game (MOBA) used in this study, reports that their game has as many as 27 million daily active players, 67 million players who are active on a monthly basis, and over 7.5 million simultaneously active players during the peak time every day (Riot Games, 2014).

Video games lend themselves to research. As a phenomenon concerning such a large proportion of the world's population, game research should be taken with appropriate rigor. Additionally, as virtual microcosms that have many parallels to the real world, games can function as a research playground - virtual laboratories or sandboxes for quasi-experiments that would be difficult or impossible to carry out in the real world. For example, Balicer(2005) used multimassive online games to model infectious disease dissemination, and in-game economies have frequently been used as a playground for research in economics (e.g. El-Shagi & von Schweinitz, 2014; Bainbridge, 2007).

Cognitive psychology is no exception in that it, too, can benefit from research on video games and players – even for research topics that have little to do with games. One example of such a research topic is team cognition. Teams are a special type of a group - a "distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal, object or mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership", (Salas, Dickinson, Converse & Tannenbaum, 1992, p. 4). Because genuine teams can be relatively difficult to come by and recruit for research on team cognition, research on computer supported cooperative work and social psychology arenas on small group behavior and decision making tend to be done on groups, not teams (Cooke, Gorman & Winner, 2006). Video game teams can offer a large pool of accessible, genuine teams for research on both areas, and additionally, can come with different levels of "teamedness", which can be useful for statistical control. For instance, in League of Legends, players can queue for ranked games in three different ways: alone or with one friend, to be paired up with random players of similar skill to form a team of five players; in a pre-made team of three players, who all know each other and queued together; or in a similar premade team of five players.

Team-based video games can provide researchers with expert teams that are often intrinsically motivated by research participation. PC game players who participate in computerized experiments are actually participating on the same platform and tools that they use when collaborating with their team. Since many team-based games contain different ranking systems - such as that for player-versus-player (pvp) performance in League of Legends - the level of performance of players participating in these games is already quantified and represented by their pvp ranking or rating. Additionally, in many games, a leveling or tier system allows for controlling the level of the players' experience in the particular game, and different games offer additional performance measures that can be used as an addition to pvp ratings. In the case of League of Legends, this could be e.g. resources gathered in game, the average proportion of own deaths in game compared to the number of enemies slain, and so on.

E-sports psychology. Finally, with the emergence of highly competitive professional gaming, video games should be researched with the same perspective as sports. Professional gamers go through rigorous training routines, practicing several hours a day, and attend

competitions similar to traditional athletes. Undoubtedly, they can benefit from techniques similar to those used in sport psychology, and the obvious differences between the tasks and performance between e-sports and normal sports only highlight the necessity of researching e-sports individually, as an addition to traditional sports psychology.

Multiplayer Online Battle Arenas and League of Legends

Multiplayer online battle arenas, also called Action Real Time Strategy games (ARTS) are a video game genre with a surprisingly long history and a vast and enthusiast community. Research on MOBA-style games has greatly increased in the recent couple of years, but the area is still relatively unexplored by academics (Ferrari, 2013). In a typical MOBA, two teams of players combat against each other, with every player controlling one character with various abilities and advantages that improve over the course of a game as the player gains experience points and purchases better items by gathering in-game resources. The objective of the game is to destroy a structure in the opposite team's base, with the assistance of friendly artificial intelligence (AI) controlled units that assist the team and march forward towards the enemy base along set paths. Throughout the game, players gain resources and experience by killing enemy players' characters or these AI-units (minions) spawned at the enemy team's base, or by destroying the enemy team's defensive structures. The resources can then be used to purchase upgrades that make the team more powerful, with the intention to overpower and beat the enemy team.

Game play in League of Legends. League of Legends follows the gameplay of a pretty typical MOBA, where the player - called the 'summoner' - controls a character - called a 'champion', and engages in games against a team of other players or AI opponents. There are currently a total of 121 champions, all of which can be classified to types, or roles, according to their strengths and weaknesses and their purpose in game play. During the game, the player's champion becomes more powerful by gaining levels and resources from killing the opposing team's champions, controlled by other players or AI, and minions that regularly spawn on the enemy team's side of the map.

Most MOBAs only have one arena that players can compete upon, and the layout for that arena typically does not change, unless it needs to be patched due to some game play balance issues. League of Legends offers its players a few maps with static layouts, each with slightly different play style. The most popular of the maps is called the Summoner's Rift

(Figure 1), which is divided into three "lanes" - the top, middle and bottom lane - which are paths followed by the spawned minions. Between the lanes there is an area called the "jungle", where a number of neutral monsters reside. The jungle is sometimes considered a lane of its own, and some champions can work as 'junglers', roaming in the jungle and collecting rewards and buffs by killing the neutral monsters, assisting freely on the other lanes. There are also two bases - one for each of the two teams - where players spawn at the beginning of the game and after dying, where they can purchase items, and where the nexus that must be destroyed to finish the game resides. Experienced players who engage in games in a fixed team typically "specialize" in one or two preferred roles or locations, with relatively little variability from game to game.



Figure 1. A map explaining the layout and the different areas in the Summoner's Rift, drawn by Narishm (2013). Reproduced with permission.

On the Summoners Rift, Twisted Treeline, and Howling Abyss maps, the game is over once one team manages to destroy a structure called the 'Nexus' in the enemy team's base. Other maps have variations of rules and objectives, such as Dominion on the Crystal Scar, where each team's Nexus weakens if the opposing team controls three or more of the five towers on the map.

League of Legends offers a few different options for game play. All players can engage in "Normal" games - unranked games against either a team of players or opponents controlled by artificial intelligence, where the game outcome does not impact players' in-game ranking, but does reward experience points - which gradually increase the player's level - and influence points, which can be used to purchase access to new champions or to make cosmetic changes to the champions the player already owns. All players begin at level 1, and gain levels by playing normal games. The benefit of gaining levels is unlocking so-called Mastery Points and Rune Page slots which allow the player to boost certain aspects of their gameplay by improving the overall strength of their champion during battles. Once the player reaches the maximum level of thirty and has access to at least sixteen champions, they are eligible for ranked games.

Ranked games are the competitive alternative of normal games - a draft mode game type, where players battle in teams and get placed on a game ladder according to their wins and losses against other players. Once they begin with ranked games, players begin with ten placement games to establish a hidden match-making rating, which is used to match them with equally skilled opponents and adjusted according to wins and losses.

Three independent play modes exist for ranked games, and so players can have a total of three different rankings. Two modes are for premade teams, where players queue in organized teams, and the third mode is for non-premade teams. For the premade teams, players can queue in a team of three people (three-versus-three), or in a team with five people (five-versus-five), and in the non-premade teams, players queue either alone (Solo-queue) or with one other person (Dual-queue), and the team is topped up to a total of five players by adding other solo- or dual-queued random players with similar match making rating. Players belong to six tiers, the lowest being Bronze, and may progress through Silver, Gold, Platinum, Diamond and Challenger, depending on their success against opponents of different skill level. Each tier, other than Challenger, is further split into five divisions. Players must win a

best-of-three Division Series to move up a division, or a best-of-five Promotion Series to move up a tier. Once the player gains a tier, they cannot lose it except by four weeks of insufficient play in the particular ranked mode.

Reaching level thirty can take a few months of casual play and due to the level requirement, most ranked players should already have sufficient experience to master the basics of the game play. The tiers and divisions can be used as a measure of game play performance, and since players' rankings are derived from PvP encounters, this measure of game performance is the outcome of direct comparisons between players and their teams. From this perspective, the differences between ranked League of Legends players can be used to research what characteristics impact the performance outcomes of individuals with a tangible level of expertise in their particular domain. Essentially, these are not so much comparisons of expert and novice players, but comparisons between players of different skill with a considerable level of experience.

Intention Of The Study

When it comes to studying cognitive skills and video game play, there is quite a large body of literature looking into video game training effects on different cognitive abilities (e.g. Maillot, Perrot and Hartley (2012); Feng, Spence and Pratt (2007); Guha, Jereen. DeGutis & Wilmer (2014); Best (2012); Green & Bavelier (2003); Green & Bavelier (2007)), and comparisons between expert and novice players as well as between players and non-players (e.g. Anderson, Bailey and West (2012); Boot, Framer, Simons, Fabiani and Gratton (2008); Dye, Green and Bavelier (2009), Green & Bavelier (2003); Green & Bavelier (2007); Colzato, van Leeuwen, van den Wildenberg and Hommel (2010); Karle, Watter and Shedd (2010)). There is much less literature on what predicts in-game performance. Some research exists on how a team's characteristics or behavior during game play relates to its performance. For example, Drachen et al. (2014) found out that the different ways in which the spatio-temporal behavior – the tendency of the players within a team to move around on the game area as a response to events that were taking place in game – vary across teams with different levels of skill. They discovered that spatio-temporal behavior of teams in a multiplayer online battle arena game (MOBA) called DotA2 is highly related to the team's skill level both in terms of changing position around the terrain, and in the distance between the characters played by the team members. Pobiedina, Neidhardt, Calatrava Moreno and Werthner (2013)

used data from DotA2 and its community to discover that the factors influencing a team's success (ranked by significance) were the number of team members that identified as each other's friends -- the social ties within the team; the team's "hero score" -- how well the chosen heroes performed when it came to their win-loss ratio or the amount of resources they could gather in game; and the team's experience score as an aggregate measure of experience of the team members.

There is much less research, again, on how cognitive skills relate to player performance. Some research exists on how cognitive skills influence team performance in non-game contexts, with the finding that individual members' greater cognitive ability is generally associated with better team performance (e.g. Heslin, 1964; LePine, Hollenbeck, Ilgen & Hedlund, 1997; Hollenbeck, Moon, Ellis, West, Ilgen, Sheppard, Porter & Wagner, 2002). This research has yet to be extended to the domain of video games, and there is barely any research on whether and how cognitive skills contribute to performance across experienced players with different performance ratings.

It is meaningful to look at contributions of cognitive skills because they bridge between behavioral factors of performance - e.g. the spatio-temporal functions described by Drachen et al. (2014) - and the social factors of performance identified by Pobiedina et al. (2013), as it could be that social processes within the team and the players' cognitive skills are what drive the performance on the behavioral level. Studies in cognitive skills and player performance are also useful to drive decision making on which cognitive skills to assess in video game training studies. A difference in a cognitive skill between players of different performance ratings implies that that particular skill is relevant to that particular game or genre. Once such a skill is identified, its use is meaningful in training studies, where researchers can then establish, whether playing the game improves that particular skill, or if performance differences occur due to underlying differences in cognitive skills irrespective of gameplay.

In this study, I wanted to explore, whether cognitive skills and team cohesion relate to performance in League of Legends players. I was interested in performance in both the solo-queue setting, where players group up with random people, and in the team-queue setting, where the players' ratings reflect how well the team members work together against other teams. I also had the intention to uncover cognitive skills that are relevant for game play in

MOBAs, as this could be used to guide instrument selection for training studies; and to assess the usability of web-based experiments for research on cognitive skills in game studies.

Methodological Challenges

There are a few methodological challenges in game studies, especially in a scenario where the researcher is interested in team-based phenomena, or in case where an experimental approach, rather than just a survey approach is necessary. One of the challenges in researching video game teams is that they can be considered a subset of distributed virtual teams. The typical mode of communication and collaboration during team games is through text and Voice Over Internet Protocol – players are rarely in the same room and collaboration happens in the virtual space. Because team members rarely occupy the same space, and because they have limited contact, it can be difficult to recruit a number of participants from the same team. Additionally, especially in Europe, the teams are often international, which has to be taken into account in instrument selection. This is more of a challenge for survey research where questionnaire validation is an issue, and could be mitigated by recruiting more homogenous teams, which in many cases might actually mean having to exclude a large chunk of the player base.

Web-based experiments. Recruiting extremely distributed teams and gaining sufficiently large, representative samples of an international participant pool is a near impossible mission for laboratory-based research without a major international operation involving multiple research groups and significant funding. However, a reasonable alternative that still allows the researcher to carry out an experimental design is simply to move the experiment to the virtual space. Web-based experiments have a steady footing in psychological science and are the perfect – if not the only – solution for experimental research on large, international player bases. Well designed web experiments have the potential to reach nearly anyone with online access and they can open doors to addressing hypotheses that simply could not be tested in a laboratory environment. But as always, the major advantage of increased scope has to be weighted and considered in light of the shortcomings and challenges of setting up and distributing web-based assessment systems. The insecurities of web-based systems are the most dire when it comes to timing accuracy – whether for stimulus presentation or response logging. Choosing research methodology, as always, depends on the nature of the research question.

Another problem in web-based research is that the participant pool is self-selected. This has more significant implications to some type of research than to other, and might even be desirable if the experimenter wishes to gather a large pool of very particular participants in their ‘native’ environment – if the research is explicitly targeted at web users. The participant pool is also relatively uncontrollable; it is difficult to absolutely exclude the possibility of multiple participation, although this can be controlled to some extent by simply asking if the participant has taken part in the study before, by generating unique user identification and preventing multiple entries with the same ID.

Drop-outs represent another issue caused by self-selectiveness in the participant pool. A significant proportion of participants halt their participation, and there is good reason to believe that the missing values introduced by partial participation are non-random by nature – whether or not someone decides to continue through a task is most likely a function of either the current task, or the previous tasks. Participants drop out because of things such as tiredness, frustration and poor performance, and so the participants who do finish the battery are likely to be characteristically different from those who dropped out, by the very traits measured in the experiment. One has to consider whether or not this is acceptable, taking into account their research problem, and whether it can be considered a fair return for the increased scope.

Another major shortcoming is reduced experimental control. Users all participate at their own locations. As an addition to unknown environmental noise, distractions, and unknown psychophysical and other user states, participants all have different operating systems, hardware, software and system loading. The experimenter can gain some basic information about the user's device, browser and operating system e.g. by extracting user agent strings from HTTP request headers, but there is no way to really know what's happening on the participant's side, and so it becomes difficult, if not impossible, to control the obscuring variance it may introduce. If the laboratory environment can be scrutinized, polished and tested until it is relatively noise-free, the web-experiment environment can only instruct the participants on how to participate in a way that minimises confounding noise.

With web-based experiments, all of the participants work on the tasks through their own computers, which means that the data may be confounded by any number of combinations of different hardware layouts, operating systems and software. Hardware considerations involve processor types and clock speeds, input and output devices with screen

refresh rates and keyboard and mouse latency, and they are the most important when it comes to reliably timing stimulus presentation and logging response times. Output devices are also important, when it comes to presenting the stimulus in the same way for all the participants. For instance, for many visual tasks, screen size and resolution might be a confounding factor. When it comes to input, different mouse types and keyboard layouts could be a psychomotorically confounding factor since giving a simple response might be more cumbersome for some participants than for others – for instance, some participants might complete the experiments on a laptop without a mouse, using only the trackpad.

Hardware and software differences also make for some interesting interactions. For instance, a number of web experiment solutions so far have been written in Flash, a decreasingly popular solution in web-development, which requires a browser plugin that many users don't have or choose to disable, and has relatively poor cross platform compatibility. Another popular language in pre-existing web-based experiments is Java, which is currently not at all the best idea for client-side applications due to security-related vulnerabilities – the US department of homeland security even issued a statement prompting all consumers to disable Java in their browsers 'unless absolutely necessary' (Perlroth, 2013). Operating system and browser constraints not only confound the data but may reduce the reach of the study – if the experiment doesn't run on the user's system, they will be prevented from participating altogether. Web-experiments can also suffer from system load (Eichstaedt, 2001), especially with users multitasking with weaker hardware, so users should be prompted to close other programs before participating. As always, compliance is not guaranteed.

Finally, depending on the research problem, these differences can inflate either alpha or beta error probability. In the typical scenario, the random noise would drown an experimental effect of small magnitude, inflating the beta error probability. An inflated alpha error probability is also possible – if high quality hardware is a factor that reduces response times in game as well as in cognitive tasks in a web-based experiment, the effect could be inappropriately attributed to psychomotor or processing speed, whereas it might be entirely governed by lower latency hardware. In other words, it could be that players with better hardware would get better scores on the web experiments because of system speed and responsiveness, and the same players could fare better at video games for the same reason.

There are a few ways to mitigate these issues, and the key to overcoming them is to minimize them with good design and additional features. I employed some strategies to avoid

a few of the problems discussed here, and present them under the description of the experimental procedure. Since Java and Flash are decreasingly popular in web development and pose several problems on client-side, and since JavaScript has a majority in scripting client-side functionality – a total of 88% of websites use JavaScript client-side, 12,3% use Flash and merely 0,1% use Java (W3Techs, 2014) – I decided to explore the usability of JavaScript as a technology for web-based experiments in psychology.

Method

Participants

The participants were a total of 278 League of Legends-players. Out of the entire pool, 218 players were considered for this study, as the rest were under level 30 or reportedly did not engage in ranked games. The final pool of participants ranged between ages 45 and 15 ($M = 21.42$, $SD = 4.15$). The players were mostly from the North America ($N=113$) and Western EU ($N=74$), with a minority of players from EU-North and East ($N=17$), Latin America ($N=5$), New Zealand ($N=7$) and Turkey ($N=2$). 206 of the participants engaged in ranked games in the solo bracket, 51 played 5-versus-5 games, and 18 3-versus-3 games. Out of the team players, 47 also played in the solo bracket.

Instruments

All of the instruments were written in JavaScript (including jQuery and AJAX) on top of the Node.js runtime environment, with Express 4.0 web-framework and Sequelize ORM between the application and a MySQL database. Modified plugins of the jsPsych-library (de Leeuw, 2014) were used for some of the experiments. Source code for all of the experiments, the plugins, the back end structure and descriptions of the dependencies for the experiment can be found in Pöllänen (2014). The experiment was hosted on a Digital Oceans virtual server in San Francisco.

Inventories

The Group Environment Questionnaire. The Group Environment Questionnaire (GEQ – Carron, Widmeyer, & Brawley, 1985) was chosen as a measure of group cohesion because it is well established with a long history of use in sports psychology and group research, with recent evidence suggesting adequate multilevel factorial validity (Fletcher & Whitton, 2014). An instrument from sports psychology was preferred over those from organizational psychology, as interviews with players indicated that the items meant for athletic teams were considered more relevant to their gameplay experience. The players' satisfaction with the instrument drove the decision of not developing a new instrument for the purpose, but rather attempting to repurpose GEQ. The items were modified slightly to match the context of virtual games (e.g. using the word "players" instead of "athletes").

GEQ is a general, rather than situation-specific measure of cohesion in sports teams. It consists of the following four sub-scales forming a four-factor model of cohesion (Brawley, Carron, & Widmeyer, 1987):

- Group Integration–Social (GI-S) conceptualizes a team member's assessment of the group's closeness, similarity and bonding as a social unit - for example, “members of our team do not stick together outside of practices and games”.
- Group Integration–Task (GI-T) conceptualizes the member's assessment of the group's closeness, similarity and bonding around the group's task - for example, “our team is united in trying to reach its goals for performance”.
- Individual Attraction to the Group–Social (ATG-S) conceptualizes to the member's notions of the social interactions and personal acceptance within the team, for example “some of my best friends are on this team”.
- Individual Attraction to the Group–Task (ATG-T) conceptualizes a member's feelings about personal involvement related to the group's common goals and productivity - for instance, “I do not like the style of play on this team”.

GEQ dimension reduction. In order to assess the usablity of the Group Environment Questionnaire (Carron, Widmeyer, & Brawley, 1985), I first analysed its structure using a varimax-rotated principal components analysis using the R psych package (Revelle, 2014). The results did not support the four-factor structure in the original inventory, so I searched for optional solutions with different factor analytic procedures. An exploration through parallel analysis, optimal coordination and acceleration factor in R suggested realistically between one and three factors. A comparison between different solutions netted very similar results, the strongest of which was a Promax-rotated two-component solution (Table 1), with a task-related component that subsumed items from ATG-T and GI-T, and a socially oriented component that subsumed ATG-S and GI-S.

Table 1. *Component loadings from a two-component promax principal component analysis on the GEQ items.*

item	PC1 – task	PC2 – social	communality	unique variance
GEQ2			0.22	0.78
GEQ4	0.72		0.48	0.52
GEQ6	0.55		0.31	0.69
GEQ8	0.85		0.65	0.35
GEQ1		0.45	0.2	0.8
GEQ3		0.57	0.34	0.66
GEQ5		-0.84	0.61	0.39
GEQ7		0.47	0.3	0.7
GEQ9		-0.6	0.32	0.68
GEQ11	0.6		0.34	0.66
GEQ13		0.62	0.39	0.61
GEQ15		-0.75	0.54	0.46
GEQ17		0.79	0.63	0.37
GEQ10	-0.72		0.45	0.55
GEQ12	-0.63		0.41	0.59
GEQ14	0.67		0.47	0.53
GEQ16	-0.6		0.51	0.49
GEQ18	0.58		0.35	0.65

Note. Negative loadings are from reverse-scored items.

The correlation between the components was moderate ($r = 0.36, p < 0.05$), and they explained 22% and 19% of the variance, respectively. Items 2 and 11 behaved differently from the original inventory – 2 did not have a loading above 0.3 on either component, and 11 loaded on the task-related component, despite of originally being an item on the social scale. Item number two sounds "I'm not happy with the play time I get", and it could simply be that it lacks validity for the gaming context, where players do not get benched the same way as they do in traditional sports. Item number 11 is the following: "Members of our team would rather go out on their own than get together as a team." In the original inventory, this refers to socialization with the team. In an online games context, this item is easier to interpret so that the individual team members rather play solo games than queue together with the team, making it a task-related item.

The two components also had sufficient internal consistency, with Cronbach's α for the first, task-related component at 0.83 and for the second, social component at 0.79. An inspection of the distribution of the scores in the player population, however, indicates that at least in the current sample, the inventory does not discriminate between players sufficiently, as they were likely to estimate their teams in a very positive light.

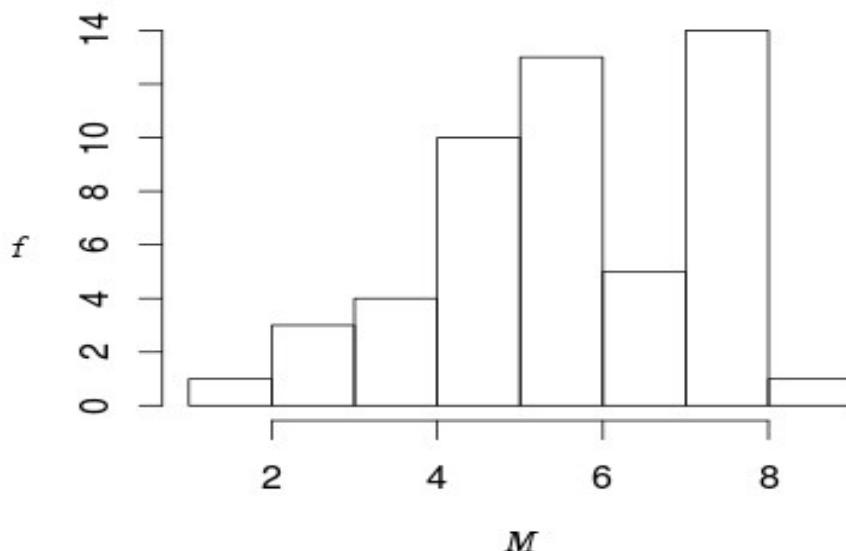


Figure 2. Average scores on the task-oriented GEq component from all players who engage in team games.

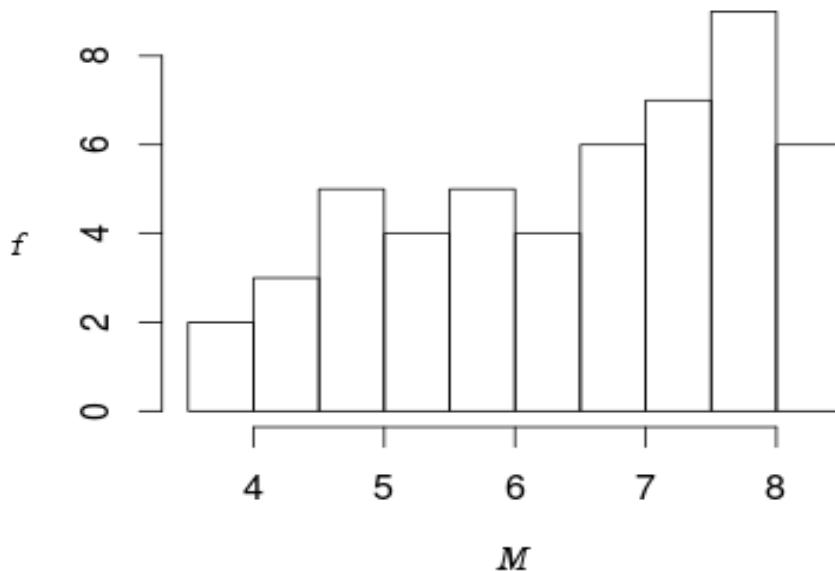


Figure 3. Average scores on the socially oriented GEQ component from all players who engage in team games.

The NASA Task Load Inventory. Nasa TLX (Hart & Staveland, 1988) was selected as a well-established, short task load inventory with good metric characteristics (Hart & Staveland, 1988; Hart, 2006). The instrument originally consisted of two parts. In the first part, six sub-scales are presented on a single page, with the following description of each of the scale:

- Mental Demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
- Physical Demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?
- Temporal Demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
- Performance: How successful were you in performing the task? How satisfied were you with your performance?
- Frustration: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

- Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

The items are rated on a 100-points range with 5-points increments. In the original version, the second part of the inventory creates individual weighing by importance for each of the six subscales by prompting the subjects to compare the categories pairwise based on their perceived importance for the task load. The estimated task loads are then weighed according to their importance. The version used for this study, however, consists of only the first part of the original test, without the pairwise comparisons. This modification was done to simplify the study design, with evidence indicating that this procedure (referred to as the Raw TLX) is as sensitive as the original instrument (Hart, 2006).

During the test, the participants were given the instruction to think about their past games in their most frequently played role and position, and to drag the slider on the scale that best indicates their typical experience of the task load during a game.

Cognitive tasks. A major difficulty in selecting instruments for the cognitive tasks was a lack of prior literature on the topic of differential performance in expert video-game players. I used previous literature on video game training effects and literature comparing expert gamers with novices to guide my decision-making, and decided to program four cognitive tasks: the Eriksen flanker task, a mental rotation task, a spatial span task, and the Tower of London.

The Eriksen Flanker task. Two conflict-tasks have previously been used in game studies: varying versions of the Eriksen flanker task (Eriksen & Eriksen, 1974) and the Stroop-task. The Stroop task has been used before by e.g. Maillot, Perrot and Hartley (2012); and Anderson, Bailey and West (2012). Maillot et al. (2012) found that older adults in their study improved significantly in their performance on resisting the interference following video game play. Anderson et al. (2010) compared players with different levels of experience and discovered no difference between less experienced and more experienced gamers in the Stroop interference effect, but did find a negative association between video game experience and proactive cognitive control.

Since my participant pool is international, I decided for the flanker task, as the fact that the participants are not all native speakers of English could influence the amount of interference in the Stroop task. A variation of the flanker task -- Attentional Network Test, or

ANT -- was used by Dye, Green and Bavelier (2009), who discovered that action video game players made faster correct responses to incongruent trials. Best (2012) used a children's version of the ANT to discover that exergaming enhances children's speed in resolving interference from conflicting visuospatial stimuli.

The Flanker task used for this experiment was programmed roughly in a similar manner to the Flanker task in the PEBL library (Mueller & Piper, 2014), which was based on work of Stins, Polderman, Boomsma and deGeus (2007). During the task, the participant is presented with images of five arrows in a row. Their task is to press the arrow key corresponding to the direction of the middle arrow, which points either to the same direction as the arrows flanking it (a congruent trial) or to the opposite direction (incongruent trial). The participants went through a total of forty-eight trials - twelve trials for congruent and incongruent stimuli of each direction – in a randomized order.

The measures gained on the Flanker task are the speed and accuracy in response to congruent vs. incongruent trials. The participants were also scored on the difference in response time between correct congruent and incongruent trials

Mental Rotation. Different mental rotation tasks have previously been used in game research, for different purposes. Boot et al. (2008) discovered that trained players were no faster but were more accurate than non-trained players in a mental rotation task, and that playing Tetris led to the greatest improvement. Okagaki and Frensch (1994) also used Tetris, but discovered an improvement in mental rotation time, rather than accuracy. In their study on older adults, Maillot et al. (2012) discovered no improvement of game training on their visuospatial measures (including the mental rotation task). Guha, Jereen, DeGutis and Wilmer (2014) also did not find an training effect for performance in a mental rotation task from action video game play, whereas Feng et al. (2007) reported that video game play did create a training effect and reduced gender differences in a mental rotation task.

The mental rotation task used in this study had three types of two-dimensional stimuli. Stimulus A consisted of one long line, perpendicularly intersected by two short ones. Stimulus B and C were similar to the stimulus A, except that they had one and two diagonal lines added to them, respectively. During the task, the participants were presented with an unrotated version of the target stimulus, and a version of the stimulus that was either an identical or mirrored version of the target, and rotated 0, 60, 120, 180, 240 or 300 degrees. The two

images were displayed side-by-side, and the target stimulus was randomly displayed on either left or right side. The participant was instructed to assess as quickly and accurately as possible, whether the rotated image was identical or a mirror image of the target, and to press Q if they were identical and P if mirrored.

The participants were scored on speed and accuracy for unrotated and rotated stimuli, on the difference between their response times to rotated and unrotated stimuli, and on the relative difference between their response times to rotated and unrotated stimuli (rotated response time - unrotated response time)/unrotated response time.

Spatial Span. Visuo-spatial working memory tasks have been used in game studies before, e.g. by Maillot, Perrot and Hartley (2012); Boot, Framer, Simons, Fabiani and Gratton (2008); and Okagaki and Frensch (1994). Okagaki and Frensch (1994) discovered that playing tetris improves reaction times in mental rotation and spatial visualization tasks. In their study on older adults, Maillot et al. (2012) found no improvement of video-game training on a spatial span task. In their visual short-term memory test, Boot et al. (2008) discovered that expert players far outperformed non-gamers in a spatial span measure, but there was no difference between longitudinal groups, so the training effect was not clear. On their working memory operation span test, as well as on the Corsi block-tapping test, Boot et al. (2008) found no difference between experts and novices, and on their spatial 2-back task, experts were faster but no more accurate, with longitudinal groups improving with practice. The jury is out on the characteristics of training effects on spatial span tasks, but as training effects are not the subject of this study, the spatial span task was included into the battery, as Boot et al.'s (2008) findings imply some importance of visuo-spatial working memory functions for gameplay.

The spatial span task was programmed similar to that of Owen and Hampshire (n.d.) as a variation of the Corsi block tapping task (Corsi, 1972). The stimuli were displayed as rectangles in a four-times-four matrix of elements, where a number of the rectangles corresponding to the trial difficulty would be highlighted in a sequence. After the stimulus was generated, the participant had to click on the boxes in the same order in which they had been highlighted. The first stimulus had a length of four blocks, trial difficulty was lowered after two failures on the same stimulus length and increased after successful completion, and the task was concluded after four failures. Number of trials completed and the maximum correct item difficulty as a visuo-spatial digit span was logged for each participant.

The Tower of London. Shalice (1982) developed the Tower of London (TOL) task as a measure of planning and problem solving skills in patients with frontal lobe damage. During the Tower of London task, the participant is presented with a set of three pegs of different length, with three colored balls on them. They are then presented a model of a different layout using a similar set of balls and pegs, and they have to reposition the balls in order to get from the first layout to the second layout with as few moves as possible. In Shalice's study, problem solving ability was determined by the average number of excess moves made to the problem - the larger the number, the poorer the problem solving ability. Since Shalice's study, however, several versions of the test have been developed for use on both clinical and the healthy population, on a range of different difficulties.

Boot, Framer, Simons, Fabiani and Gratton (2008) used the London Tower test in their video game training study. Unfortunately, they did no comparison between expert and novice gamers, but only looked at their longitudinal training groups, all of which improved the same. I decided to use the Tower of London despite of its low presence in prior literature, because gameplay in League of Legends often involves the strategic planning and execution in a setting where speed and accuracy is important.

In the version of Tower of London in this study, the participant was presented with two parallel images - the first one representing the original layout, and the second one the final layout. They were then asked, how many times they would have to move a ball to get from the original layout to the final layout, performing as few moves as possible. The response was recorded as a button press on the corresponding number on the keyboard. There were three training images and twenty-two trial images, ranging between one move and six moves by difficulty. The participants were scored by the most difficult item they completed, and by the average difference between the participant's response and the actual number of moves - as the average number of error in the participant's estimation.

Dimension reduction on the cognitive tasks. As there was a significant number of predictive variables from the cognitive tasks, I decided to perform dimension reduction with principal component analysis on them, as well. A four-component varimax-rotated solution netted the best results, shown in Table 2.

Table 2. Component loadings, communalities and unique variance in each score on the different cognitive tasks in a four-component principal component analysis with varimax rotation.

	RC1 – flanker speed	RC2 – mental rotation	RC3 – flanker accuracy	RC4 – planning	communalities	unique variance
flanker	congruent RT	0.91			0.86	0.14
	<i>p</i> correct, congruent		0.63		0.52	0.48
	incongruent RT	0.93			0.90	0.11
	<i>p</i> correct, incongruent		0.80		0.70	0.30
	mean RT	0.97			0.95	0.05
	inc. RT – cong.RT	0.45		-0.51	0.55	0.45
	<i>p</i> correct, total		0.90		0.82	0.18
	conflict adaptation RT	0.87			0.80	0.20
	conflict adaptation <i>p</i> correct		0.66		0.52	0.48
	rotated RT	0.90			0.87	0.13
mental rotation	rotated <i>p</i> correct	0.40			0.20	0.80
	unrotated RT	0.36			0.28	0.72
	unrotated <i>p</i> correct	0.45			0.21	0.79
	RT at 120 degrees	0.84			0.74	0.26
	<i>p</i> correct at 180 degrees				0.09	0.91
SS	rotated RT – unrotated RT	0.92			0.85	0.15
	(rotated RT – unrotated RT) / unrotated RT	0.80			0.67	0.33
	spatial digit span	-0.34			0.22	0.79
	<i>p</i> correct		0.87		0.80	0.20
Tower of London	most difficult correct item		0.69		0.49	0.51
	mean incorrect estimate		-0.90		0.83	0.17

Note. Loadings below .3 are suppressed. RT = response time, *p* correct = proportion of correct responses, conflict adaptation – difference between response time in a pair of incongruent and congruent flanker stimuli, where the congruent stimulus follows the incongruent, SS = spatial span task.

On the principal components, the speed aspects of the flanker task performance loaded on the first component, which explained a 19% of the total variance in the cognitive tasks. There was a negative loading from the Spatial Span task, which did not load on any other component, and had a lot of unique variance. I thought the test performed suspiciously poorly, but left its negative loading on the component due to empirical reasons. In a way, it indicates that the first component has to do with processing or allocation of cognitive resources that prioritises speed over precision.

The second component had loadings from the mental rotation items (both speed and accuracy), explained 16,7% of the variance, and can be considered an indicator of efficacy in two-dimensional mental rotation. The third component corresponds to accuracy (proportion of correct responses) on the flanker task, explaining 12,6% of the total variance, and the fourth component has loadings from the Tower of London test and indicates planning, explaining 11,2% of the variance on the cognitive tasks.

All of the participants were scored on the four components using regression through the automated scoring option in the psych package for R (Revelle, 2014). Since there was good reason to believe the missing values were non-random (whether or not a person chose to drop out from the experiment was very likely a function of the current or the previous task), I chose not to perform missing value imputation. Unfortunately, this reduced the power of the analyses, as it significantly reduced the size of the participant pool.

Procedure

The participants were recruited on various online forums, MOBA- and gaming-related online communities and news sites, and the official boards of Riot Gaming. All of them took part in a uniform experimental battery, the completion of which took between twenty and thirty minutes, in the following task order:

1. Informed consent and instructions on participation
2. Demographic items
3. GEQ
4. NASA TLX
5. Eriksen flanker task

6. Mental rotation task
7. Tower of London task
8. Finish page with an opportunity to contact the researcher for feedback or comments

There are a few precautions and design decisions I made to mitigate some of the shortcomings of web-based experiments. First, the informed consent included a section on rules of participation, where I informed participants on what measures I'd like them to take in order to guarantee some basic usability of the data. As a part of the informed consent, the participants were given the following general instructions:

Since this is a web experiment, the results can be impacted by things such as system load and by uncontrollable events in your environment, and there are a few precautions I'd like you to take:

- Please minimize the number of programs running in the background.
- Keep the number of open tabs in your browser window to a minimum.
- Make sure you will not be bothered during the cognitive tasks.
- Try to find a relaxing and distraction-free environment.
- Please complete the experiment in a single sequence. You will not be able to return to your session later!
- Short breaks during the instruction screens are fine.
- If you are on a laptop, you can complete all of the tasks on a track pad, but I recommend using a mouse for your own convenience.

Upon submitting their agreement to the informed consent, a version 4 universally unique identifier (UUID) was generated for the participant and appended to the URL. This user ID was always submitted together with data from the each task in experiment. If the participant attempted to submit data twice with the same UUID – if a full entry with that ID parameter already existed in the relevant table in the database – a duplicate ID error was thrown, and the participant was redirected to the front page with an appropriate error message. Similarly, there was a redirect to the landing page with an error message for missing ID if a page was accessed with no UUID in the request.

In the demographics form, I mitigated submission of invalid data by validation with Sequelize – the ORM I used between the application and the MySQL database – and by

utilizing appearing and disappearing form elements depending on the level and play style of the participant. For instance, if the participant reported being below level 30, they had no option to select that they played ranked games. The sub-menus prompting for division and tier only appeared if the participant selected that they did indeed play that particular mode of ranked games. Additionally, if relevant data was missing, the form could not be submitted and relevant messages about missing data were appended to their respective elements.

Access from mobile devices was blocked to the experiment by redirecting requests from mobile devices to a web page with a short message explaining that participation was only possible on a computer. I planned all of the experiments so that they could be completed with just a keyboard and a track pad in case there were participants who used a laptop without a mouse.

Most of the cleanup happened after the data was collected. I logged each trial on every experiment separately, and first excluded trials where response times were below 300 milliseconds. I then calculated these trials for every participant, and if there were many such trials throughout the task, I would exclude the results for that task for that participant. I also excluded tasks for participants who had an unreasonably long response time on some trial, since that implied that they had taken a break during the task. I excluded results where correctness of responses was inappropriately low (at 0.6 or below accuracy), and I excluded duplicate cases, where a participant from the same region with the same summoner name and other demographics had two full entries under two different UUIDs. This meant that they actually went through the effort of completing the entire battery twice – quite the dedication!

Once the data had been cleaned up and aggregated, I assessed the structure of the GEq and performed dimension reduction with principal component analysis on the cognitive tasks. I computed Spearman's correlation coefficients for the performance measures, the survey results and scores on the cognitive tasks, explored regression solutions using stepwise analysis and constructed a couple of preliminary regression models to dissect the data.

Results

Correlational Analysis

At first, I constructed correlation matrices that can be used to inspect the relationships between the ratings, GEQ and TLX items and cognitive tasks.

Table 3. *Spearman correlation coefficients between solo rating, TLX scores and the cognitive components for all players engaging in solo games.*

	Solo rating	mental demand	physical demand	temporal demand	performance	effort	frustration	task load	RC1	RC2	RC3
Solo rating											
mental demand	0.00	-0.10									
physical demand	0.05	0.19	0.42***								
temporal demand	-0.14	-0.02	0.30***	0.21*							
performance	0.16	0.16	0.01	0.06	0.01						
effort	0.04	-0.17	0.47***	0.29***	0.37***	0.20*					
frustration	-0.18*	-0.24	0.17*	0.07	0.03	-0.13	0.28***				
task_load	-0.08	-0.06	0.70***	0.61***	0.55***	0.24**	0.72***	0.45***			
RC1/flanker speed*	-0.10	-0.46*	-0.04	0.15	-0.11	-0.18	-0.23	0.01	-0.11		
RC2/mental rotation	-0.01	-0.16	0.05	-0.02	-0.09	0.07	-0.04	-0.22	-0.14	0.03	
RC3/flanker accuracy	0.14	0.10	-0.12	0.05	-0.19	-0.13	-0.10	-0.05	-0.14	0.12	-0.27*
RC4/planning	0.14	0.16	-0.11	0.08	0.24	0.04	0.09	-0.17	-0.02	-0.04	-0.05

Note. The flanker speed component is reversed, in that it a high score implies high response times. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

The only significant correlation is between solo rating and frustration. This is likely due to the reduced statistical power due to sample size loss after not performing missing value imputation in coming up with the principal component solution and scoring the cognitive tasks.

Table 4. Spearman correlation coefficients between 5v5, 3v3 and solo rating, TLX scores, GEQ scores and the cognitive components for all players engaging in in team games

	5v5 rating	solo rating	Highest team rating	mental demand	physical demand	temporal demand	performance	effort	frustration	task load	GEQ task	GEQ social	RC1	RC2	RC3
5v5 rating															
solo rating	0.65***														
team rating	0.93***	0.64***													
mental demand	-0.26	0.00	-0.07												
physical demand	0.20	0.05	0.22	0.39***											
temporal demand	0.01	-0.14	-0.08	0.29***	0.22**										
performance	0.08	0.16	0.12	0.01	0.06	0.00									
effort	-0.33*	0.04	-0.08	0.45***	0.24**	0.34***	0.22**								
frustration	-0.34*	-0.18*	-0.23	0.15	0.06	0.04	-0.12	0.28***							
task load	-0.24	-0.08	-0.09	0.68***	0.59***	0.56***	0.25**	0.70***	0.46***						
GEQ task	0.11	0.06	-0.05	-0.11	-0.09	-0.10	0.23**	0.07	-0.20*	-0.08					
GEQ social	-0.13	-0.18*	-0.14	-0.05	-0.02	-0.15	-0.01	0.02	0.12	0.02	0.28***				
RC1/ flanker speed*	-0.29	-0.10	-0.36	-0.02	0.13	-0.09	-0.21	-0.18	0.01	-0.10	-0.10	-0.11			
RC2/ mental rotation	0.20	-0.01	0.03	-0.01	0.02	-0.09	0.05	-0.04	-0.19	-0.14	-0.02	-0.04	0.08		
RC3/ flanker accuracy	0.23	0.14	0.02	-0.07	0.02	-0.18	-0.13	-0.10	-0.05	-0.14	0.05	0.09	0.20	-0.24*	
RC4/ planning	-0.13	0.14	-0.03	-0.08	0.06	0.22	0.05	0.07	-0.16	-0.03	0.31**	0.16	0.00	-0.10	0.01

Note. The flanker speed component is reversed, in that it a high score implies high response times. 3v3 rating has been excluded due to small sample size. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

There are significant, moderate negative correlations between the effort and frustration on the Task Load index, and performance for people who engage in team games. A nearly significant negative effect exists between the flanker speed component and 5v5 and solo rating, but in the current data set, the components simply do not appear to directly relate to performance – this probably calls for resampling. The task component on GEq appears to relate to self-estimation of performance and lower frustration in ranked games, and it also has a moderate positive correlation with the fourth cognitive component, related to planning in the Tower of London, as well as the social component of team cohesion. It makes sense that team members with better planning abilities would also have a more strategic and task-oriented play-style.

Regression

As there was little guidance from previous literature or theoretical grounds on the order and combination of variables to include into regression, I used stepwise analysis to explore meaningful combinations of predictors. This was done by the stepAIC function in the MASS R package (Venables & Ripley, 2002).

In order to predict solo rating for all players engaging in solo-based ranked games, a few models were plausible. The strongest model predicted a small amount of variance in solo rating from a combination of the third ($\beta = 0.22, p = 0.08$) and fourth ($\beta = 0.18, p < 0.05$) cognitive components (flanker accuracy and planning) together with frustration ($\beta = -0.24, p = 0.06$) on TLX ($R^2 = 0.16, F(3,57)=3.52, p < 0.05$).

Predicting team ratings, or five-versus-five results was more difficult, as there were several plausible models that seemed to explain 100% variance – for instance, one between task load ($b = 0.99, p < 0.0001$) and the second cognitive component ($b = 0.05, p < 0.0001$) – mental rotation ($R^2 = 1.0, F(2,20)=2.464e+3, p < 0.0001$). Notice that task load explains nearly all of the variance in the results – a similar model can be repeated with the third and the second cognitive component where cognition, as in the first role, contributes to the model by only a fraction, and where the TLX score appears to overpower the entire regression. This is likely due to the small sample size due to missing values on the cognitive tasks, and where on one hand it is disappointing that the data works so poorly for regression, it is alleviating to notice that despite of being such a short inventory, TLX seems to function well in the context.

Discussion

Overall, the results are inconclusive and afford few useful inferences, other than that different self-reported aspects of task load appear to play a significant role in both solo- and team-based player performance in League of Legends, and that cognitive skills also relate to performance, albeit not in an inambiguous or direct manner that would be readily interpretable from the data. The response time component had the highest direct – although non-significant – connection to team players' performance, but the rotation and accuracy components functioned better – although still fairly poorly – in the regression.

In order to get a better grip of the phenomenon under research, a few changes should be made. First of all, it is necessary to repeat the study and attempt to get a sizeable pool of participants that would work as a representative cross-section of the different layers of performance in the player population. A large chunk of the sample was lost due to non-random missing data, and it is important to come up with methods to reduce drop-out as well as an approach in imputation that would sufficiently patch up the holes created by the inevitable proportion of participants who do drop out. A couple of ways to mitigate the drop-out would be to shorten the battery, to incentivise participation, and to take more advantage of adaptive testing.

The spatial span task simply did not appear to work together with the rest of the cognitive tasks, and it did also did not seem to correlate at all with neither solo ($r = 0.03, p > 0.05$) nor 5v5 rating ($r = 0.14, p > 0.05$). Generally, I believe a better-guided selection of cognitive tasks is definitely in order. Seeing that the mental rotation task did work to some degree in the regressions, I'd modify it by utilizing adaptive testing and three-dimensional stimuli. For the Tower of London test, I'd also utilize adaptive testing, together with object manipulation (moving the balls on the pegs with a mouse) instead of static images – the way the task is constructed in the PEBL test library (Mueller & Piper, 2014) – and a solution like this should be relatively easy to construct in JavaScript with jQuery elements. Instead of the Spatial Span task, I'd look into how task switching might relate to performance, and I'd integrate a go/no go component to the Flanker task.

It was positive to note that Nasa TLX worked so well for the purpose of this study. It is a very short inventory and only adds a single short page to the battery, and so it makes a significant contribution to the experiment while adding very little time and complexity.

Another use for TLX could be to explore task load during the cognitive tasks themselves, instead of during gameplay.

The two-component structure and sufficient reliability of GEQ was also a positive discovery, however more research needs to be undertaken to find out whether the lack of discrimination might be an issue in a video game context. It would also be reasonable to look into how communication style or different conative factors contribute to performance, instead of team cohesion, as this is something that could relate to task load through event appraisal and would also be relevant to solo players, who do not have the opportunity to develop team cohesion in the numerous randomized groups they take part in.

Finally, my intention was also to explore the usability of web-based methods for game studies. It appears that web-based methods are indeed suitable for this kind of research, as they allow sampling of a population that would simply be out of reach for other type of assessment. The participants left numerous messages on the finish page after completing the tasks, most of them expressing enthusiasm about the experiment and asking for more information about the tasks and the expected results. On one hand, informing participants is necessary, and engaged and enthusiastic participants are a positive factor. On the other hand, while researching performance, inadequate blinding is one of the major problems (Andrews & Murphy, 2006), and volunteering too much information about the objective can damage the research outcome.

Carrying out web-based research is certainly challenging. Web-based experiments suffer from all of the shortcomings of traditional survey research – participant drop out, invalid responses, submitting false data, multiple entries, self-selective samples... Additionally, researchers also face with other challenges, such as invalid response times, technical challenges and bugs created by platform incompatibility, unpredictable confounding variables and so on. Another challenge is that in research for cognitive psychology, batteries are longer and more strenuous than surveys, and motivating participants becomes a key issue. Regardless, there are many ways to control some of the problems. Many of the issues leave tangible behavioral artifacts in the data that can be addressed as long as the data is not entirely aggregated before it gets piped into the database – and for a number of research problems, there is simply no alternative for web-based studies. Additionally, because of their natural scope, web-based experiments are great for prototyping and piloting before carrying out a more controlled laboratory task – for which the same instrument can be used, since nothing

prevents one from using the same battery in a laboratory setting. They offer some additional conveniences that cannot be utilized in laboratory-based studies: participants can complete tasks at their own pace, in their own space, without the confounding presence of the experimenter, with familiar tools and at a convenient time, and so on. Web technologies provide the experimenter with an interesting toolbox, with e.g. the option to use APIs or login systems to triangulate data and to extract more reliable demographic information – if the participant agrees to this, of course.

From the experience I have gained with this study, I'd conclude that the key to successful web-based research is the combination of a good dose of creativity, thorough consideration of details, and oversampling. There is no iterative product development when it comes to scientific research, but the path from a pilot or a test run to an active study feels shorter when the entire battery is contained in the virtual space. Here, version control really comes in handy. Elementary peer-review and sharing source code and experimental stimuli is simple through reading, cloning and forking repositories, and performing another round of sampling with a modified experiment is as simple as modifying the code and pushing the changes. I believe that web-based experimentation not only widens the scope and diversity of samples and research questions, but significantly contributes to the emergence of creative and open behavioral science.

References

- Anderson, C. A., Bailey, K. & West, R. (2010). A negative association between video game experience and proactive cognitive control. *Psychophysiology*, 47, 34/42.
- Andrews G., Murphy K. (2006). "Does video-game playing improve executive function?" in *Frontiers in Cognitive Sciences*, ed. Vanchevsky M. A., editor. (New York, NY: Nova Science Publishers, Inc.), 145–161. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3171788/>
- Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*, 317, 472-476.
- Balicer, R. (2005). Modeling Infectious Diseases Dissemination Through Online Role-Playing Games. *Epidemiology* 18 (2) pp. 260–261
- Best, J. R. (2012). Exergaming immediately enhances children's executive function. *Developmental Psychology*, 48(5), 1501-1510.
- Blizzard Entertainment. (2010, October 7). World of Warcraft® subscriber base reaches 12 million worldwide. Retrieved from: <http://eu.blizzard.com/en-gb/company/press/pressreleases.html?101007>
- Boot, W. R., Framer, A. F., Simons, D. J., Fabiani, M. & Gratton, G. (2008). The effects of video game playing on attention, memory and executive control. *Acta Psychologica*, 129, 387-398.
- Carron, A. V., Widmeyer, W. N., & Brawley, L. R. (1985). The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of Sport Psychology*, 7, 244-266.
- Colzato L. S., van Leeuwen P. J. A., van den Wildenberg W. P. M., Hommel B. (2010). DOOM'd to switch: superior cognitive flexibility in players of first person shooter games. *Frontiers of psychology*, 1(8).

Cooke, N.J., Gorman, J.C. & Winner, J.L. (2006). Team cognition. In F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, and T. Perfect, *Handbook of Applied Cognition*, 2nd Edition Wiley.

Corsi, P.M.(1972), Human memory and the medial temporal region of the brain, Unpublished doctoral dissertation. McGill University.

de Leeuw, J. R. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*. Advance online publication.
doi:10.3758/s13428-014-0458-y

Drachen, A., Yancey, M., Maguire, J., Chu, D., Yuhui Wang, I., Mahlmann, T., Schubert, M. & Klabajan, D. (2014) Skill-Based Differences in Spatio-Temporal Team Behaviour in Defence of The Ancients 2 (DotA 2) (In proceeding) Proceedings of the IEEE Games, Entertainment, and Media (GEM) Conference 2014.

Ducheneaut, N., Dmitri, W., Nickell, E., Xiog, L., Yee, N. & Zhang, Y. (2006). From tree house to barracks: The social life of guilds in World of Warcraft. *Games and culture* 1(4), 338-361

Dye, M.W.G., Green, C.S., & Bavelier, D. (2009). The development of attention skills in action video game players. *Neuropsychologia*, 47(8-9), 1780-1789.
doi:10.1016/j.neuropsychologia.2009.02.002.

Eichstaedt, J. (2001). An inaccurate-timing filter for reaction-time measurement by JAVA-applets imple-menting internet-based experiments. *Behavior Research Methods, Instruments, and Computers*, 33(1), 179–186.

El-Shagi, M. & von Schweinitz, G. (2014). *The Diablo 3 Economy: An Agent Based Approach* [PDF document]. Retrieved from:
<http://econ.henu.edu.cn/upload/20141016/14134680789594.pdf>

Entertainment Software Association (ESA) (2011). *The 2011 Essential Facts About the Computer and Video Game Industry*. Retrieved 20.12.2011 from
http://www.theesa.com/facts/pdfs/ESA_EF_2011.pdf.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.

Feng, J., Spence, I. & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological science*, 18(10), 850-855.

Ferrari, S. (2013) From Generative to Conventional Play: MOBA and League of Legends, Proceedings of DiGRA 2013: DeFragging Game Studies, Volume 1 (1), 1-17.

Fletcher, R. B. & Whitton, S. M. (2014). The Group Environment Questionnaire: A Multilevel Confirmatory Factor Analysis. *Small Group Research*, 45(1), 68-88. doi: 10.1177/1046496413511121

Green C. S., Bavelier D. (2007). Action video game experience alters the spatial resolution of attention. *Psychological Science*, 18(1), 88-94.

Green, C. S., & Bavelier, D. (2003). Action video-game modifies visual selective attention. *Nature*, 423, 534-537.

Guha, A., Jereen, A., DeGutis, J. & Wilmer, J. (2014). No action video game training effects for multiple object tracking or mental rotation. *Journal of Vision*, 22(14).

Hart, S. (2006). Nasa-Task Load Index (Nasa-TLX); 20 Years Later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 50, 904-908.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Human Mental Workload* (P. A. Hancock and N. Meshkati (Eds.)), 139-183. North-Holland: Elsevier Science.

Heslin, R. (1964). Predicting group task effectiveness from member characteristics. *Psychological Bulletin*, 62, 248-256.

Hollenbeck, J.R., Moon, H., Ellis, A.P., West, B.J., Ilgen, D.R., Sheppard, L., Porter, C.O., & Wagner J.A. (2002). Structural contingency theory and individual differences: Examination of external and internal person-team fit. *Journal of Applied Psychology*, 87, 599-606.

Karle J. W., Watter S., Shedd J. M. (2010). Task switching in video game players: benefits of selective attention but not resistance to proactive interference. *Acta Psychologica (Amst.)* 134(1).

LePine, J.A., Hollenbeck, J.R., Ilgen, D. R., & Hedlund, J. (1997). Effects of individual differences on the performance of hierarchical decision-making teams: Much more than g. *Journal of Applied Psychology*, 82, 803-811.

Maillot, P., Perrot, A. & Hartley, A. (2012). Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. *Psychology and Aging*, 27(3), 589-600.

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of neuroscience methods* (222), 250–259.

N. Pobiedina, J. Neidhardt, M. d. C. Calatrava Moreno & H. Werthner (2013). Ranking factors of team success in Proc. of the 22nd international conference on World Wide Web companion. International World Wide Web conferences Steering Committee, 2013, pp. 1185-1194.

Narishm. (2013). A map of the Summoner's Rift map on League of Legends with an overlay dividing it into easy to remember regions. Retrieved 29.11.2014 from <http://narishm.deviantart.com/art/League-of-Legends-Summoner-s-Rift-Regions-Map-413345149>

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology*, 15(1), 33-58.

Owen, A. M. & Hampshire, A. (n.d.). Spatial span ladder [software for online web-based testing]. Cambridge Brain Sciences Inc, University of Western Ontario, Canada. Retrieved from <http://www.cambridgebrainsciences.com/browse/memory/test/spatial-span-ladder>.

Perlroth, N. (2013). Department of Homeland Security: Disable Java ‘Unless It Is Absolutely Necessary’. Bits – the business of technology, a New York Times blog. Retrieved from: <http://bits.blogs.nytimes.com/2013/01/14/department-of-homeland-security-disable-java-unless-it-is-absolutely-necessary/>

Pöllänen, T. I. (2014). Test battery of JavaScript cognitive tasks written using jsPsych for my master's thesis. Github repository. Retrieved from <https://github.com/tuuleh/masters-battery>.

Revelle, W. (2014) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych> Version = 1.4.8.

Riot Games. (2014). “Our Games”. Retrieved from: <http://www.riotgames.com/our-games>

Salas, E., Dickinson, T. L., Converse, S.A. & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R.W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-20). Norwood, NJ: Ablex.

Shalice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 298, 199-209.

Stins, J. F., Polderman, T. J. C., Boomsma, D. I., & de Geus, E. J. C. (2007). Conditional accuracy in response interference tasks: Evidence from the Eriksen flanker task and the spatial conflict task. *Advances in Cognitive Psychology* 3 (3), 389–396.

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

W3Techs (2014). Usage of client-side programming languages for websites. Retrieved 2.12.2014 from http://w3techs.com/technologies/overview/client_side_language/all