Spring 2019

Exploring the Gamification Paradox: Why Does Improved Engagement Not Lead to Improved Performance?

Katarzyna Maria Sliwinska
San Jose State University

EXPLORING THE GAMIFICATION PARADOX:
WHY DOES IMPROVED ENGAGEMENT NOT LEAD TO IMPROVED PERFORMANCE?

A Thesis

Presented to

The Faculty of the Department of Psychology

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Katarzyna Sliwinska

May 2019

The Designated Thesis Committee Approves the Thesis Titled

EXPLORING THE GAMIFICATION PARADOX:
WHY DOES IMPROVED ENGAGEMENT NOT LEAD TO IMPROVED PERFORMANCE?

by

Katarzyna Maria Sliwinska

APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

SAN JOSÉ STATE UNIVERSITY

May 2019

Evan Palmer, Ph. D.	Department of Psychology
Sean Laraway, Ph. D.	Department of Psychology
Susan Snycerski, Ph. D.	Department of Psychology

ABSTRACT

EXPLORING THE GAMIFICATION PARADOX:
WHY DOES IMPROVED ENGAGEMENT NOT LEAD TO IMPROVED PERFORMANCE?
by Katarzyna Sliwinska

Gamification is the application of game elements to non-game environments (Deterding, 2012), and is often used to engage people and make their experiences more enjoyable in areas ranging from fitness and education to psychological research. Previous studies have shown that adding gamification to new environments can result in increased motivation (Hamari, Koivisto, & Sarsa, 2014). However, increased motivation from gamification does not seem to increase performance in terms of accuracy or response times (Hawkins et al., 2013). This research study examined this “gamification paradox” by testing performance of 87 participants on a visual search task both with and without gamification elements. We found no difference in terms of intrinsic motivation between participants in the gamified and non-gamified conditions. Additionally, the two conditions did not significantly differ in their performance. However, we did find that motivation was related to performance in terms of accuracy. We also found that our point formula altered participant behavior, such that participants emphasized accuracy over response time. These findings suggest that game elements, such as points, can affect participant behavior. However, because the implementation of gamification failed to sufficiently motivate participants, we were unable to see whether gamification can increase participant performance

ACKNOWLEDGMENTS

First and foremost, I would like to thank my thesis advisor, Dr. Evan Palmer, for his continuing support, patience, and humor, throughout this process. Dr. Palmer provided me with valuable and constructive feedback during the development of this research. His contagious passion for psychological research has been inspiring and encouraging. Without his guidance this work would not have been possible. Second, I would like to thank my committee members, Dr. Susan Snycerski and Dr. Sean Laraway, for their thoughtful comments and for their valuable time. I would also like to thank my research assistants, Sumaria Ilag and Jennifer Li, who have spent countless hours running participants and collecting data for this research project. Without them, I would not have been able to complete the study.

Next, I would like to thank my family for their support and encouragement to pursue a master's degree. To my parents, Elżbieta Kołodziej and Paweł Śliwiński, who have raised me with curiosity, compassion, and confidence to pursue my endeavors and venture optimistically to where life takes me. To my brother, Jakub Śliwiński, and his family, Martyna and Tymek, for keeping my life full of love and humor, even from a distance. Last but certainly not least, I would like to thank Dr. Beata Czajkowska and Dr. David Skully, for their thoughtful advice and their financial support throughout this journey. You are my real-life role models and you inspired me to continue my education and pursue my career with confidence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Gamification, or the application of game elements to non-game environments (Deterding 2012), has been widely used in various settings to increase human productivity and motivation (Fitz-Walter, 2016). The gamification market was estimated to be worth $1.6 billion in 2015 and predicted to reach $22.9 billion by 2022 (P&S Market Research, 2017). Gamification has been attracting high and sustained interest in online searches since 2010 according to Google Trends (Figure 1) and research on the topic has also increased in the same time frame (Figure 2).

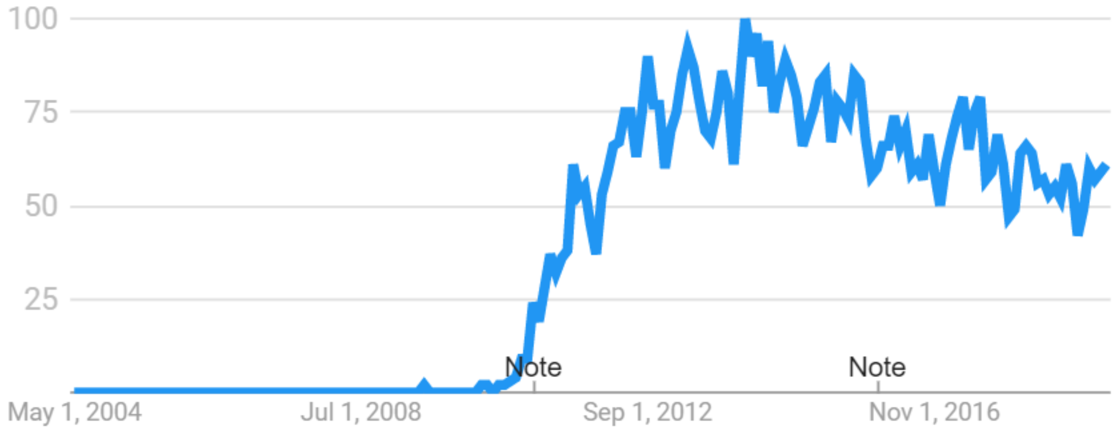


Figure 1. A steady increase of interest in the term "gamification" over time (Google Trends, 2017).

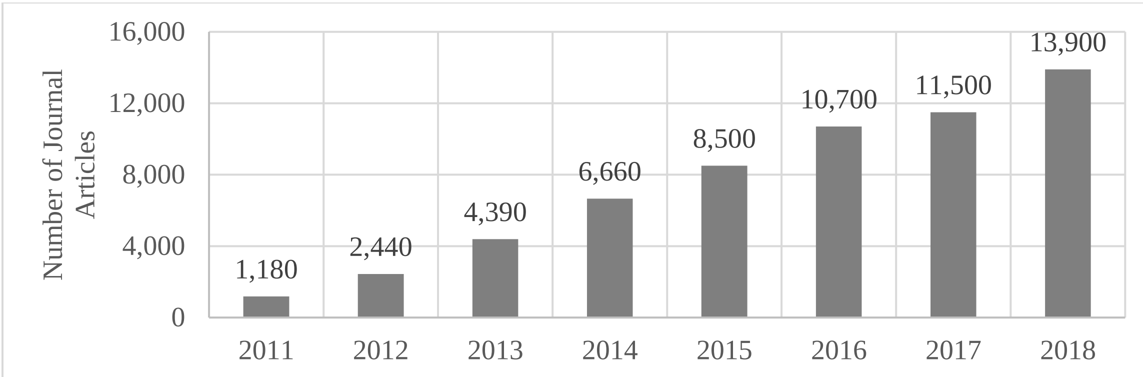


Figure 2. The number of journal articles on the topic of gamification based on search hits in Google Scholar.

Gamification has become popular in a variety of fields ranging from business (Frost & Sullivan, 2017), to healthcare and research (Mesko, 2017), and academics (Davis, 2014) because it engenders greater motivation from users in a variety of contexts (Banfield & Wilkerson, 2014; Deterding, 2012; Fotaris et al., 2016; Peng et al., 2012; Rodrigues, Oliveira, Costa, 2016).

However, despite being more motivated and engaged in gamified tasks, people do not seem to perform any better in those tasks (Hawkins et al., 2013). Given that increased motivation and external incentives normally lead to improved task performance (Cerasoli, Nicklin, & Ford, 2014) and given that gamification leads to increased motivation (Hamari, Koivisto, & Sarsa, 2014), why does gamification not lead to improved task performance?

This study examined several possible explanations for why improved engagement from gamification might not lead to improved performance. One possibility was that previous studies eliminated bad performers or bad trials during data cleaning and screening that would have shown such differences in performance. Another possibility was that previous implementations of gamification did not increase performance because the elements used did not specifically reward improved performance. Finally, the third possibility was that gamification improves motivation but not performance, regardless of data cleaning or strength of implementation.

This introduction is divided into five sections: principles of gamification, examples of gamification, the gamification paradox, study relevance, and the experiment. In the first section, we will discuss the theories behind gamification, such as the self-determination theory and operant conditioning theory. The second section will provide the applications of gamification in the industry, in academics, and in research. In the next section, we will discuss the idea of the gamification paradox and why gamification may not improve performance. Next, we will discuss the relevance of the study, following which we will introduce the experiment and our hypotheses.

Principles of Gamification

In this section we will analyze the leading theories behind gamification. First, we will discuss gamification in terms of its relation to games and Self-Determination Theory. After that we will discuss gamification principles from the perspective of behavioral theory, specifically operant conditioning.

Since gamification uses elements taken directly from games (points, badges, avatars), understanding how and why specific game elements engage and motivate players within games is central to understanding how gamification works. It has been hypothesized by Ryan and Deci (2000; 2017) that the principles of Self-Determination Theory (SDT), which help to explain human engagement and enjoyment in activities, also apply to games. SDT proposes that supporting the human needs of relatedness, competence, and autonomy leads to optimal functioning and can foster the highest quality of motivation (Ryan & Deci, 2000; 2017). Relatedness refers to the feeling of connection with others – in the context of games it can mean other players or even characters in the game itself. Feelings of competence occur when people are being challenged appropriately and can showcase their skills, fostering the feeling of success. Players experience autonomy in games when they have opportunities to make their own choices, for example selecting an avatar or choosing where to travel next in the virtual world (Ryan & Deci, 2017). Games that satisfy all three needs tend to be more enjoyable and more engaging for the player, as demonstrated by Tamborini, Bowman, Eden, Grizzard, and Organ (2010). In their study, the authors found that SDT's basic need satisfaction accounted for over 50% of the experienced enjoyment of the bowling simulator game used in their study. Additionally, they manipulated elements of the game to see how they affected each need (relatedness, competence, and autonomy). The results showed, for example, that offering more multiplayer options increased the satisfaction of the relatedness need. Autonomy need was satisfied through players' perceived game skill and natural mapping of the controls. This relationship was confirmed through statistical analysis, which showed a positive path coefficient between autonomy and self-efficacy, and between autonomy and natural mapping. In conclusion, taking care to satisfy the basic needs of relatedness, competence, and autonomy through specific game elements can lead to more engagement and enjoyment experienced by the player (Tamborini et al., 2010). Because gamification is the application of game elements to non-game environments, it follows that simply using appropriate elements – those that satisfy SDT needs – could result in increased engagement.

Zichermann and Cunningham (2011) speculated that gamified systems are also powerful because they engage the dopamine system in our brains, which is associated with learning and pleasure. Experiences considered rewarding, valuable, or surprising trigger a dopamine release, which then becomes associated with the activity (Robinson, Sandstrom, Denenberg, & Palmiter, 2005). Therefore, gamified systems reinforce engagement by continuously providing rewards that seem valuable to the user and because they satisfy the human needs of relatedness, competence, and autonomy.

Another explanation for the mechanics behind gamification comes from behavior theory, specifically operant conditioning. According to Skinner (1938) and many others, a behavior will be repeated when it is associated with a strong reinforcer. Reinforcers can take the form of simple rewards, such as those seen in gamified environments (points, badges, high scores, etc.). Reinforcing high performance is possible when the reinforcement (reward) reduces the aversiveness towards high effort (Eisenberger, 1992; Eisenberger & Cameron, 1996). Therefore, rewards commonly seen in gamified tasks may lead to behavior association and the repetition of said behavior. Rewards that are specifically associated with performance will also result in increased effort and performance.

A meta-analysis by Cameron and Pierce (1994) found that, overall, rewards do not decrease intrinsic motivation and in fact can increase intrinsic motivation. They found that performance-independent rewards and completion-dependent rewards are the only type of rewards that may negatively influence intrinsic motivation and in turn performance. Quality-dependent rewards, also called performance-contingent rewards, on the other hand, reinforce behavior based on increased effort and therefore may lead to both increased intrinsic motivation and increased performance (Cameron & Pierce, 1994; Pierce, Cameron, Banko, & So, 2003).

Examples of Gamification

Gamification in industry. One of the most successful commercial applications of gamification was by Fitbit, with 22.3 million smartwatches sold worldwide (Gordon, 2017). The purpose of adding gamification into their product was to make people more engaged in exercising by making it fun. The company incorporated game elements such as badges, points, and leaderboards into their user interface to foster engagement among consumers and encourage them to get more exercise (Figure 3). This implementation resulted in the company's immense success among the industry of wearable fitness products. Their current revenue increased to $1.8 billion in a span of 5 years (Gordon, 2017).

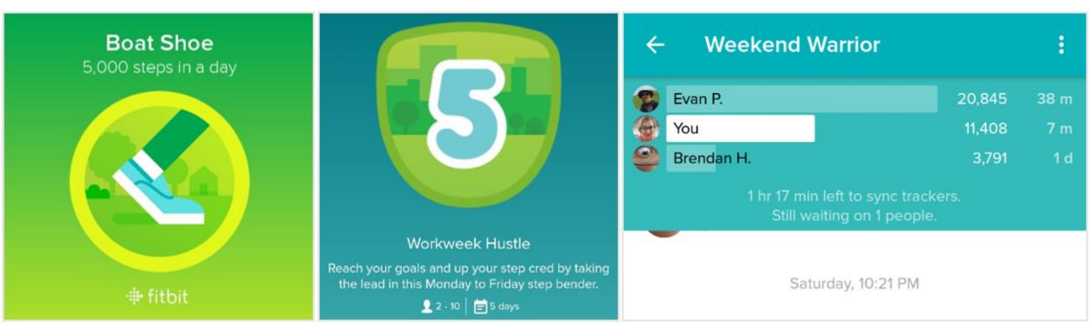


Figure 3. Example of game elements in the Fitbit fitness mobile application (from left to right: badge for taking 5,000 steps in a day; a challenge to compete with friends, a leaderboard).

Similarly, Six to Start, a game development company based in London, wanted to encourage exercising by making running fun. In 2012, the company created the *Zombies, Run!* fitness mobile application. The application turned running into a game, set in a post-apocalyptic world filled with zombies. Runners take on a role of a zombie apocalypse survivor by listening to the app narrative. At some point they start hearing groans and gutteral breathing, which indicates they are being chased by zombies and must run faster! By adding stories, missions, and collectibles, Six to Start, gamified the action of running and motivated millions of players worldwide to download their app (Jordan, 2017). *Zombies, Run!* continued to attract players, and as of 2017 it has 250,000 monthly active players.

Another example of commercial use of gamification was an online banking application designed by Rodrigues, Oliveira, and Costa (2016). The specific game elements included in the application were avatar customization options, game-like graphics, and a story that framed banking as a soccer game. The gamified version of the banking software was rated as more enjoyable and easier to use than the non-gamified version. The authors concluded that enjoyment and ease-of-use should ultimately lead customers to use the e-banking applications more and to become more loyal to the product. The goal of the implementation was to make e-banking easier and more engaging; adding game elements to the software accomplished that goal.

To test whether gamification affects customer behavior, Hamari (2013) implemented badges in an online trading community (*Sharetribe*). The service allows users to borrow and lend, as well as sell and buy goods in smaller communities called tribes. Users were able to unlock these badges by performing typical actions within the service, such as logging into Sharetribe five days in a row. The badges were then displayed on each user's profile. The result of the implementation showed that users who monitored their badges showed higher user activity than users who did not monitor or have the badges enabled. In conclusion, customer behavior was affected by the addition of badges for the users who were interested in them.

Gamification has also made its way into the automotive industry. An initiative by Volkswagen attempted to change people's behavior by making the desirable driving behaviors fun (Volkswagen, 2009). The car company made following the speed limit a game by adding incentives in the form of a lottery for safe drivers (*Speed Camera Lottery*). This innovation reduced the average speed of passing drivers by 20%.

To reduce fuel consumption, smartphone app developers and automobile manufacturers began developing ways to encourage better driving habits (Gibson, 2015). Honda, Toyota, and Ford have developed their own on-board systems, which keep track of car mileage and show driving efficiency. Ford's SmartGauge system, provides visual feedback in the form of growing green leaves indicating better fuel economy while driving (Wojdyla, 2008).

In summary, gamification has been applied in many industries, including fitness, banking, and the automotive industry. Gamification can engage customers by making regular tasks, such as getting exercise or following the speed limit, both fun and motivating.

Gamification in academia. Gamification has also been applied to educational contexts. In general, gamification interventions have been successful in increasing student engagement and overall enjoyment of their learning environment, though there are some important exceptions. It is important to note that the successful implementations all have had similar characteristics; namely, they applied the principles of relatedness, competence, and autonomy from Deci & Ryan's (1985) Self-Determination Theory of intrinsic motivation.

For example, when gamification techniques were applied to a programming class (Fotaris et al., 2016) the approach turned out to be motivating and enriching for both students and teachers. The gamification design for the programming class was complex and multimodal. Teaching was done in multiple forms through gamified online lectures and assessment programs. When developing the gamified teaching program, Fotaris et al. (2016) applied SDT's principles of relatedness, competence, and autonomy, through competitive game play and collaborative problem solving. To foster relatedness, collaborative play was presented in the form of a "Who Wants To Be A Millionaire?" (*Millionaire*) game procedure. Millionaire is a game show in which players answer multiple-choice questions to win the top prize of $1,000,000 (in the case of the class, students played for points). The game involves three "lifelines": (1) asking a friend for help, (2) asking the audience to vote on the answer, or (3) eliminating 2 out of 4 answer choices. The same rules applied in the classroom version of the game. To satisfy the need for competence, students were awarded points by attending lectures and participating in learning games, as well as through earning achievements through the online learning platform *Codecademy*. Through the quiz competitions of Millionaire and "Kahoot!" (a multiple choice clicker activity) students also experienced freedom of choice, which produced the feeling of autonomy. In Millionaire, students picked whether they wanted to answer a question, get a hint, poll the class, or ask a friend. In Kahoot!, autonomy was fostered through the choice of nickname and avatar. In the end, students' enjoyment and the feeling of confidence in their learning grew over time in the class, additionally students in the gamified class had better academic performance than the students in the control class (Fotaris et al., 2016).

Students' intrinsic motivation was increased through gamification in a system administration course developed by Banfield and Wilkerson (2014). In this case, the game element used was a story, which created a more relaxed and fun atmosphere for students during learning. It is worth noting that the students were free to choose how to complete the gamified activity, fostering autonomy. Additionally, the story in the form of a case study provided a challenging and interesting environment to learn in but did not require a specific approach to successfully solve the tasks (competence and autonomy). At the end of the gamified task, the students had an opportunity to talk about their approaches to solving the tasks, which provided a sense of relatedness. In the end, the students in the gamified group had an increased sense of intrinsic motivation and self-efficacy. Banfield and Wilkerson did not report academic performance of students in either condition.

When game elements were added to courses without the consideration of student needs, however, the implementation had negative results. Hanus and Fox (2015) gamified a communications course by adding points, badges, and leaderboards. The gamification elements were posed as mandatory activities with specific instructions on how to complete them. Students did not have a choice about whether to engage in the gamification or in the way they engaged in it. Thus, the gamification intervention did not allow for participant autonomy, which is one of three human needs that games satisfy according to SDT (Ryan & Deci, 2000; 2017). As a result, participants' motivation had decreased and in turn so did their final exam scores. The lack of autonomy could have contributed to the negative effects of this implementation.

A systematic literature review conducted by Dicheva et al. (2014) revealed that the most common game elements used in the gamification of educational programs were badges, points, and leaderboards. The majority of the analyzed papers (18 out of 34) reported positive results in terms of student engagement and quantity of work without a reduction in their quality. The analysis demonstrated that applying gamification in educational contexts may lead to mixed results, but is likely to yield more positive results, especially in terms of student engagement.

Gamification in research. Lumsden et al. (2016) found that many researchers, especially in the cognitive psychology domain, use gamification in their experiments to increase long- and short- term engagement, usability, ecological validity, and stimulation. There are many possibilities for the application of gamification in research environments. The following studies have shown the advantages of gamification in increasing participant motivation, engagement, and enjoyment.

Porter (1995) used an arcade game called Save the Whale instead of a traditional experimental task to demonstrate that gamified tasks can be meaningfully used in psychological research studies. The game involved players maneuvering a whale to eat plankton or destroy kayaks. Performing either of these two actions awarded points to the player. The plankton task was a version of a random tracking task, while the kayak task was predictable motion trajectory tracking task. The game promoted player autonomy by allowing the players to choose their own strategies and by steering the whale. It satisfied the need for competence through varied difficulty of the tasks and feedback from points. Porter’s findings confirmed that performance on gamified tasks can be successfully used as a measure in laboratory experiments. It is important to note that Porter did not include a control group without gamification, which would have allowed elimination of alternative explanations. The proposed study, however, will consist of both a gamified and a non-gamified version of the same visual search task.

Brewer et al. (2013) motivated children to complete experimental tasks through the use of points and prizes. Originally, the study attempted to collect data on child-computer interaction without added gamification. The non-gamified tasks required children to draw six gestures on a touch-screen with their finger and to touch a target square as quickly and accurately as possible. Without game elements the task was repetitive and uninteresting to children, which resulted in low completion rates. The gamified versions of the tasks awarded points for completed gestures and correctly targeted squares. At the end of the tasks children received prizes based on how many points they earned while completing the tasks. After adding gamification, children's motivation to complete the tasks increased, which resulted in study completion rates rising from 73% to 97%.

Similarly, to increase motivation and engagement of study participants, Miranda and Palmer (2014) used points and sound effects. The gamified experiment was a visual search task, in which participants had to determine whether the line in a circle of the target color was horizontal or vertical. Each correct response awarded points to the player, with more points being awarded for faster responses. The player could also see a previous high score on the screen, their total current score, and the streak of correct responses, which awarded bonus points. The sounds were played during the delivery of bonus points. The points were found to be the motivating factor, while the sound effects functioned as a powerful reinforcer that served the same purpose as money in other studies. Overall the gamified task was perceived as enjoyable by study participants, however, no differences were found in their performance between the gamified and non-gamified conditions.

Peng, Lin, Pfeiffer, and Winn (2012) incorporated autonomy- and competence-supportive game features into an exercise game, which resulted in increased participant motivation, engagement, and enjoyment. Peng et al. designed an exercise game titled *Olympus*. The game used a story of ancient Greek athletic training to drive gameplay. To promote autonomy, the game allowed players to customize and upgrade their characters and choose responses to non-player characters. The game offered a dynamic difficulty mechanism and in-game feedback, which promoted competence. The authors suggested that the choice elements fulfilled the human need for autonomy, and the dynamic difficulty and feedback satisfied the need for competence. The study also demonstrated that game elements designed to specifically satisfy the human needs for competence and autonomy have positive effects on motivation and engagement of participants (Peng et al., 2012). In summary, game features can be used to motivate, engage, and increase people's enjoyment.

The Gamification Paradox

Previous studies on motivation have demonstrated that higher motivation is associated with higher performance. Most notably, Cerasoli, Nicklin, and Ford (2014) conducted a 40-year meta-analysis of studies linking motivation to performance. By evaluating studies from various domains, such as education, employment, and sports, they were able to show that motivation predicts performance (ρ = .21-45). However, as expanded upon below, the same has not been consistently found for individuals motivated by gamification, as evidenced by the Hawkins et al. (2013) and Miranda and Palmer (2014) studies. This seems paradoxical since games foster playfulness, which increases motivation (Paras, 2005) and it is generally assumed that more motivated individuals desire to and tend to do better on tasks (Atkinson & Litwin, 1960; Nicholls, 1984).

Hawkins et al. (2013) showed that while gamification did increase participant engagement, it did not improve performance. The Hawkins et al. (2013) study involved two versions of two simple cognitive tasks. In the first experiment, participants were required to make judgments about the number of dots appearing in squares. They had to select the square that accumulated the dots faster, as quickly and accurately as possible.

When the target square was correctly identified, the participant would receive visual feedback in the form of a green border around the selection. When one of the distractors was misidentified as the target, its border would turn red and the correct square would highlight green. Both versions of the task included this form of feedback. The gamified version of this task also included a story, graphics, and audio-visual feedback for both correct and incorrect answers.

In the second experiment, participants had to decide which side (left or right) had more targets in a row based on previous knowledge of how the targets were distributed (Figure 4). To optimize their performance on the task, participants had to make changes in their decision making based on the “payoff” of their previous decision. Feedback in the form of a green tick was displayed on the selection with a spotted square and a red X was displayed on the selection without the spotted square. The payoff squares appeared in batches. Sometimes more payoff squares would appear on the left and sometimes they appeared on the right. In the gamified condition, the spotted squares were replaced with ghosts. The positive feedback in the form of a green circle was displayed on the selection of a ghost and the negative feedback was displayed on the selections without a ghost. The participants were told about a point-based scoring system, which counted each captured ghost as one point. The points were visible to the participant during the task and they appeared after each block.

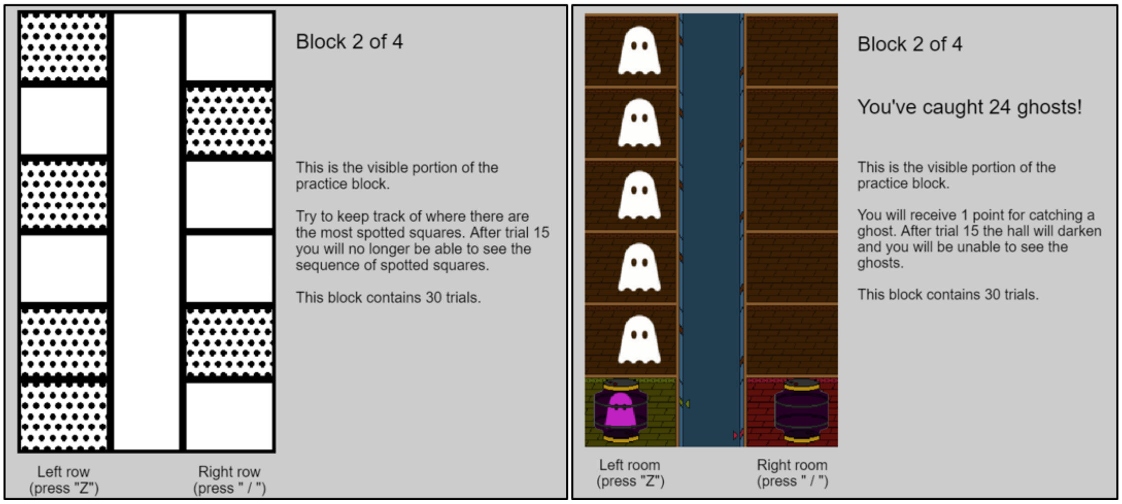


Figure 4. The non-gamified target square distribution task (left) and the gamified version of the same task (right).

At the end of each experiment the participants answered 11 questions about their experience with the corresponding task. The questions related to the understanding of the task, as well as participants' subjective enjoyment, motivation, interest, and effort during the task. Both experiments resulted in no differences between performance on the gamified and standard tasks and minor differences in the motivation and engagement of participant in the two conditions.

The study by Miranda and Palmer (2014), which gamified a visual search task with points and sounds, also showed a lack of correlation between motivation and performance. While participants' motivation was higher in the gamified conditions, this motivation did not lead to better performance as measured by response time and accuracy. No significant differences in performance were found between the gamified and non-gamified versions of the tasks.

There may be a variety of reasons why motivated participants did not perform better than non-motivated ones.

Why Might Performance Not Improve in Gamified Tasks?

There are several possible explanations for why performance did not improve in the Hawkins et al. (2013) and Miranda & Palmer (2014) studies. It is possible that the poor performers or bad trials were removed from analysis during data cleaning and screening. Hawkins attempted to account for this by showing that the proportion of participants excluded did not differ between conditions in Experiment 1 ($\mathcal{X}^2 = 0.08$, $p = .77$) and in Experiment 2 (6.8% of outliers in the nongame and 5.8% of outliers in the gamified condition). Nonetheless, data cleaning and screening could have affected the results through the exclusion of individual trials, which accounted for 2.46% of the total number of trials or because of the outliers excluded from the analysis. Because of the possible effect of data cleaning on the study results, the current study will examine the experimental results both with and without data cleaning. Miranda & Palmer (2014) excluded participants that had response times or accuracy scores that were more than two standard deviations worse than the rest of the sample during the training phase. They also excluded RTs < 200 and > 3,000 ms. In the current study, we based our data cleaning methods on the methods established by Miranda & Palmer (2014).

Another reason for the lack of improvement in performance among the gamification participants in Hawkins et al. (2013) could be that the game elements chosen were not impactful enough on their own to make a difference. Graphics and audio-visual feedback are game elements not tied to performance. This can be seen in the self-reported experience measure, which showed very little difference between the conditions in each of the questionnaire answers. The overall significant differences in rated experience was due to significant differences in just two out of eleven survey questions. The significant differences were for how interesting the graphics were and whether participants enjoyed the task. This suggests that the gamification used in the experiment did not affect participants on dimensions other than enjoyment and interest in the graphics. If the experiment had used additional game elements, such as points, streaks, bonuses, leaderboards, or high scores (game elements tied to performance) the results might have shown increased performance as well as increased engagement and motivation. Points, sounds, high scores, and leaderboards seem to have a stronger effect on motivation than graphics because they satisfy the three needs of SDT (Ryan & Deci, 2000; 2017) and can be considered rewarding reinforcements (Skinner, 1938). The current study will therefore use elements tied to performance as well as graphics in the gamified condition to assess the hypotheses.

Lastly, it could be that gamification just does not improve performance. It may create more motivated participants, but their performance is not dependent on the motivation that resulted from gamification. This study examined whether gamification does indeed result in increased motivation and whether this leads to improved performance.

Study Relevance

Understanding how gamification affects peoples' motivation and engagement will be important for industry professionals, who aim to motivate people to use their product, keep them engaged during use, and make the whole experience more enjoyable. Boring tasks, such as lengthy surveys, certain therapeutic interventions (e.g. behavioral interventions for clinical populations), fitness plans, etc., would all benefit from the addition of motivating and engaging factors that come with gamification.

In the research domain, the results of this study will provide a method for implementing gamification in various experimental tasks and show its potential effects on data collection and participant performance. The method may also be used as a tool for engaging participants in studies. Currently, research labs rely on undergraduate psychology students' participation or money incentives. If gamification could improve motivation in participants, they might be more likely to recommend the study to a friend, increasing participant recruitment (Miranda & Palmer, 2014).

Statement of Purpose

The purpose of this experiment was to determine whether gamified tasks positively affect participant motivation and engagement during experiments, and whether that ultimately leads to improved participant performance, after accounting for data cleaning and screening. To test these ideas, we used two versions of a traditional visual search task: a standard task and a gamified version of the task. The visual search task was a spatial-configuration search task, in which participants had to search for a randomly oriented target T among randomly oriented distractor Ls. Set size and target presence was also varied since these factors are well-known to affect performance and served as a manipulation check for the study. Target presence and set size was randomly distributed among the trials. Set sizes was either 10 or 20 items per trial, with half of the trials being small (10 items) and half being large (20 items) set sizes. Based on previous visual search research, we expected to see slower response times for larger set size trials (e.g., Wolfe, 1994) and for target absent trials (e.g., Treisman & Gelade, 1980; Chun & Wolfe, 1996).

Gamification, the between-subjects independent variable, was defined as the application of game elements to the experimental task. The gamification elements for the task included points, high scores, sounds, a story, and graphics. Points and high scores, the game elements tied to performance, were used as quality-dependent rewards to motivate correct responses on each trial (Cameron & Pierce, 1994). Points also allowed participants to feel a sense of accomplishment, which fostered the need for competence (Deci & Ryan, 1985). High scores allowed the participant to feel connected to previous players, fostering relatedness. A story, graphics, and sounds were used to emotionally engage the participant in the visual search task and make the task appear more game-like (Fullerton, 2008).

Participant motivation was defined as motivation to perform an action because it is inherently enjoyable and interesting, as measured by the Intrinsic Motivation Inventory (Ryan, 1982). Visual search performance was defined as the percent of correct trials and reaction times on the task, as well as the total score achieved by the participant during the visual search.

To evaluate the data cleaning & screening hypothesis, we adopted normal data cleaning procedures. This means counting the number of trials that would typically be screened out, based on reaction times less than 200 ms or greater than 3,000 ms, as well as counting the number of participants who would typically be removed from the study because their overall proportion correct is more than two standard deviations below the mean. Analyses of the data was performed both with and without the excluded trials to assess the effect these procedures on the gamification paradox.

Hypotheses

The first set of hypotheses evaluated the implementation of the visual search task (Figure 5). If the visual search task performance was as expected from the literature, we should have seen a difference in participant performance between the two set size displays, as well as a difference between target present and target absent displays. Generally, miss errors are more common than false alarm errors in visual search tasks (Wolf, 1998). Therefore, if our implementation of the visual search task was successful, we would see that target presence affected participant performance.

Hypothesis 1: Participants will have better visual search performance in the lower set size condition.

Hypothesis 1a: Average response times will be lower in the smaller set size condition.

Hypothesis 1b: Average proportion correct will be higher in the smaller set size condition.

Hypothesis 2: Target presence will affect visual search performance.

Hypothesis 2a: Average response times will be lower in the target-present condition.

Hypothesis 2b: Average proportion correct will be lower in the target-present condition.

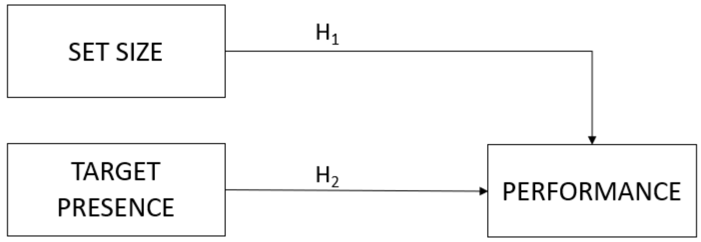


Figure 5. Model of the set size and target presence relationships to participant performance.

The second set of hypotheses focused on the main purpose of the study, whether gamification affects participant motivation and performance. Additionally, we also tested data cleaning and screening effects. See Figure 6 for hypotheses model.

Hypothesis 3: Gamification is related to motivation such that participants will have a higher total score on the Intrinsic Motivation Inventory in the gamified condition.

Hypothesis 4: Motivation is related to performance, such that participants with higher total scores on the Intrinsic Motivation Inventory will perform better in the visual search task.

Hypothesis 4a: IMI score will show a negative correlation with average response times.

Hypothesis 4b: IMI score will show a positive correlation with average proportion correct.

Hypothesis 5: Gamification is related to performance such that participants will have better visual search performance in the gamified condition.

Hypothesis 5a: Average response times will be lower in the gamified condition.

Hypothesis 5b: Average proportion correct will be higher in the gamified size condition.

Hypothesis 6: The gamified version of the task will yield fewer trials that would be excluded using normal data cleaning and screening procedures.

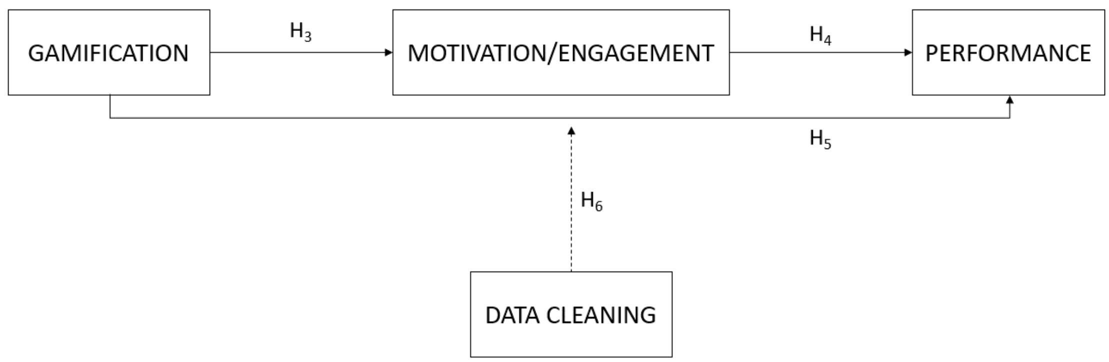


Figure. 6. The main constructs of interests (gamification, motivation, performance, and data cleaning effects) and the relationships between the constructs are laid out in the above diagram.

Methods

Participants

G*Power 3.1 software was used to conduct an a priori power analysis. The *ANOVA: Repeated measures, between factors* formula was run with a moderate partial eta squared (η_p^2) effect size value (.30), alpha level of .05, power level of 0.85, two groups, four measurements, and default correlation among measurements of 0.5. The calculated total required sample size was N = 66. To increase power and account for participants who would decide to leave the study before completing all of its parts, a larger sample was recruited. A sample of 87 participants aged 18 to 28 (M = 18.8, SD = 1.17) were recruited through the San Jose State University's SONA system and awarded partial course credit for their participation. The study required participants to have normal or corrected-to-normal vision (glasses/contacts acceptable). The participants included in the analyses were 34.1% (29) male and 65.9% (56) female.

Most studies on gamification did not report effect sizes. Studies on motivation and performance with indirect incentive reported a wide range of effect sizes with the calculated average of .30 (Cerasoli, Nicklin, & Ford, 2014). In the visual search domain, studies typically "involve 5 to 15 subjects performing a few hundred trials each" (Wolfe, 1998, pg. 33). Therefore, a larger total sample of 87 participants was deemed to be sufficient to cover the hypotheses for the remaining tests.

Materials

The experiment was programmed in the MATLAB 2017A (Natick, MA) using the open-source Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997, Kleiner et al, 2007). Stimuli were presented, and responses gathered on one of three Mac Mini computers (two 1.4 GHz, one 2.3 GHz) with 4 GB RAM and Apple Extended Keyboards. The Mac Mini computers were attached to three identical 21" Dell P2317H monitors at 1024 x 1200 pixel resolution running at 60 Hz. Participants wore Amazon Basics headphones while performing the visual search task, to help isolate possible environmental noises. Participants sat approximately 50 cm from the computer monitor and responded to the tasks by pressing appropriate keys on identical keyboards.

The Intrinsic Motivation Inventory (IMI), a self-report measure developed by Deci and Ryan (1985) and validated by McAuley, Duncan, and Tammen, (1989) was administered using the Qualtrics website. Overall coefficient alpha of .85 for the test indicated adequate reliability. See Appendix A for this scale.

Design

The study utilized a 2 (target present/absent) x 2 (set size) x 2 (gamification) mixed design, with target presence and set size manipulated within subjects and gamification manipulated between subjects. Target presence and set size were randomized within each block of trials following the method of constant stimuli. To avoid carry-over effects of the gamification intervention, each participant was assigned to one of two groups: the gamified group or the control group. The gamified group (G) included the "treatment" in the form of gamification, while the control group (NG) received no treatment, and thus lacked gamification. The gamified portion of the visual search task included a story (Table 1 & Appendix B), points, correct trial streak counter, and a genuine high score from a previous participant (Fig *7*). The control condition did not include any gamified elements.

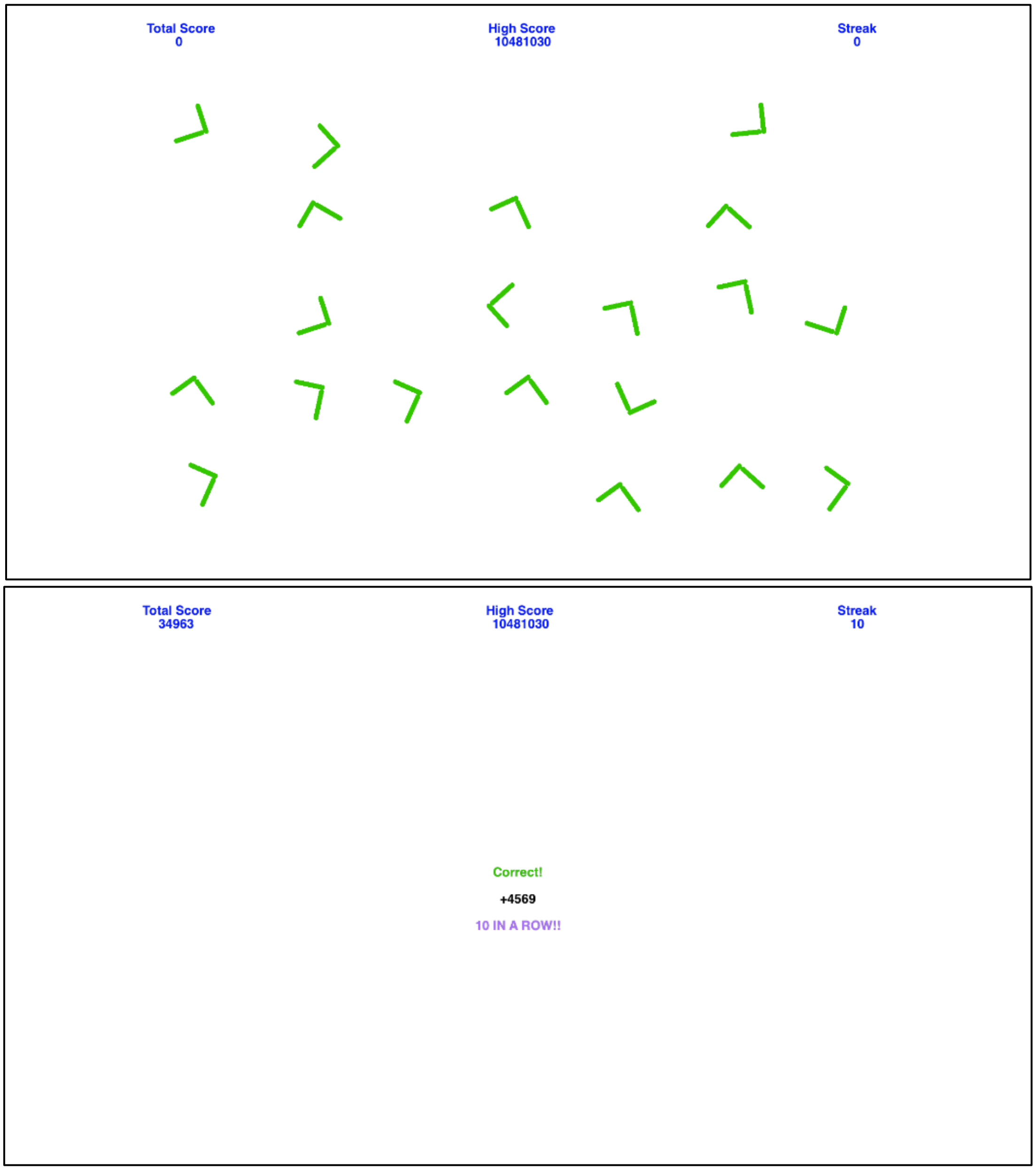


Figure. 7. The gamified version of the visual search tasks for a T among Ls. Participants in both conditions saw the same search stimuli, but the gamified version included a story, points, high scores, and sound effects. For additional screenshots see Appendix C.

The experimental task was a simple search for a T among Ls with set sizes consisting of 10 or 20 items (Figure 7). Half of the displays contained a target and half did not, and participants' task was to indicate whether the target (T) is present in the display via keypress.

Game elements for the gamified condition included points, a real high score displayed at the top of the screen based on the score of previous participants, sounds corresponding to the type of answer given by the participant (correct, incorrect, streak, high score), and a story describing the task. Points were awarded based on the formula:

$$Points\ per\ trial = (2000 * P_{10}) + (3000 - RT) + (S * 100)$$

Where P_{10} is the proportion correct during the last 10 trials, initially set to 0.5, RT is the response time in ms for the current trial, and S is the number of correct trials in a row. The first part of the formula, including P, proportion correct on the last 10 trials, emphasized accuracy and did not decrease dramatically if one or two trials were missed. The second portion of the formula including RT provided more points the faster the participant responded, as long as the response was correct. The streak component S allowed bonus points to continue to accumulate indefinitely, but decreased dramatically once the streak was broken, therefore rewarding accuracy.

Post-tests of motivation, as measured by the Intrinsic Motivation Inventory (IMI), followed the last block of trials (see Appendix A). The IMI measured intrinsic motivation on six dimensions: interest and enjoyment, perceived competence, effort and importance, pressure and tension, perceived choice, and value and usefulness. Both conditions responded to the IMI questionnaire on Qualtrics.

Table 1

Game Elements Used in The Gamified Version Of The Visual Search Task

Game element	Element Description
Points	Points were given after every correct answer. The number of points given was based on participant's proportion correct on the last 10 trials, current trial RT and bonus streak multiplier based on the current number of correct trials in a row.
High score	Previous high score was displayed in the top middle of the screen. This score was based on the highest score achieved by the previous participant sitting at that computer.
Sounds	Sounds were played after correct and incorrect responses as well as when a new high score was achieved, and after a streak of 10 correct responses in a row.
Story	A short description of the "game" was displayed before the gamified task (Appendix B).
Feedback	The participant was informed whether their answer was correct or incorrect by displaying "CORRECT" or "INCORRECT" on the screen. Feedback in the form of sounds and points was also be given based on the response given by the participant.

Note: Also see Appendix D.

Procedure

Participants were welcomed into the lab and seated in front of a computer. They were asked to read and sign the informed consent document if they agreed to participate in the study. Participants were also informed that they were allowed to leave the study at any point without penalty. Next, each participant read the task instructions presented on the computer screen. After the participants familiarized themselves with the instructions, they begun the practice phase (1-2 min) during which they learned the visual search task. After the practice phase, the experimental phase lasted for 30-40 minutes. At the end of the experimental phase the participant responded to a brief survey. Lastly, the participants were debriefed by the experimenter.

The task included three blocks, with each non-practice block consisting of 300 trials. Each trial lasted approximately 3 seconds and the time between each trial was 1.5 seconds. The first block of 40 trials was a practice round with immediate feedback for correct and incorrect trials. In the gamified condition, the practice block also contained information on how the points, bonus points, and high scores work. It also introduced the story (see Appendix B). The following blocks differed based on condition. The gamified condition received feedback in terms of game mechanics (points, streaks, high scores, and sound effects), the control condition only received feedback through "CORRECT" or "INCORRECT" text displayed in the middle of the screen, depending on their response. For a list of game elements used in previous studies see
Appendix D.

One questionnaire followed the visual search task. The questionnaire was conducted using Qualtrics and measured participant motivation and engagement during the experimental task. The questionnaire was a version of the Intrinsic Motivation Inventory (see Appendix A for sample of this questionnaire) with some demographic questions including a measure of prior game experience (Appendix E).

Dependent Measures and Data Cleaning

Visual search performance was operationalized as the average response time on correct trials and the error rate for each condition. In typical visual search studies, RT data are cleaned and screened to exclude extremely short or long RTs (e.g., Miranda & Palmer, 2014). Participants were allowed to respond quickly without being told to slow down to collect information on possible data cleaning effects (measured by extremely quick or slow RTs and number of missed or incorrect trials). Some studies that use a visual search task have a 5 second delay penalty and a message that discourages fast responses (e.g. Miranda & Palmer, 2014). In this study, we eliminated the delay, thus allowing participants to respond quickly without penalty, if they chose to do so. We predicted that participants in the non-gamified condition will be more likely to “blow off the task” in this manner than participants in the gamified condition, which would then be reflected in our data cleaning and screening analysis.

Data which would normally end up being removed from statistical analysis during the clean-up phase was instead quantified to account for the data clean-up effect. This included identification of participants with extremely short or long response times (number of trials < 200 ms or > 3,000 ms) and performance on the visual search task equivalent to or lower than three standard deviations below the mean (high error rate).

Results

Visual Search Performance

Each hypothesis was evaluated using an appropriate statistical test (Table 2). A 2 (Target Presence [Present, Absent]) x 2 (Set Size [10, 20]) x 2 (Gamification [Yes, No]) mixed analysis of variance (ANOVA) was conducted to asses hypotheses one (set size), two (target presence), and five (gamification). The repeated measure had only two levels therefore the assumption of sphericity was met.

Table 2

Summary of The Statistical Analysis Plan

H	Hypothesis	IVs	DVs	Statistical Test	Effect Size
1	Set size is related to performance such that participants will have better performance in the lower set size condition.	Set size (20, 10)	Response time Proportion correct	ANOVA	η_p^2
2	Target presence will affect participant performance	Target Presence (TP, TA)	Response Time Proportion Correct	ANOVA	η_p^2
3	Gamification is related to motivation such that participants will report higher levels of motivation in the gamified condition.	Gamification (Yes, No)	Motivation (IMI score)	*t*-test	*Cohen's d*
4	Motivation is related to performance, such that participants that report higher levels of motivation on the Intrinsic Motivation Inventory will perform better in the visual search task.	Motivation (IMI score)	Response time Proportion correct	Pearson correlation	r^2
5	Gamification is related to performance such that participants will have better performance in the gamified condition.	Gamification (Yes, No)	Response time Percent correct	ANOVA	η_p^2
6	The gamified version of the task will yield fewer trials that would be excluded using normal data cleaning and screening procedures.	Gamification (Yes, No)	Number of trials that would have been excluded	*t*-test	*Cohen's d*

Visual search response time analysis. As predicted, there was a statistically significant main effect of target presence, $F(1, 82) = 317.43$, $p < .001$, $\eta_p^2 = .80$, meaning that participants spent more time searching when the target was absent than when it was present. Additionally, and also as expected, there was a main effect of set size, $F(1, 82) = 475.96$, $p < .001$, $\eta_p^2 = .85$, indicating that participants spent more time searching for the target in the larger set size displays of 20 items than in the smaller set size displays of 10 items. There was a significant interaction of target presence by set size, $F(1, 82) = 245.99$, $p < .001$, $\eta_p^2 = .75$, indicating that on target absent trials, participants were slower when searching larger set size displays than smaller set size displays. Similarly, on target present trials, participants were slower to respond when searching for the target in the larger set size than in the smaller set size displays. This can be seen in Figure 8, where the slopes of target present and target absent trials are different, with the target present trials having a shallower slope than target absent trials. Additionally, there was a significant three-way interaction of target presence by set size by condition, $F(1, 82) = 4.37$, $p < .05$, $\eta_p^2 = .05$, which, based on the slope of the target absent trials in gamified condition (Figure 8), appears to be driven by participants in the gamified condition responding slower on target absent trials, particularly for the larger set size.

There was no significant main effect for condition $F(1, 82) = .99$, $p = .322$, $\eta^2_p = .001$, meaning that no significant differences were found between the gamified and non-gamified conditions in response times.

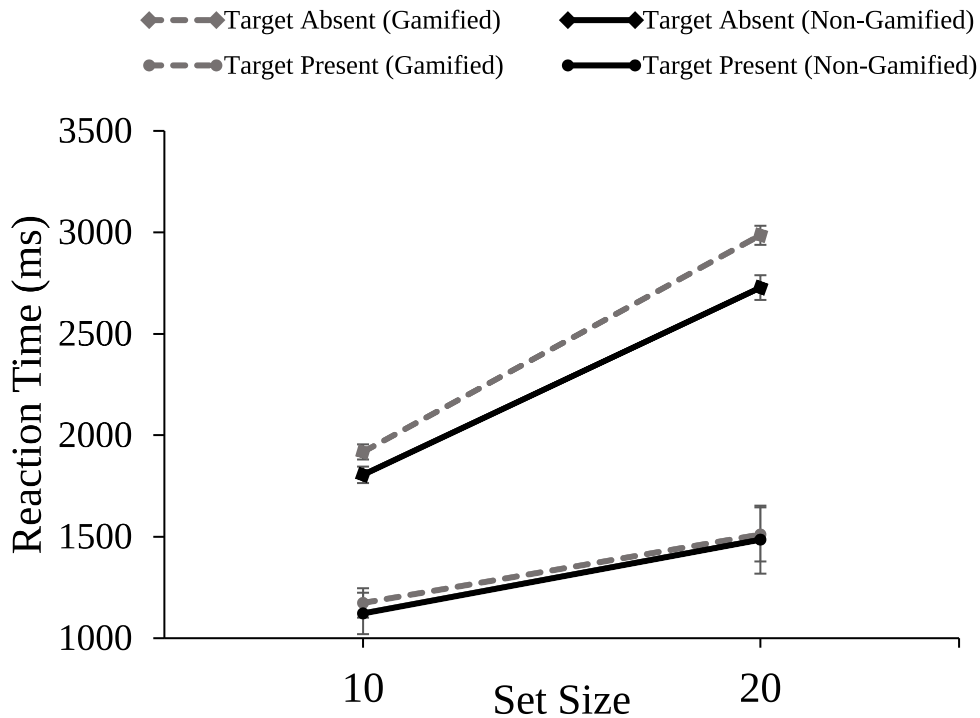


Figure 8. Reaction time for target present and target absent trials in small and large set size displays. Different colors represent different conditions (Grey = Gamified; Black = Non-Gamified). Reaction time was longer on target absent trials and in larger set size displays. Error bars represent standard error of the mean.

Visual search accuracy analysis. A second 2 x 2 x 2 mixed ANOVA was conducted to compare the mean differences between groups for accuracy. There was a statistically significant main effect of target presence, $F(1, 82) = 171.28$, $p < .001$, $\eta_p^2 = .68$, indicating that error rates were lower on target absent trials than on target present trials. There was also a main effect of set size, $F(1, 82) = 42.25$, $p < .001$, $\eta_p^2 = .34$, meaning that participants tended to commit fewer errors on trials with smaller set size displays than on trials with larger set size displays. Additionally, there was a significant interaction of target presence by set size, $F(1, 82) = 40.38$, $p < .001$, $\eta_p^2 = .33$, reflecting that set size had a larger impact on target absent than target present trials. This can be seen in the differences between the slopes for target present and target absent trials, with target absent trials having a much shallower slope than target present trials (see Figure 9). Lastly, there was a significant interaction of target presence by set size by condition, *F*(1, 82) = 4.40, $p < .05$, $\eta_p^2 = .05$. As seen in Figure 9, participants in the non-gamified condition had a higher error rate on target present trials in both small and large set size displays as compared to the error rates in target present trials in non-gamified condition. The error rates for target absent trials were similar for both conditions and set sizes.

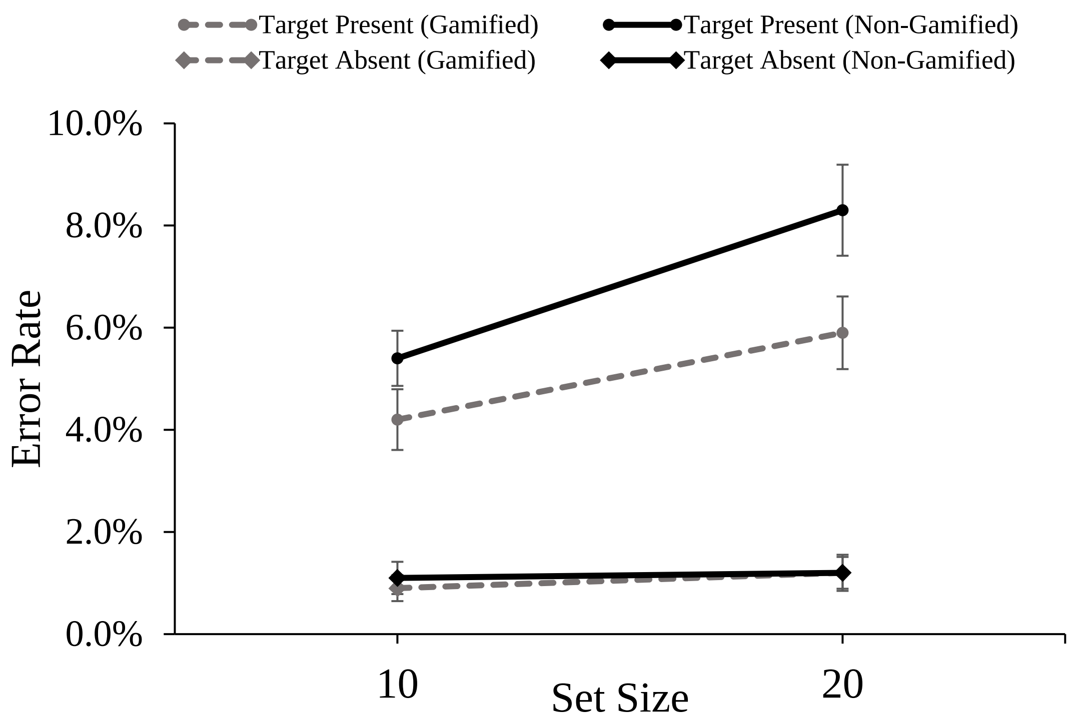


Figure 9. Error rate for target present and target absent trials in small and large set size displays. Different colors represent different conditions (Grey = Gamified; Black = Non-Gamified). Error rate was lower for trials with target absent and trials with smaller set size displays. Error bars represent standard error of the mean.

Intrinsic Motivation Analysis

t-tests were used to assess the third and sixth hypotheses. First, a *t*-test was conducted to examine whether mean differences exist between the motivation of participants in the gamified condition versus the non-gamified condition.

Each IMI subscale score was calculated by averaging the item scores for the items on each subscale. The total motivation score was calculated by combining each subscale score and averaging it by condition (see Figure 10). An independent-samples *t*-test was conducted to compare the motivation of participants in the gamified condition and the non-gamified condition. No significant differences were found between the gamified (M = 27.00, SD = 3.97) and non-gamified (M = 27.26, SD = 4.96) conditions in terms of motivation, t(82) = - 0.26, p = .792, d_{diff} = - 0.06.

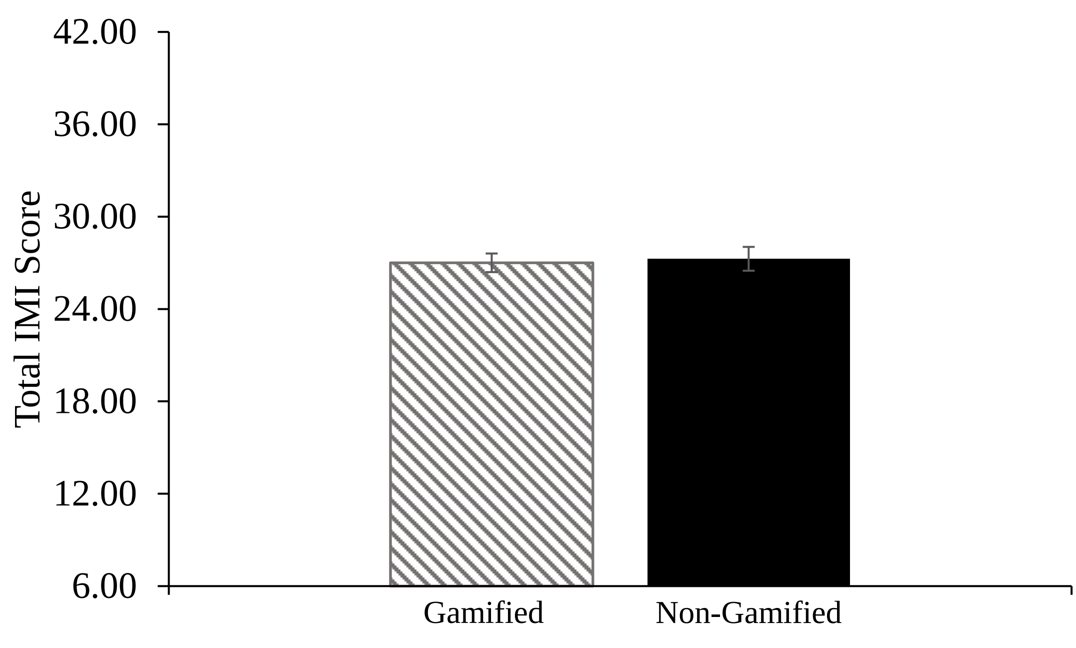


Figure 10. Average Intrinsic Motivation Inventory scores for each condition. The two conditions did not significantly differ in intrinsic motivation. Highest score possible was 42, while lowest possible score was 6. Error bars represent standard error of the mean.

Motivation and Performance

Bivariate (Pearson) correlations were used to assess the relationship in the fourth hypothesis between motivation and performance (see Table 3). One correlation test was performed on the IMI scores and participant reaction times, and another correlation was performed on the IMI scores and average proportion correct responses. Motivation significantly correlated with accuracy, $r(82) = .245$, $p < .05$, meaning that more motivated participants were more accurate and committed fewer errors on the visual search task. Motivation did not significantly correlate with response time, the other measure of performance.

Table 3

Bivariate (Pearson) Correlations for Motivation and Performance

	Accuracy	Response Time	Total Score	Total IMI
Accuracy	—			
Response Time	0.405***	—		
Total Score	0.594***	0.398***	—	
Total IMI	0.245*	-0.043	0.128	—

* p < .05, ** p < .01, *** p < .001

Data Cleaning & Screening Effects

The sixth hypothesis was assessed using a *t*-test between the number of trials that would be excluded from each participant in the gamified versus the non-gamified condition. A significant Levene's test ($p < .05$) indicated a violation of the equal variance assumption, therefore, Welch’s approximation was used to compare the number of participants to be excluded based on participant accuracy and reaction time (two standard deviations below the mean) in the gamified and non-gamified conditions. A total of four participants were to be excluded from each condition. In the gamified condition, three participants would be excluded based on accuracy and one participant based on reaction time. In the non-gamified condition, one participant would be excluded based on accuracy and three based on reaction time. No significant differences were found between the gamified and non-gamified conditions in terms of participants to be excluded based on accuracy, t (69.71) = 0.981, p = .330, d_{diff} = .213. Similarly, no significant differences were found between the gamified and non-gamified conditions in terms of participants to be excluded based on reaction time, t (63.44) = -1.056, p = .295, d_{diff} = -.232. In terms of trials, in the gamified condition 400 trials would be excluded overall and 336 trials would be excluded in the non-gamified condition. No significant differences were found between the two conditions in terms of overall number of trials to be excluded, t (81.81) = 0.263, p = .793, d_{diff} = .057 (see Table 4).

Table 4

Independent Samples t-test for Data Cleaning and Screening Effects

	n	*M*	*SD*	*t*	df	*p*	Cohen's d
Excluded Participants by Accuracy							
Gamified	43	0.07	0.258	0.981	69.714	0.330	0.213
Non-Gamified	41	0.024	0.156				
Excluded Participants by Reaction Time							
Gamified	43	0.023	0.152	-1.056	63.442	0.295	-0.232
Non-Gamified	41	0.073	0.264				
Excluded Trials Overall							
Gamified	43	9.302	19.30	0.263	81.811	0.793	0.057
Non-Gamified	41	8.195	19.30				

Note. Welch's t-test.

Exploratory Analyses

Additional exploratory analyses were conducted to further understand the relationships between variables regardless of condition (see Table 5). In particular, Bivariate (Pearson) correlations were used to assess the relationships between performance (accuracy, response time, streak score), previous game experience (GEM), and IMI Subscales (choice, pressure, competence, enjoyment, effort, and value). See Table 5 for correlations data. Total in-game score and streak length were collected in the background during the experiment in both conditions to make these analyses possible. There was a significant positive correlation between accuracy and response time, $r(82) = .405$, $p < .001$, indicating that as participants spent more time on each trial their accuracy increased. Response time also significantly correlated with total score, $r(82) = .398$, $p < .001$, meaning that when participants spend more time on a trial they were more likely to achieve a higher score in game. Unsurprisingly, total score also positively correlated with accuracy, $r(82) = .594$, $p < .001$, indicating that as participants got more accurate, their score increased. Participants' longest streak length positively correlated with accuracy as well, $r(82) = .623$, $p < .001$, confirming that streak length is linked with accuracy. Streak length was also positively correlated with response time, $r(82) = .482$, $p < .001$, indicating that the longer participants spent on a trial the more likely they were to achieve a longer streak length. Unsurprisingly, streak length was also significantly positively correlated with total score, $r(82) = .955$, $p < .001$, demonstrating that a longer streak length resulted in a higher total score. Previous game experience was significantly negatively correlated with response time, $r(82) = .224$, $p < .05$, indicating that participants who had more previous game experience tended to respond faster. Additionally, previous game experience positively correlated with the IMI value subscale, $r(82) = .233$, $p < .05$, suggesting that participants who had more previous game experience found the visual search task more valuable.

The competence subscale of the Intrinsic Motivation Inventory was found to significantly correlate with three performance measures: accuracy, $r(82) = .286$, $p < .01$, streak length, $r(82) = .319$, $p < .01$, and total score, $r(82) = .340$, $p < .01$. This suggests that participants who did well on the visual search task as measured by accuracy, streak length, and total score, also self-reported that they did well on the task. The feeling of competence also significantly correlated with effort, $r(82) = .248$, $p < .05$, meaning that participants who felt more competent at the task also felt that they put in more effort into the task. Competence also significantly correlated with enjoyment, $r(82) = .266$, $p < .05$, such that when self-reported competence increased so did self-reported enjoyment. Choice subscale was positively correlated with enjoyment, $r(82) = .348$, $p < .01$, effort, $r(82) = .234$, $p < .05$, and value subscales, $r(82) = .219$, $p < .05$, indicating that participants who felt they had more choice also felt more enjoyment from the task, put in more effort into the task, and found the task to be valuable. Enjoyment also correlated with effort, $r(82) = .466$, $p < .001$, and value, $r(82) = .647$, $p < .001$, suggesting that when participants enjoy themselves more they tend to put more effort into the task and find the task more valuable or participants who put more effort into the task tend to enjoy it more. Effort also positively correlated with the value subscale, $r(82) = .435$, $p < .001$, indicating that when participants felt they put in a lot of effort into the task they also felt the task was more valuable or participants who felt the task was valuable tended to put more effort into the task.

Table 5

Bivariate (Pearson) Correlations of Additional Performance Measures (Streak, Points), Previous Game Experience (GEM), and Subscales of the IMI.

	Performance				IMI					
	Accuracy	Response Time	Streak Length	Total Score	Choice	Pressure	Competence	Enjoyment	Effort	Value
Performance										
Accuracy	—									
Response Time	0.405***	—								
Streak Length	0.623***	0.482***	—							
Total Score	0.594***	0.398***	0.955***	—						
IMI										
Choice	0.113	-0.073	0.078	0.047	—					
Pressure	-0.068	-0.090	0.060	0.052	-0.039	—				
Competence	0.286**	0.053	0.319**	0.340**	0.040	-0.093	—			
Enjoyment	0.168	-0.100	0.107	0.084	0.348**	0.001	0.266*	—		
Effort	0.118	-0.044	0.086	0.027	0.234*	0.189	0.248*	0.466***	—	
Value	0.116	-0.071	0.045	0.049	0.219*	0.069	0.207	0.647***	0.435***	—
Previous Game Experience										
GEM	-0.126	-0.244*	-0.141	-0.110	-0.037	-0.066	0.158	0.154	0.079	0.233*

* p < .05, ** p < .01, *** p < .001

Discussion

Main Hypotheses

The purpose of this experiment was to examine the gamification paradox, that is, to determine whether gamification can positively influence intrinsic motivation and in turn participant performance. The first set of hypotheses (hypotheses one & two) evaluated the implementation of the visual search task. As predicted in the first hypothesis, participants had better visual search performance on the lower set size displays in both response time and proportion correct. Participants tended to respond faster and commit fewer errors on smaller set size displays. This indicated that the visual search task implementation was successful and aligned with previous studies (Wolfe, 1994; 1998). Target presence also had an effect on participant performance in terms of response time and proportion correct. As predicted by the second hypothesis, participants tended to look longer for the target and commit fewer errors on target absent trials. This was consistent with previous literature, which showed that target presence affects participant reaction time and accuracy (Chun & Wolfe, 1996; Treisman & Gelade, 1980).

The second set of hypotheses (hypotheses three through six) focused on the main purpose of the study, specifically whether adding gamification elements to a task would affect participant motivation and engagement and whether the increased motivation would result in increased performance. The third hypothesis stated that gamification would be related to motivation such that participants would report higher levels of motivation in the gamified condition. There was no difference between the gamified and non-gamified conditions in motivation scores as reported on the IMI. Participants in both conditions were moderately motivated (M_G = 27.00, M_N = 27.26), which means the gamification implemented did not motivate participants in the gamified condition any more than the task itself in the non-gamified condition. One explanation for this could be that the task was not difficult enough to see the difference. The main task consisted of a simple search for a T among Ls in relatively small set sizes (10 & 20). A more difficult task could result in frustration and boredom in the non-gamified condition, and a welcome challenge in the gamified condition.

Our study found that motivation was related to performance in terms of accuracy, meaning that more motivated individuals had a lower error rate and therefore were more accurate on the visual search task. This suggests that overall intrinsic motivation affects performance in terms of accuracy. Response time was not correlated with motivation, therefore we found only partial support for the fourth hypothesis. It seems that our implementation of gamification was too weak to motivate participants. The task we chose may not have been difficult enough to accentuate the differences between the conditions. Even so, the fact that motivation did correlate with a performance measure provides support for the link between motivation and performance.

The fifth hypothesis focused on the relationship between gamification and participant performance. Ideally, participants would try to maximize their performance on both accuracy and speed. However, the longer participants spend on a trial the more visual information they will collect, and thus their error rate will decrease. The participants then face a choice whether to respond slowly and become more accurate, or to respond quickly but make more errors. This speed-accuracy trade-off is often seen in participant performance on visual search tasks. In this study we found evidence of a speed-accuracy trade-off, such that participants chose to respond slower to increase their accuracy. While no between subject effect was found for response time or proportion correct, a significant interaction of target presence by set size by condition was detected for both response time and proportion correct. Participants in the gamified condition responded slower on target absent trials than participants in the non-gamified condition, particularly on large set size displays. Additionally, participants in the gamified condition had lower error rates (higher accuracy) on target present trials than participants in the non-gamified condition. This appears to be driven by the point awarding formula we used in the experiment. The formula emphasized accuracy over speed by rewarding unbroken streaks of correct answers and awarding points based on the proportion of correct responses on previous 10 trials. Although the formula also awarded points for shorter response times, accuracy was emphasized more, resulting in participants adjusting their strategy to earn more points in the gamified condition. These results suggest that there is a behavioral effect of gamification elements that reward participants.

The sixth hypothesis focused on data cleaning effects, stating that the gamified version of the task would yield fewer trials that would be excluded using normal data cleaning and screening procedures. No differences were found between the two conditions in the number of participants that would have been excluded from analyses. Additionally, the same tests were run with the trials excluded, yielding similar results, therefore the trials were kept in the analyses. A possible explanation for the lack of support for this hypothesis lies in the difficulty of the visual search task. The task may have been too easy for participants as evidenced by low error rates and high proportion correct. If the task were to be more difficult, we may see more drop-off in accuracy and more participants may stop trying to perform well on the task, especially in the non-gamified condition in which they were not incentivized to do well.

Exploratory Analyses

The purpose of the additional analyses was to explore the relationships between variables of performance (accuracy, response time, streak, and score), previous game experience (GEM), and IMI Subscales (choice, pressure, competence, enjoyment, effort, and value). Bivariate (Pearson) correlations revealed multiple significant relationships among the variables. Unsurprisingly, all performance predictors were highly correlated with each other, which is likely due to the nature of the task as well as our points awarding formula. Three of the four performance measures were also correlated with self-reported competence, which supports the IMI creators' claim that the subscale is a positive predictor of behavioral measures of intrinsic motivation (Deci & Ryan, 1985).

The value subscale was found to correlate significantly with previous game experience, which suggests that participants with more game experience found the task to be more valuable to them. This could be because the visual search task in general resembled a game and participants who have more game experience find more value in game-like tasks than participants with no prior game experience.

The remaining correlations found within the IMI subscales are to be expected within a measure of intrinsic motivation and therefore will not be analyzed in detail.

Why did performance not improve?

This study aimed to address the reasons why gamification did not previously improve performance, while improving motivation. The first explanation was rooted in data cleaning and screening effects. This study found that data cleaning and screening did not affect the results of the study enough to explain the lack of difference in performance between the two conditions.

The second explanation focused on the type of gamification elements used in previous studies. Specifically, Hawkins et al. (2013) chose to use game elements that were not tied to performance and therefore did not reward higher performance behavior. This study addressed this problem by incorporating points, streaks, and high scores, into the reward mechanism of the gamification. Although, this study was unsuccessful in increasing intrinsic motivation of participants in the gamified condition, it was able to affect their behavior through the performance-based rewards.

The last explanation stated that gamification just does not improve performance. Although we did not find differences between conditions in terms of performance or motivation, we did find that motivation correlated with participant accuracy but not response time. Given that our point formula emphasized accuracy, we could be seeing an effect of gamification on participant performance behavior. This suggests that gamification does indeed affect performance, however the effect is difficult to detect, and our implementation was not able to fully capture it.

Implications

While we did see a minor behavioral effect of gamification on performance, we did not see increased participant motivation in the gamified condition. The lack of a motivational benefit of our implementation suggests that adding gamification elements to a simple task may not be very useful if the goal is improved motivation. Researchers, educators, and industry professionals alike should consider that improving intrinsic motivation via gamification may not be as simple as adding rewarding game elements to a task. Professionals who wish to use gamification in their product should consider the desired purpose of the gamification as well as how it fits into the product. Careful research into each gamification element should be done before blindly adding it into the design of the product.

The scope of this study was limited to examining the overall effect of gamification on performance and intrinsic motivation. Further research into each gamification element should be done in the future to establish a database of their possible benefits and drawbacks. This study demonstrated that gamification can affect behavior in a controlled task, without necessarily affecting participant intrinsic motivation, therefore contributing to the overall knowledge about gamification.

Limitations

Our study found some effects of gamification elements on participant behavior, however the differences between conditions were non-significant. It is possible that the visual search task required too little effort for participants to lose interest in it in the course of an hour. When the task is too easy, effects of motivation may not be as visible as when the task is more difficult. Intrinsic motivation might affect participant performance when the task at hand is more challenging and requires continued effort. Additionally, we might be seeing a novelty effect of the experimental context, which naturally motivated participants in the non-gamified condition. For example, if the task were boring and tedious enough on its own, participants in the non-gamified condition might lose interest and start skipping trials, while participants in the gamified condition might continue their engagement with the task thanks to the game elements designed to motivate participant performance.

Another limitation to this study was the design of the point awarding formula. The formula did not equally emphasize response time and accuracy, which resulted in participants in the gamified condition adjusting their response strategy to focus on accuracy. A more balanced point awarding formula could result in a different participant behavior during the visual search task.

While the IMI has been previously validated and is generally considered a good measure of intrinsic motivation, it does rely on self-report. There are a couple limitations of this self-report measure, such as social desirability bias and reference bias. Participants might be inclined to respond in a way that will be viewed favorably by the researcher and thus might choose to rate a statement higher than they would normally. Similarly, different participants may interpret the same statement in different ways with different reference points, which makes comparison of self-reported motivation more difficult to interpret.

Future Directions

As gamification becomes more and more popular in the modern world it is important to study its effects on human behavior. While this study did not establish a clear link between gamification motivating participants and in turn increasing their performance, some effects of the implementation were seen. Specifically, our point awarding formula was able to influence participant behavior during the visual search task. Future research should consider examining the effects of specific game elements one by one. For example, seeing whether these results can be replicated without additional game elements, such as sounds or story, or whether the gamification elements only work when combined.

Additionally, we should examine whether the difficulty of the task had an effect on sustained motivation, which resulted in similar motivation among the two conditions. A future study could use the same visual search task and gamification elements but with raised difficulty of the task. Adding additional distractors in various shapes and colors, as well as increasing the set size of the display could sufficiently increase the difficulty of the task and therefore emphasize the differences between motivation in the two conditions. Increasing the difficulty of the task would also increase the effort required to perform the task well, thus the gamification rewards would aim to reduce the aversiveness towards high effort.

Another possibility would be to use a different task that can be more dramatically improved upon and one that requires more sustained effort. The task chosen for this study was a simple visual search task, in which performance cannot be improved passed a certain point. That is, even highly motivated individuals will not be able to spot the target faster than they are perceptually capable. Seeing that the error rates were very small in both conditions, a different task, that requires more skill and effort might result in more obvious differences between conditions.

References

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers and Education*, *83*, 57–63. doi: 10.1016/j.compedu.2014.12.012

Banfield, J., & Wilkerson, B. (2014). Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. *Contemporary Issues in Education Research (Online), 7*(4), 291. doi: 10.19030/cier.v7i4.8843

Brewer, R., Anthony, L., Brown, Q., Irwin, G., Nias, J., & Tate, B. (2013). Using gamification to motivate children to complete empirical studies in lab environments. *Proceedings of the 12th International Conference on Interaction Design and Children, (July 2016),* 388–391. doi: 10.1145/2485760.2485816

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis, 64(3), 363–423. doi: 10.3102/00346543064003363

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4). doi: 10.1037/a0035661

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*(1), 39–78. doi: 10.1006/cogp.1996.0002

Davis, V. (2014, March 20). Gamification in education. Retrieved September 20, 2017, from https://www.edutopia.org/blog/gamification-in-education-vicki-davis

Deci, E. L. & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York, Plenum. doi: 10.1007/978-1-4899-2271-7

Deterding S. (2012). Gamification: Designing for motivation. *Interactions*, *19*(4), 14-17. doi: 10.1145/2212877.2212883

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal Of Educational Technology & Society, 18*(3), 75-88.

Eisenberger, R. (1996). Detrimental effects of reward. *American Psychologist, 50*(11), 1153-1166. doi: 10.1037/0003-066X.51.11.1153

Ferro, L. S., Walz, S. P., & Greuter, S. (2013). Towards personalised, gamified systems. *Proceedings of The 9th Australasian Conference on Interactive Entertainment Matters of Life and Death - IE '13, (September),* 1–6. doi: 10.1145/2513002.2513024

Fitz-Walter, Z. J. (2016). Examples. Retrieved September 20, 2017, from http://www.gamificationgeek.com/resources/examples

Fotaris, P., Mastoras, T., Leinfellner, R., & Rosunally, Y. (2016). Climbing up the leaderboard: An empirical study of applying gamification techniques to a computer programming class. *Electronic Journal of E-Learning*, *14*(2).

Frost & Sullivan. (2017). Adoption of integrated gamification technologies boosts growth in automotive ecosystem. Retrieved September 20, 2017, from http://www.prnewswire.com/news-releases/adoption-of-integrated-gamification-technologies-boosts-growth-in-automotive-ecosystem-300513816.html

Fullerton, T. (2008). Game design workshop: A playcentric approach to creating innovative games. *Technology*. doi: 10.1007/s13398-014-0173-7.2

Gibson, D. K. (2015). Fuel economy: The video game. Retrieved March 1, 2018, from http://www.bbc.com/autos/story/20151028-the-gamification-of-fuel-economy

Gordon, K. (2017). Fitbit – statistics & facts. Retrieved October 02, 2017, from https://www.statista.com/topics/2595/fitbit/

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. In Proceedings of the Annual Hawaii International Conference on System Sciences. doi: 10.1109/HICSS.2014.377

Hawkins, G. E., Rae, B., Nesbitt, K. V, & Brown, S. D. (2013). Gamelike features might not improve data. *Behavior Research Methods*, 45(2), 301–18. https://doi.org/10.3758/s13428-012-0264-3

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, 31(1), 6–41. doi: 10.1016/S0165-0173(99)00023-5

Jordan, J. (2017). The slow, steady progress of Zombies, Run! | Pocket Gamer.biz | PGbiz. Retrieved February 8, 2018, from http://www.pocketgamer.biz/interview/66669/the-slow-steady-progress-of-zombies-run/

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. JMIR Serious Games, 4(2), e11. doi: 10.2196/games.5888

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. Psychological Methods, 17(3), 437–455. doi: 10.1037/a0028085

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525–534. doi: 10.1016/j.chb.2015.08.048

Mesko, B. (2017, July 21). The top 15 examples of gamification in healthcare. Retrieved September 20, 2017, from http://medicalfuturist.com/top-examples-of-gamification-in-healthcare/

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, 60, 48-58. doi: 10.1080/02701367.1989.10607413

Miranda, A. T., & Palmer, E. M. (2014). Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behavior Research Methods*, 46(1), 159–72. doi: 10.3758/s13428-013-0357-7

Ong, D., Yeng, C. Y., Hong, C. W., & Young, K. T. (2013). Motivation of learning: An assessment of the practicality and effectiveness of gamification within a tertiary education system in Malaysia. *World Academy of Researchers, Educators, and Scholars in Business, Social Sciences, Humanities and Education (In Association with the Academy of World Finance, Banking, Management and IT) Conference Proceedings*, 1(1), 131–146.

P&S Market Research. (2017). Gamification market to touch $22.9 billion by 2022: P&S Market Research. Retrieved February 2, 2018, from https://globenewswire.com/news-release/2017/08/03/1071818/0/en/Gamification-Market-to-Touch-22-9-Billion-by-2022-P-S-Market-Research.html

Palmer, E. M., Clausner, T. C., & Kellman, P. J. (2008). Enhancing air traffic displays via perceptual cues. *ACM Transactions on Applied Perception*, 5(1), 4. doi: 10.1145/1279640.1279644

Paras, B., & Bizzocchi, J. (2005). Game, motivation, and effective learning: An integrated model for educational game design. In DiGRA 2005: Changing Views: Worlds in Play, 2005 International Conference.

Pavlas, D., Jentsch, F., Salas, E., Fiore, S. M., & Sims, V. (2012). The Play Experience Scale. *Human Factors*, 54(2), 214–225. doi: 10.1177/0018720811434513

Peng, W., Lin, J.-H., Pfeiffer, K. A., & Winn, B. (2012). Need satisfaction supportive game features as motivational determinants: An experimental study of a Self-Determination Theory guided exergame. *Media Psychology*, 15(2), 175–196. doi: 10.1080/15213269.2012.673850

Pierce, W. D. (2004). Rewards, Task Difficulty, and Intrinsic Motivation: A Test of Learned Industriousness Theory. *The Alberta Journal of Educational Research* 50(3), 317–320.

Pierce, W. D., Cameron, J., Banko, K. M., & So, S. (2003). Positive effects of rewards and performance standards on intrinsic motivation. *The Psychological Record*, 53, 561–579. doi: 10.1007/BF03395453

Phan M. H., Keebler J. R., & Chaparro B. S., 2016. The development and validation of the Game User Experience Satisfaction Scale (GUESS). *Human Factors*. doi: 10.1177/0018720816669646

Porter, D. B. (1995). Computer games: Paradigms of o*pportunity*. *Behavior Research Methods, Instruments & Computers*, 27(2), 229–234. doi: 10.3758/BF03204737

Reiners, T., & Wood, L. C. (2015). Gamification in education and business. *Gamification in Education and Business*, 1–710. doi: 10.1007/978-3-319-10208-5

Robinson, S., Sandstrom, S. M., Denenberg, V. H., & Palmiter, R. D. (2005). Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. *Behavioral Neuroscience*. Howard Hughes Medical Institute, University of Washington, American Psychological Association. doi: 10.1037/0735-7044.119.1.5

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450. doi: 10.1037/0022-3514.43.3.450

Ryan, R. M., & Deci, E. L. (2017). Self-Determination Theory: Basic psychological needs in motivation, development, and wellness. Retrieved from https://ebookcentral-proquest-com.libaccess.sjlibrary.org

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), 68. doi: 10.1037/0003-066X.55.1.68

Skinner, B. F. (1938). The Behavior of organisms: An experimental analysis. New York: Appleton-Century.

Taylor, G., Singer, M. J., & Jerome, C. J. (2009, October). Development and evaluation of the game-based performance assessment battery (GamePAB) and game experience measure (GEM). In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 53, No. 27, pp. 2014-2018). *SAGE Publications*.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Volkswagen. (2009). The Fun Theory. Retrieved February 9, 2018, from http://www.thefuntheory.com/

Washburn, D. A. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*., 35(2), 185–93. doi: 10.3758/BF03202541

Wimmer, G. E., Braun, E. K., Daw, N. D., & Shohamy, D. (2014). Episodic Memory Encoding Interferes with Reward Learning and Decreases Striatal Prediction Errors. *The Journal of Neuroscience*, 34(45), 14901 LP-14912. Retrieved from http://www.jneurosci.org/content/34/45/14901.abstract

Wojdyla, B. (2008). Ford SmartGauge LCD Instrument Panel Brings Futuristic Look, Green Leaves To 2010 Hybrids. Retrieved March 1, 2018, from https://jalopnik.com/5070371/ford-smartgauge-lcd-instrument-panel-brings-futuristic-look-green-leaves-to-2010-hybrids

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238. doi: 10.3758/BF03200774

Zichermann, G., & Cunningham, C. (2011). Gamification by design: implementing game machanics in web and mobile apps. Cambridge, MA: Oreilly & Assoc. Inc.

Appendix A

Intrinsic Motivation Inventory (IMI)

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*(3), 450.

1	2	3	4	5	6	7
Not True at all			somew hat true			very true

1. I didn’t really have a choice about doing this task.
2. I did this activity because I wanted to.
3. I think doing this activity could help me to _____.
4. I did not feel nervous at all while doing this.
5. I was very relaxed in doing these.
6. I am satisfied with my performance at this task.
7. I enjoyed doing this activity very much.
8. I think I am pretty good at this activity.
9. I tried very hard on this activity.
10. I thought this activity was quite enjoyable.
11. I think this is important to do because it can _____.
12. This was an activity that I couldn’t do very well.
13. While I was doing this activity, I was thinking about how much I enjoyed it.
14. I was anxious while working on this task.
15. I was pretty skilled at this activity.
16. This activity did not hold my attention at all.
17. I felt very tense while doing this activity.
18. I felt like it was not my own choice to do this task.
19. I felt like I had to do this.
20. I believe doing this activity could be beneficial to me.
21. I believe this activity could be of some value to me.
22. This activity was fun to do.
23. I think this is an important activity.
24. I would be willing to do this again because it has some value to me.
25. I felt pressured while doing these.
26. I did this activity because I had to.
27. I would describe this activity as very interesting.
28. It was important to me to do well at this task.
29. I did this activity because I had no choice.
30. I didn’t put much energy into this.
31. I believe I had some choice about doing this activity.

32. After working at this activity for awhile, I felt pretty competent.
33. I put a lot of effort into this.
34. I didn't try very hard to do well at this activity.
35. I think that doing this activity is useful for ______.
36. I think I did pretty well at this activity, compared to other students.
37. I thought this was a boring activity.

<Scoring>

Intrinsic Motivation Dimensions

<u>Interest/ Enjoyment</u>: 7, 10, 13, 16, 22, 27, 37

7. I enjoyed doing this activity very much.
10. I thought this activity was quite enjoyable.
13. While I was doing this activity, I was thinking about how much I enjoyed it.
16. This activity did not hold my attention at all. *
22. This activity was fun to do.
27. I would describe this activity as very interesting.
37. I thought this was a boring activity. *

<u>Perceived Competence</u>: 6, 8, 12, 15, 32, 36

6. I am satisfied with my performance at this task.
8. I think I am pretty good at this activity.
12. This was an activity that I couldn't do very well. *
15. I was pretty skilled at this activity.
32. After working at this activity for awhile, I felt pretty competent.
36. I think I did pretty well at this activity, compared to other students.

<u>Effort/ Importance</u>: 9, 28, 30, 33, 34

9. I tried very hard on this activity.
28. It was important to me to do well at this task.
30. I didn't put much energy into this. *
33. I put a lot of effort into this.
34. I didn't try very hard to do well at this activity. *

<u>Pressure/ Tension</u>: 4, 5, 14, 17, 25

4. I did not feel nervous at all while doing this. *
5. I was very relaxed in doing these. *
14. I was anxious while working on this task.

17. I felt very tense while doing this activity.
25. I felt pressured while doing these.

Perceived Choice: 1, 2, 18, 19, 26, 29, 31

1. I didn't really have a choice about doing this task. *
2. I did this activity because I wanted to.
18. I felt like it was not my own choice to do this task. *
19. I felt like I had to do this. *
26. I did this activity because I had to. *
29. I did this activity because I had no choice. *
31. I believe I had some choice about doing this activity.

Value/ Usefulness: 3, 11, 20, 21, 23, 24, 35

3. I think doing this activity could help me to _____.
11. I think this is important to do because it can _____.
20. I believe doing this activity could be beneficial to me.
21. I believe this activity could be of some value to me.
23. I think this is an important activity.
24. I would be willing to do this again because it has some value to me.
35. I think that doing this activity is useful for _____.

*** Reversed Items**: 1, 4, 5, 12, 16, 18, 19, 26, 29, 30, 34, 37

Appendix B

Story displayed for the gamified condition.

Imagine you are a microbiologist looking at water samples. Your task is to determine

whether a sample is contaminated with harmful T-shaped bacteria

or whether the sample is clean and contains only harmless L-shaped bacteria.

You will be looking at many samples.

Some samples will have the T-shaped bacteria while some will not.

When you find the T-shaped bacteria

press the “Quote” button to indicate “target present.”

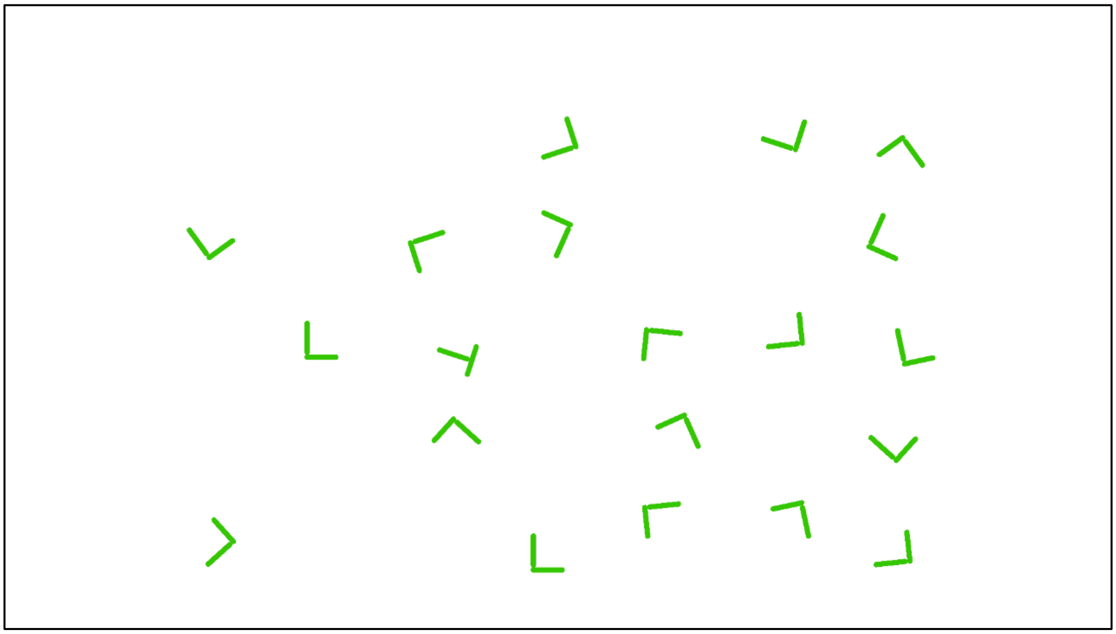
When you determine that the T-shaped bacteria is absent

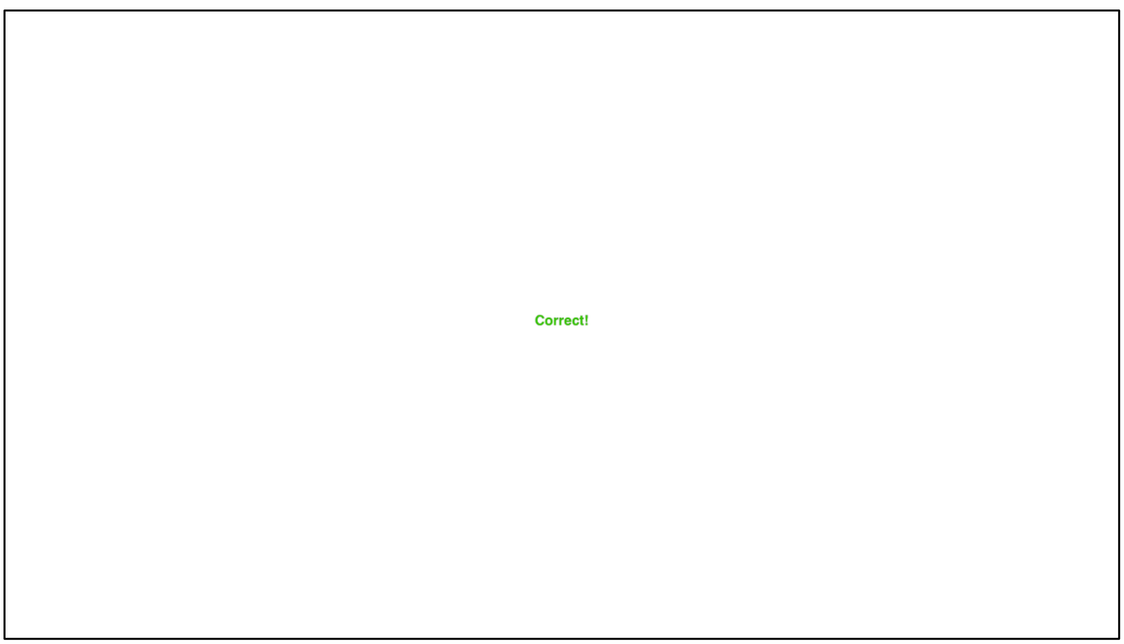
press the “A” button to indicate “target absent.”

After indicating your choice, a new sample will appear.

Appendix C

Additional screenshots of the visual search task (non-gamified).





Appendix D

Game elements previously used in gamification.

Game elements	Description	References
Story	A narrative which unfolds during a game or gamified activity. It may engage player emotionally.	Banfield and Wilkerson (2014) Hawkins et al. (2013) Peng, Lin, Pfeiffer, and Winn (2012) Rodrigues, Oliveira, and Costa (2016) Zombies, Run!
Points	Used to determine the score or achievement in a game or gamified activity. Awarded for performing specific actions.	Brewer et al. (2013) Fotaris et al. (2016) Hawkins et al. (2013) Hanus and Fox (2015) Miranda and Palmer (2014) Porter (1995)
High Scores	Used to compare scores within and between players.	Fotaris et al. (2016) Hanus and Fox (2015)
Sounds	Tones played after achieving a goal or after specific action was performed.	Miranda and Palmer (2014) Zombies, Run!
Feedback	Informs the player of the result of their actions.	Fitbit Miranda and Palmer (2014)

Appendix E

Game Experience Measure (GEM)

Taylor, G., Singer, M. J., & Jerome, C. J. (2009, October). Development and evaluation of the game-based performance assessment battery (GamePAB) and game experience measure (GEM). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 27, pp. 2014-2018). SAGE Publications.

Answer the questions below to characterize your daily experience or habits with video and computer games. For each question, circle the appropriate choice that most accurately describes your experience. Answer questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

1. What is your level of confidence with video games in general?

Very Low Low Average High Very High

2. How many *hours per week* do you currently play video games (average of the past 6 months)?

0-9 10-19 20-29 30-39 40+

3. What is the maximum number of *hours per week* you've ever spent playing video games?

0-9 10-19 20-29 30-39 40+

4. About how many *times* have you read a video game magazine or website to find out tips to improve your gaming skill?

0-9 10-19 20-29 30-39 40+

5. How often do you play the following types of games:

Action (e.g., Street Fighter, Contra)	Never	Rarely	Monthly	Weekly	Daily
Adventure (e.g., Myst, Fable)	Never	Rarely	Monthly	Weekly	Daily
Music (e.g., Guitar Hero, Dance Dance Revolution)	Never	Rarely	Monthly	Weekly	Daily
Platform (e.g., Mario Bros., Sonic the Hedgehog)	Never	Rarely	Monthly	Weekly	Daily
Puzzle (e.g., Minesweeper, Tetris)	Never	Rarely	Monthly	Weekly	Daily

Racing (e.g., Need for Speed, Test Drive)	Never	Rarely	Monthly	Weekly	Daily
Role-Playing (e.g., Final Fantasy, Pokemon)	Never	Rarely	Monthly	Weekly	Daily
Shooter (e.g., Doom, Halo)	Never	Rarely	Monthly	Weekly	Daily
Sports (e.g., Madden Football, FIFA Soccer)	Never	Rarely	Monthly	Weekly	Daily
Strategy (e.g., Command and Conquer, Civilization)	Never	Rarely	Monthly	Weekly	Daily

6. Indicate your experience with the following game controllers:

	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert
	None	Very Little	Average	High	Expert