Engagement Factors in Games with a Purpose

J.A.L. Farias, F.V. Brasileiro

Federal University of Campina Grande, Dept. of Computer Science, Brazil

Abstract

Games with a Purpose (GWAPs) are applications that use games to harness human intelligence to perform various computational tasks. That is, users perform computations as a byproduct of being entertained while playing a game. Such games are increasing in popularity and it is critical to examine how to maximize their performance, understanding why people play this type of game. In the present research, we aim at identifying the influence of some factors in the engagement of users when playing GWAPs. For that, we have created a collaborative GWAP to perform tagging in audio files and have run an experiment using this game. Results suggest that the player profile is the most relevant factor in defining the level of user engagement in GWAPs, while factors such as gender and age are not very relevant. Further, using statistical tests, we found that explorer users are the ones that are most engaged in these applications. Our findings should provide useful suggestions for designing GWAPs in the future.

Keywords: Human computation games; games with a purpose, tagging; playability; quality; motivations; needs, crowd computing, game design.

Authors' contact:
José A. L. Farias:
joseantonio@copin.ufcg.edu.br
Francisco V. Brasileiro:
fubica@dsc.ufcg.edu.br

1. Introduction

In recent years, there has been a remarkable increase in the popularity of social computing. At the same time, innovative social computing applications have appeared, giving rise to new approaches that empower users to collaborate, communicate and contribute content. One such example is the so-called Game With A Purpose (GWAP) [Ahn and Dabbish 2008], which offers entertainment, but is also meant to accomplish tasks or solve problems. Such games could serve as motivators for users to contribute their brain power to a given endeavor. That is, while users are entertained by playing a game, they are also performing computations as a byproduct of gameplay [Go et al. 2010]. The use of games in novel ways, such as in GWAPs, is not surprising, given that gaming has burgeoned into a significant industry worldwide.

GWAPs have been employed relatively successfully in areas such as image tagging [Ahn and Dabbish 2004], location-based annotation authoring [Lee at al. 2010], and ontology creation [Siorpaes and Hepp 2008]. One of the most successful game is the FoldIt [Cooper et al. 2010], a GWAP designed for human-assisted protein folding, taking advantage of humans' innate spatial reasoning abilities, with the goal of producing accurate protein structure models through gameplay. Improperly folded protein conformations are posted online as puzzles for a fixed amount of time, during which players interactively reshape them in the direction they believe will lead to the highest score. To play, players manipulate digital 3D protein structures, trying to find a configuration with the most tightly packed (and lowest energy) shape for each protein.

For instance, the ESP Game [Ahn and Dabbish 2004] is one of the earliest examples of GWAPs. Two unrelated players are tasked to create matching keywords to randomly presented images within a given time limit. Points are earned based on specificity of the keywords, and coupled with a countdown timer, these elements add excitement and hence motivation for players. Thus, while players have fun with the game, which in turn potentially translates into sustained usage, the matching keywords can be used as tags for the images that can improve the performance of image retrieval algorithms. In the context of GWAPs, the resulting computations in the ESP Game are the keywords that describe images.

Today there are two major problems in designing GWAPs: one is associated with the **quality** of the information that players can produce, i.e. the effectiveness and relevance of the work done by humans to complete the task. The other problem is related to the **amount** of information generated by the game itself and is directly related with the number of players that play the game, and the time these players spend playing the game. In this paper we focus on the second problem.

The amount of information generated by the game is directly related the amount of people playing the game. However, the number of people engaged in the game, is associated with the capacity the game has to attract and retain the players. Them, is critical to the success of GWAPs an understanding of their appeal to different users, and why users are motivated to engage in them. Studies on collaboration and communication technologies have suggested that different features afford different resources to users, which in turn may

affect their perception and subsequent usage patterns [Lee 2010]. Because GWAPs have different features, they potentially offer different benefits to players that could influence sustained usage.

People play games for a variety of reasons and much research has been conducted on their motivations for doing so. [Bartle 1996] proposed four types of players in multiplayer computer games: Killers, Achievers, Socializers and Explorers. These were derived from interviews with players. Subsequently, Bartle organized these player types into four quadrants defined by two dimensions of behavior: acting on versus interacting with the game elements, and focusing on other players versus the virtual world in one's actions. By understanding player types and their motivations, Bartle argued that game developers should provide features to appeal to all player types in order to be commercially successful.

Being an emerging area that yields promise in terms of widespread usage and utility, there is active research in GWAPs. To date, typical areas have focused on the design and implementation of GWAPs [Ho et al. 2009], and their evaluation in terms of usability and performance [Ahn and Dabbish 2004]. However, there are few studies that attempt to identify what kind of people are attracted and actually engage in GWAPs. The literature is unequivocal in stating that games that are perceived positively by players and those that satisfy their motivational needs will more likely attract continued usage [Ha et al. 2007]. But compared to traditional games, it is still not clear why there are so few people playing GWAPs.

The goal of the present research is to know how factors such as sex, age and "type" of user impact the user's engagement in a GWAP. The type of player has been defined through a form that captures a user's perception about yourself and maps into a typology that shall be mentioned later. Our approach is based on gathering empirical evidences through the execution of an experiment. We have deployed an experimental GWAP into a large social network (Facebook) to determine: (1) who are the people that are attracted to the game and who are the people that actually have been engaged; and (2) what are the factors that lead a person to decide his/her engagement to the game.

We argue that studying these issues in the context of GWAPs is necessary for the following reasons. Firstly, the growing popularity of GWAPs necessitates a deeper understanding of how users perceive them so that, ultimately, games can be designed to better cater to different profiles of users. Secondly, GWAPs are different from games for pure entertainment. It is critical to understand what defines their players, so that better games can be designed, maximizing the number of people playing it, and therefore, maximizing the results offered by the game.

This paper is organized as follows: we review the related literature in Section 2 and describe the experiment in Section 3; results are shown in Section 4; validation considerations are discussed in Section 5; we conclude the paper with our final remarks and directions for future work.

2. Related Work

Despite the high potential of GWAPs and the numerous features of suggested techniques, only a few studies have been conducted to explore the important factors that influence the engagement of GWAPs. Here, [Ahn and Dabbish 2008] approach the problem of designing a GWAP, and the difficulty of designing a game that is at the same time efficient in solving the problem that gives the game a purpose other than amusing its players, and attractive to humans [Ahn and Dabbish 2008]. The authors show some successful cases and suggest some design principles which would be more appropriate for certain types of computational problems, but always in terms of quality of the result. Our research complements this work by investigating the quantity aspect of a GWAP efficiency.

[Goh and Lee 2011] evaluated the quality of the results obtained from the same GWAP created in two different ways, a competitive and collaborative model, and tried to establish which approach is best suited for games whose work involve information tagging. According to the authors, the type of the game has no impact on results. The authors confirm the need to know the player's profile, but again, they are concerned with the impact of player profiles only on the quality of the results. In another study, [Go et al. 2010] compared the performance of different GWAPs, but users' motivations for engagement were not investigated.

Many researchers have tried to clarify why people play video games. [Hunicke et al. 2004] has proposed an MDA (mechanics, dynamics and aesthetics) model for game design analysis that includes a list of eight kinds of fun: sensation, fantasy, narrative, challenge, fellowship, discovery, expression, and submission. [Lazzaro 2004] has listed four keys to creating emotion in video games as hard fun, easy fun, altered state, and a people factor. Bartle's [Bartle 1996] four player categories, based on multi- user dungeon (MUD) games, are achievers, killers, socializers, and explorers—a taxonomy that corresponds to player activities. Based on player responses to massively multiplayer online role playing games (MMORPGs), [Yee 2007] has extended Bartle's taxonomy to propose three major MMORPG gaming components: achievement, immersion, and social interaction. According to [Ryan et al. 2006], the pull of a game is sometimes associated with out-of-game effects. Using self-determination theory (SDT), they posit that the pull of games largely results from their ability to generate (at least in the short term) three key feelings

of well-being: autonomy (sense of willingness), competence (challenge and feeling of effectance), and relatedness (feeling of connection with other people). [Koster 2005] views game fun in terms of four categories: fun, aesthetic appreciation, visceral reactions, and social status maneuvers. In that taxonomy, fun focuses on mastering a problem mentally—that is, recognizing new patterns based on our brain's desire for stimuli. Thus, Koster's definition of a good game is one that teaches a player all aspects of the game before the player stops playing. In the following sections, we analyze how reward systems provide pleasure and satisfying experiences by classifying rewards and playing activities, and relate reward mechanics to psychological theories. As yet, there are no studies that have focused on GWAPs

3. Methodology

To accomplish our case study, we first create a GWAP that was used to attract different players' types. We use the traditional concepts found in other GWAPS [Hofer and Eport 2011] and created a game similar to TagATune [Hofer and Eport 2011], an online game designed to collect tags for music and sound clips, collaboratively. Next, we set up the game database with 28,715 audio clips from the FreeSound Database (`http://freesound.org`). Broadly speaking, the genres of music include classical, new age, electronica, rock, pop, world music, jazz, blues, heavy metal, and punk. All audio clips are provided under the Creative Commons License, allowing for much less restrictive usage than other typical music licenses.

There were two reasons for implementing custom games rather than employing existing ones. The first was to achieve better control over the games' look-and-feel to ensure a more consistent user and "friendly" experience during the case study. With existing applications such as the TagATune, customizations to the user interface could not be performed. Second, using our game allowed us unfettered access to the data generated (i.e. user profiles) for analyses, something not possible for existing games.

Then, we published our game as a regular game on Facebook and start to disseminate it among our friends and online game communities, which have spread virally, attracting more and more people and more different players' types. It is common sense that more people play the "traditional" games like *Halo* that play GWAPS. We have chosen to disseminate among communities of gamers because it is important to know who is that people and why they don't play GWAPs, in order to have a better comparison among the factors that make this difference.

The entire disclosure of the game was made highlighting the factors of entertainment and the importance of engagement for *"help Internet search engines find the audio files you want"*. **Figure 1** shows the page that invited people to play the game.

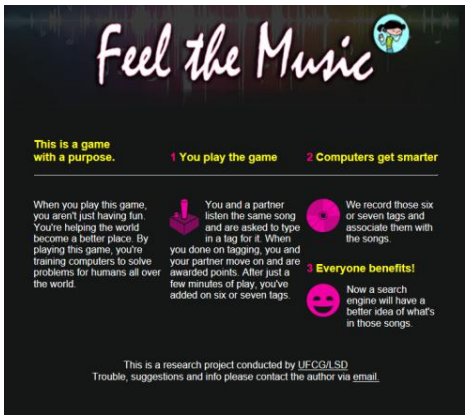

Figure 1 – *About the Game*

For each player who has accessed the game was stored their age and sex, provided by its own Facebook profile, and also the players' type, according to the original Bartle's typology. This model, which was based on observing and analyzing the behaviors people playing together in a multiplayer game, holds that there are four different kinds of play style interests, each of which is given a descriptive name: *Killers*, *Achievers*, *Explorers*, and *Socializers*.

- **Killers:** interfere with the functioning of the game world or the play experience of other players
- **Achievers:** accumulate status tokens by beating the rules-based challenges of the game world
- **Explorers:** discover the systems governing the operation of the game world
- **Socializers:** form relationships with other players by telling stories within the game world

This typology is very popular among game designers, because are useful when trying to categorize and label your players and to understand how a new design may or may not afford particular player experiences. That is, typologies are good for describing user meaning in a game.

The user type is collected through the form shown in **Figure 2**. This form is exhibited for each new player to start playing the game and was designed to capture the user type according to the Bartle's typology. Another type was added to identify people who do not consider themselves gamers, because in a social network we can find people who have never played any videogame. This new type is important because it was not defined by Bartle and could be relevant in definition of the profile of the player who is attracted to or engaged in the game.

We also stored the number of times the player plays the game, i.e., the amount of songs effectively tagged by the user. This amount of work was the factor used to categorize a player as **engaged** or **transient**. A player is **engaged** when he successfully marks more

than a song. On the other hand, a player is **transient** when it begins to play the game but never perform any task, that is, he turned down the game completely. There are other ways to evaluate the engagement, the quality of the markings made by a user, for instance, but our study only focuses on evaluating the amount of work done by the players, so the amount of tasks performed was the main factor for the whole research.

Figure 2 – *The form to collect player type*

The game collected users 'data for one month and during that time was made another game release, but not limited to Facebook only - The game was also featured on Twitter and Orkut after the first two weeks of data capture. This new release was important to attract new and different people not affected by the first release on Facebook, thereby increasing the randomness of the collected sample.

We also categorized the user age information like in **table 1** and we also group the users by your engagement status. If a user has more than one audio clip tagged, we assume this user is engaged. If not, this user is a transient user.

Table 1 – *Age categories*

Value	Age Range
1	0-12 years
2	13-17 years
3	18-29 years
4	30-39 years
5	40-60 years
6	More than 61 years

3.1 Feel the Music

Feel the Music was the game built in order to attract players and collect their profiles' information, and is an **simulation** of a game using a input-agreement mechanism [Law and Ahn 2009]. A screenshot of the interface for a round of *Feel the Music* is shown in **Figure 3**. This game still available at `http://apps.facebook.com/feelthemusic/` on Facebook.

As we are not interested in the quality of results (one of the strengths of input-agreement mechanism), the game only **simulates** the interaction necessary for the input-agreement mechanism, but in fact, the player does not actually interact with another human, but only with a simulated version of another player, that means any player can play this game even if nobody else is playing at same time, and have a real social gaming experience. The game players do not know that was a simulation and imagined they were playing with human beings, trying to work together to produce better results and earn more points and get the highest place on global leaderboard. The game itself does not make any comparison between the tags sent by the players or any other quality control on these tags, because our research was only interested in the amount of work and not the quality of work done, but the players' marks were stored for future reference.

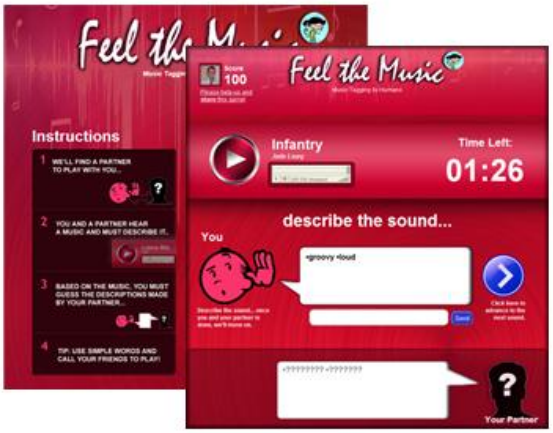

Figure 3 - *The game experiment*.

It has common features used by traditional games available on Facebook, such as an invitation interface to call friends to play, the publication on the wall about their game activities and a global ranking of the best players.

The game is simple. In each round, "two players" are given either the same audio clip. They are provided with a basic music player interface to start, stop, and adjust the volume of the audio clip to which they are listening. Each player describes the given audio clip by typing in any number of tags, which are not revealed to the partner. After both players have no more tags, the game reveals the result of the round to the players and presents the next round. Each round lasts two minutes in total. The player's score is based on the amount of tags and a random factor - Remember, the player's score has no importance for this research, but only the **amount** of rounds he plays.

4. Results

The game attracted **203** players in one month, but only **146** of them have been considered in the sample because they were not author's friends on Facebook,

which could skew the results. The sample summary is shown in figure 4.

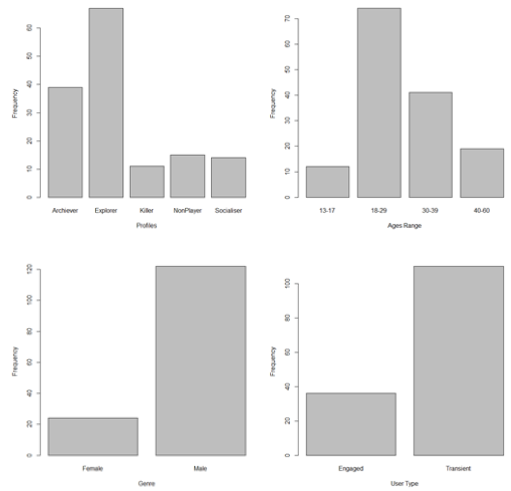

Figure 4 – *The sample summary*

Results showed that there was a predominance of **men** with **18-29** years old and been characterized as **explorer**, according to Bartle's typology. Only around **25%** of these players have been engaged in the game.

To investigate the influence of the factors with players' engagement, we start with a *fisher test* that shown there is a dependency between the player's engagement with all three factors - its sex, age and Bartle's typology. **Table 2** shows these results.

Table 2 – *The fisher test results*

Factor	p-value
Type	0.004877
Age	0.008364
Sex	0.01725

With the perception that these factors influence in user engagement, we conducted a *logistic regression* and obtain the coefficients shown in Table 3. The most direct way to associate a magnitude of change in probability due each independent variable is examine the coefficients. The coefficient with the highest value is the most influencer of that group and the sum of coefficients' group influences indicates the influence of the entire group in the engagement of the user. The group with the highest value has the greatest influence.

Table 3 – *Logistic regression coefficients*

Value	coefficient	Factor
profilesExplorer	-0,6741	
profilesKiller	-1,4827	**Type**
profilesNonPlayer	-0,6330	
profilesSocialiser	15,7254	
ages18-29	1,6113	**Age**
ages30-39	0,3470	
genresMale	1,2622	**Sex**

Looking at the results, the group Sex totals **1,2622**, the Age group totals **2,6152** and the Type group totals **12,6068**, so we can conclude that the user's Bartles' typology is the most important factor, followed by Age and Sex. Note that this order of importance is equal to the order of probability of influence that we find previously in Fisher test.

Figure 5 shows the graphs of the relations between Player Type, Genre, Sex and the player engagements status.

.

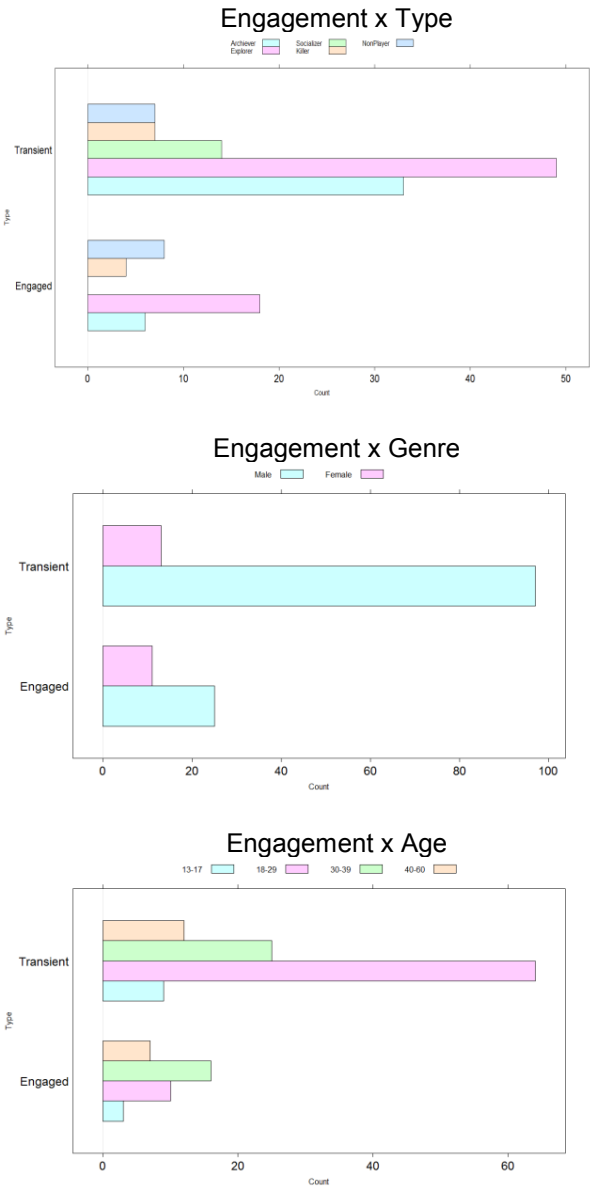

Figure 5 – *Player type and your engagement status*

5. Discussion

Surprisingly, few players in the sample defined themselves as "Socializers", which according to common sense, should be the majority among the users of a social network. This result may have been

influenced by how the game was published, primarily in communities of gamers available on Facebook. Many of these gamers are "hardcore gamers", i.e., spend many hours playing complex video games on consoles at home and do not necessarily fit the profile of the typical user of a social network, called "casual gamer", that plays only in their free time without a major engagement with video games.

Also, none of which identified themselves as "Socializers" were engaged in the game, demonstrating a high rejection of this profile to the game. One plausible explanation is that, in contrast to typical successful Facebook game, like *FarmVille*, that have shown a strong incentive to social interactions [Fang and Zhao 2010], but in our GWAP, the players relationships were ephemeral and anonymous. Put differently, the design of the game meant that collaboration was only needed for the duration of a game session and thus establishing relationships was not possible or necessary. Further, owing to the time limits imposed by the game, relationship building could not occur. In contrast, successful Facebook games require longer term use and provide a rich variety of socializing experiences for players. Evidence for this view can be found in the work of [Fang and Zhao 2010] that analyzed a variety of game genres and found that the influence of the socializers on intention to engage on game is based on the quality of social interactions offered by the game.

Another important result, almost all players who identified themselves as "non gamers" were engaged in the game. Probably for some scientific interest, altruism or by curiosity about a "serious" game with a serious purpose and not just another typical casual game available on a social network, they usually reject. This is a factor in favor of GWAPS, that even being video games can attract and engage even people not familiar to this type of entertainment. However, it should be noted that achieving sustained engagement is a complex issue because of the need to balance the twin goals of GWAPs – effective human computation and entertainment. A GWAP should support the generation of quality computations (tags in the case of audio tagging applications), but at the same ensure that entertainment is not sacrificed. From the perspective of the player, users need to be convinced that they can derive enjoyment through the game, thereby fostering sustained usage [Hsu and Lu 2007], and also persuaded that the outputs of the game have utility, thereby appealing to their sense of altruism and/ or other motivating drives.

Finally, the results showed that the engagement of players is only associated with its profile, according to Bartle's typology, i.e., Gender and Age as factors that drive the development of traditional games were not relevant to the engagement of the players in this game. This means that the design of GWAPS created in the traditional way, as our *Feel the Music*, games designers should focus their efforts on mechanics that are attractive to the types of players that wish to attract, without worrying about factors that will appeal to certain age or sex.

While our present work has yielded insights, the findings should be interpreted with caution for several reasons. One, we evaluated one particular type of GWAP in a specific domain of audio tagging. For better generalizability, it would be instructive to carry out investigations using different game designs and different domains of human computation. Two, our findings were obtained through a single experiment in a short period. Conducting a long-term study involving repeated use of the GWAPs over time would be beneficial in validating our results. Three, to provide more insights to our analysis, a wider range of factors could be included such as locale and educational level. Four, although influential, the Bartle's typology was not empirically tested and therefore using another typology can bring other insights about the relevance of this factor on user engagement.

6. Conclusion

Our work has succeeded in identifying which of all player types; the "socializers" were those who had higher rejection to the game. In contrast, players who defined themselves as "Explorers" were the most attracted and most engaged to the game. Also, the results showed that sex and age of the players seems to have low relevance on the influence on engaging players to the game.

In addition, our work suggests a number of design implications for existent GWAPs:

- Features for community-building should be designed. As discussed, typical GWAPs do not foster or need player relationships, possibly explaining why the Socializers have the higher rejection to the game. However, developing community-oriented features around such games provides an alternative. This could be in the form of discussion boards, chat rooms and social media sites where people share ideas, strategies or their experiences.
- Rewards for good performance should be considered. This is an extension of feedback, and has the potential to spur players to outdo others and promote sustained use. One such possibility is the leaderboard (as used in our game) that lists high performing players. For image tagging HCGs, this could be a list of top scoring players. For non-game variants, this might be users generating the most number of image tags.
- GWAPs should consider demonstrating how the generated computations are useful. This seems important because the involvement of non-players was quite high, apparently due to

the purpose of the game itself and it is reasonable to suggest that a more relevant purpose, them more "non-gamers" will be engaged in the game. Existing game-theoretic models for human computation have assumed that all agents participating in the system have selfish motivations. But there is evidence that users are behaving altruistically in peer-production systems such as Wikipedia, Yahoo! Answers, and YouTube. In fact, game-theoretic analysis is not constrained to selfish utility functions, and it will be useful to expand our models of what motivates users.

For future work, there is a need to evaluate the impact of the player's type in other GWAPs types and also evaluate the impact on the amount of work achieved when designing games more attractive to a specific profile. Finally, while our study has demonstrated that design a game with Bartle's typologies in mind, a gulf between engaged and transient users seems to exist, and further research needs to be conducted in designing games to realize the full potential of GWAPs.

Acknowledgements

We thank Nazareno Andrade and Raquel Queiroz for helpful conversations and feedback.

References

Bartle, R. Players who suit MUDs. Journal of MUD research, (1996).

Fang X, Zhao F. Personality and enjoyment of computer game play. Computers in Industry 2010; 61: 342–349

Goh, D. H., & Lee, C. S. (2011). Perceptions, quality and motivational needs in image tagging human computation games. Journal of Information Science, 37(5), 515-531

Goh DH, Ang RP, Chua AYK, Lee CS. Evaluating game genres for tagging images. In: Blanford A, Gulliksen J (eds) Proceedings of the 6th Nordic Conference on Human–Computer Interaction. New York: ACM Press, 2010, pp. 659–662

Ha I, Yoon Y, Choi M. Determinants of adoption of mobile games under mobile broadband wireless access environment. Information & Management 2007; 44: 276–286

Ho CJ, Chang TH, Lee JC, Hsu JYJ, Chen KT. KissKissBan: a competitive human computation game for image annotation. In: Bennett P, Chandrasekar R, Chickering M, Ipeirotis P, Law E, Mityagin A et al. (eds), Proceedings of the 2009 ACM SIGKDD Workshop on Human Computation. New York: ACM Press, 2009, pp. 11–14.

Hofer, C., & Eport, S. T. I. T. E. R. (2011). A SURVEY ON GAMES FOR KNOWLEDGE ACQUISITION. Technology

Hsu CL, Lu HP. Consumer behaviour in online game communities: a motivational factor perspective. Computers in Human Behaviour 2007; 23: 1642–1659.

Hunicke, R., LeBlanc, M. and Zubek, R. (2004). MDA: A Formal Approach to Game Design and GameResearch. Challenges in Game Artificial Intelligence. 2004 AAAI Workshop.

Law, Edith and von Ahn, Luis. Input-agreement: A New Mechanism for Data Collection Using Human Computation Games. To Appear in CHI 2009.

Lazzaro, N. (2004). Why We Play Games: Four Keys to More Emotion Without Story. XEODesign, Inc. Available at: http://xeodesign.com/xeodesign_whyweplaygames.pdf (accessed May 2012)

Lee CS, Goh DH, Chua AYK, Ang RP. Indagator: investigating perceived gratifications of an application that blends mobile content sharing with gameplay. Journal of the American Society for Information Science and Technology 2010; 61: 1244–1257.

von Ahn L, Dabbish L. (2008). Designing games with a purpose. Communications of the ACM 2008; 51: 58–67.

von Ahn L, Dabbish L. Labeling images with a computer game. In: Dykstra-Erickson E, Tscheligi M (eds) Proceedings of the 2004 Annual SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2004, pp. 319–326.

Ryan RM, Rigby CS, Przybylski AK. The motivational pull of video games: a self-determination theory approach. Motivation and Emotion 2006; 30: 347–364.

Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, & Foldit players, "Predicting protein structures with a multiplayer online game", Nature, Vol. 466, N. 7307, 756-760, 2010

Siorpaes K, Hepp M. Games with a purpose for the Semantic Web. IEEE Intelligent Systems 2008; 23: 50–60.
Lee CS. Managing perceived communication failures with affordances of ICTs. Computers in Human Behavior 2010; 26: 572– 580.

Yee N. Motivations for play in online games. Cyberpsychology & Behavior 2006; 9: 772–775