CROWDSOURCING DATA COLLECTION THROUGH MOBILE GAMIFICATION: LEVERAGING THE FREEMIUM MODEL

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Computer Science

University of Saskatchewan

Saskatoon, CANADA

By

Kristen Dergousoff

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

ABSTRACT

Classic ways of gathering data on human behavior, such as laboratory based user studies, can be time-consuming, costly and are subject to limited participant pools. Crowdsourcing offers a reduction in operating costs and access to a diverse and large participant pool, however issues arise concerning low worker pay and questions about data quality. Gamification provides a motivation to participate, but also requires the development of specialized, research-question specific games that can be costly to produce. We provide another alternative that combines gamification and crowdsourcing in a smartphone-based system that emulates the popular Freemium model of micro-transactions to motivate voluntary participation through in-game rewards, using a robust framework to study multiple unrelated research questions within the same system. We deployed our prototype framework on the Android market and gathered data over a period of 5 weeks. We compared this data to that gathered from a gamified laboratory version and a non-gamified laboratory version, and found that players who use the in-game rewards were motivated to do experimental tasks. The data showed that there was no difference between the groups for performance on a motor task; however, performance on a cognitive task was worse for the crowdsourced Android group. We discuss the possible reasons for this and provide options for improving data collection and performance on tasks.

ACKNOWLEDGMENTS

First and foremost I would like to thank my supervisor Regan Mandryk for all of her support, encouragement, and guidance throughout my degree. She has provided me with countless opportunities and experiences that have helped me grow as an academic, a professional, and a person. I would also like to thank everyone at the HCI Lab for creating such an awesome, welcoming, and enjoyable environment.

Finally, and most importantly, I would like to thank my family and my husband Kevin Munson for their unconditional and unending support (and their incredible patience) in everything I do.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMT	Amazon’s Mechanical Turk
API	Application Programming interface
GWAP	Games With a Purpose
ID	Index of Difficulty
IMI	Intrinsic Motivation Inventory
MT	Movement Time
PENS	Players Experience of Need Satisfaction

CHAPTER 1

1 INTRODUCTION

1.1 MOTIVATION AND PROBLEM

A fundamental, and limiting, step in Human Computer Interaction research is gathering data in order to understand human behaviour. Researchers often perform user studies in laboratory environments which, despite providing researchers with a large amount of control, come with a few tradeoffs. Laboratory based user studies can be costly and time-consuming and are subject to small local participant pools. In academic research, this usually consists of recruiting participants from a participant pool at the university, which confines participation in studies to the "boom-and-bust semester cycle" and limits the generalizability of the study to the demographic of young college students (Mason & Suri, 2012).

One alternative that has been proposed is to crowdsource experiments. Crowdsourcing platforms, such as Amazon's Mechanical Turk (AMT), provide researchers with the opportunity to submit online tasks and experiments that are completed by "workers", providing them with data for a relatively low cost (Buhrmester et al., 2011). Platforms such as AMT have been shown to successfully recreate experimental results (Heer & Bostock, 2010; Komarov et al., 2013) and provide additional benefits such as constant any-time access to a large and diverse participant pool, reduced operational and administrative costs, and quick experimental development (Kraut et al., 2014; Mason & Suri, 2012). However issues have been raised regarding data quality due to the lack of experimental control that laboratory based studies provide and the ethics surrounding the low monetary incentives often provided to motivate participation (Mason & Suri, 2012).

Gamification is an alternate technique proposed by researchers as a way to increase the small number of willing participants within the participant pool. Using game elements within experimental tasks has been shown to motivate large numbers of voluntary participants (von Ahn

& Dabbish, 2004) and to increase the enjoyability of tasks (Flatla et al., 2011). However, it is still unclear whether the results obtained through a gamified approach are as accurate as data gathered in traditional laboratory settings. Also gamified experiments are usually highly customized to the specific research question, requiring costly development of new games for each research problem that is addressed.

1.2 SOLUTION

In this thesis we propose to combine the motivation of gamification with the broad participant base available through crowdsourcing by creating a framework for delivering experimental tasks leveraging the mobile market and the Freemium model, in which players can play a game for free, but are given access to special content, features, or advertisement-free play for a fee. We created a framework in which players of a game gain in-game advantages (i.e., power-ups) for completing experimental tasks useful to researchers (i.e., microexperiments). We leverage the mobile game market to target players looking to simply kill time with short gameplay sessions. The main advantage of our framework is that it separates the system for completing microexperiments from the game with which it is deployed so that multiple experimental tasks can be deployed in a single game, or multiple games can be deployed to increase appeal for players and target multiple demographic groups.

1.3 STEPS IN THE SOLUTION

The first step of our work required us to design a framework allowing us to deliver experimental tasks within a mobile game, but in such a way that it was flexible and robust. In order to meet our requirements for flexibility, our framework needed to be able to handle multiple experimental tasks, be able to easily add, remove, or modify experimental tasks, allow for the game to easily be replaced without affecting the operation of the tasks and vice versa, and allow the tasks to be delivered using any form of Freemium model (in other words, they should not be bound to the method or rewards used within any particular game). To meet these requirements we took a modular approach to designing our system architecture, separating the logic for our framework into three main areas of concern: the game, the individual tasks, and the task manager. In our framework, the game and the individual tasks are isolated from each other and unaware of the other's existence. This decouples the components from each other and allows for either the game or the experimental tasks to be modified or replaced without affecting the operation of the other. The only component in our system that is fully aware of the other components is the task manager, which is in charge of communicating between the game and the tasks, informing the experimental tasks when they are required to run, and informing the game when/if the player deserves an in-game reward (due to the completion of an experimental task).

To investigate our design, we developed a prototype game called Sugar Rush, and replicated two classical experimental tasks that were deployed within a prototype of our framework. Sugar Rush is an action-based vertical platformer, developed to run on Android devices, that utilizes the device accelerometer. In the game, players are in control of a continuously bouncing cupcake, which they guide through platforms and enemies to collect candy and obtain a high score. For our prototype, we chose to emulate the Freemium model of micro-transactions, in which players use in-game currency purchased with real money to purchase in-game advantages. In our prototype, players are provided with the option to purchase useful power-ups with in-game currency earned through the completion of our quick experimental tasks (microexperiments) rather than by paying a fee.

1.4 EVALUATION

In order to evaluate the effectiveness of our technique, we investigated two main research questions

Research Question One: was the quality of the work that the players were doing similar to that received from more traditional approaches?

Research Question Two: Were players were motivated to play our game and complete the experimental tasks for our in-game rewards?

We evaluated our framework under three groups to isolate the factors of gamification and crowdsourcing and examine their effects separately on the quality of data gathered. Our three groups included a crowdsourced game which was deployed on the Android market over a period of 5 weeks, a laboratory version where participants were asked to play the game for a period of time within a controlled setting in our research lab, and a control group conducted within the laboratory where participants were given only the tasks and were not exposed to the game in any way. Data regarding player performance on the tasks was analyzed and compared between all three groups. We also analyzed usage data received from the Android market to inform the degree to which our prototype framework and chosen Freemium model motivated voluntary participation in the tasks. Finally, we deployed a survey with the laboratory participants who played the game to gather their opinions on including research tasks within mobile games.

1.5 CONTRIBUTIONS

We make two main contributions in this thesis. First, we introduce the idea of leveraging the Freemium model in free-to-play mobile games for delivering and motivating participation in research-based tasks. And second, we designed and developed a framework around this concept. We demonstrate the use of this framework with our prototype Android game, Sugar Rush, and two classic psychophysics experiments. The results of our experiment showed the following results:

- There was no difference in performance on the motor task but that performance worsened for the attention-based cognitive task,
- Participants who were exposed to the power-ups used them in about 21% of games,
- Participants indicated they were willing to do the tasks in return for earned in-game bonuses, would prefer the tasks over in-game advertising, and that the inclusion of tasks did not reduce play experience.

To follow-up on the results, we discuss ways to improve participation and performance on experimental tasks delivered using our model.

1.6 THESIS OUTLINE

The remainder of this thesis will present the related work, system design and implementation, methodology, evaluation, and results of our research as well as opportunities for increasing participation and the quality of the data received with our framework.

Chapter 2 reviews previous work in the areas of crowdsourcing and gamification, both in the public and in the academic domains. Background information on the Freemium business model and its use in mobile games is also discussed.

Chapter 3 introduces the structure of our framework as well as the design of our prototype game, Sugar Rush. First we discuss the software architecture of the framework as well as the technical integration of the tasks. Next we describe the design of our mobile game, Sugar Rush, and discuss our reasoning behind our design decisions.

Chapter 4 discusses the methodological approach taken in evaluating our framework. It also presents three versions of the framework, as well as their accompanying surveys, used in the evaluation.

Chapter 5 discusses the evaluation of the framework, outlining details of the evaluation process as well as discussing the approach taken to analyze the data.

Chapter 6 presents the results of our evaluation and analysis in detail.

Chapter 7 discusses the implications of the results on our framework and outlines design recommendations for further investigation.

Chapter 8 provides a summary of the framework and our prototype game and discusses the main findings of our evaluation as well as recommendations for future work.

CHAPTER 2

2 RELATED WORK

This chapter reviews related work done in the areas of crowdsourcing and gamification. More specifically this chapter will be discussing crowdsourcing used in the public domain as well as in research, the use of non-monetary incentives in crowdsourcing and its effect on data quality, the use of games in education and research, and the use of games in crowdsourcing. This chapter also presents background information on the Freemium business model and its application in mobile games.

2.1 CROWDSOURCING

Crowdsourcing refers to the outsourcing of a function or task to a large network of people and has become an increasingly popular practice in commercial industries and academia alike. The practice of crowdsourcing allows companies and researchers to utilize the power of the crowd to not only offset internal costs and efforts, but also to harness the creative and innovative potential that exists within a large and diverse population (Howe, 2008). Below we highlight some uses of crowdsourcing within business and research.

2.1.1 Crowdsourcing in Business

In 2008, Howe outlined multiple usages of crowdsourcing within business consisting of: the application of collective intelligence towards brainstorming or solving a problem, the creation of large bodies of creative works, maintenance and organization of large information stores, and the collective use of crowd resources (Howe, 2008).

Creative competitive events, such as the innovation and renovation efforts put on by Anheuser-Busch[1] or My Starbucks Idea[2] put on by Starbucks, are a form of crowdsourcing categorized by Howe as the *collective intelligence towards brainstorming or solving a problem*. These competitions are used as a way of leveraging the differing perspectives and experiences of the greater community to provide a wider range of ideas than possible with a small group of product researchers.

Almost everyone will be familiar with the famous information hub Wikipedia.org[3], which serves as a prime example of crowdsourcing used for *the creation of large bodies of creative works.* With projects like that of Wikipedia.org the community of users work together and volunteer their time and effort to extending the information available on the website and also to verify the information that is added – a task that would quickly grow very cumbersome for a small, dedicated group.

There are many companies with websites that allow users to provide votes or ratings on products. These features allow the community to identify which products are useful or of good quality, which helps other community members filter or sort through all of the content. This sort of feature is an example of how the crowd can work together to *filter and organize large stores of information.*

In May 1999, SETI initiated the SETI@home project, which utilized a large network of volunteered computers from around the world in order to analyze radio signals detected in space. This project garnered an astounding number of volunteers, having 400,000 people pre-register within the first year of the project being announced (Anderson et al., 2002). This massive world-wide project is an example of *the collective use of crowd resources.*

---

[1] http://www.ab-inbev.com/innovation.html
[2] http://mystarbucksidea.force.com/
[3] http://wikimediafoundation.org/wiki/Our_projects#Wikipedia

2.1.2 Crowdsourcing in Research

Within research, crowdsourcing has been utilized by many domains to assist in various tasks such as solving computationally complex problems (Cooper et al., 2010; Kawrykow et al., 2012), generating large amounts of data (von Ahn & Dabbish, 2004; von Ahn & Dabbish, 2008; Freitas et al., 2010), and performing large-scale experimentation (Komarov et al., 2013).

In 2011, Franklin et al. created CrowdDB, a database management system (DBMS) that integrates human workers with a traditional query-based database to enhance and increase its capabilities. In this project, the authors combined traditional database technologies and tools (such as SQL) with crowdsourcing capabilities (Amazon Mechanical Turk API) into a hybrid DBMS. Developers used CrowdDB as they would a normal DBMS through their custom query language CrowdSQL. CrowdDB would first attempt to find the requested information within the underlying DBMS and if the information was not available it would submit the request as a HIT on AMT to be completed by a crowdsourced worker, as a way to increase the capability and accuracy of the query results.

Vonikakis et al., (2014) utilized crowdsourcing and the Microworkers platform to identify selection and preference patterns for images and generated a model of *Image Appeal*, which could then be used to predict preferences for unseen photos and used by applications that, for example, could automatically parse through and compile images into photo albums. In their study, crowdsourced workers were asked to create photo albums for a particular subject from a given album of images.

2.1.2.1 Citizen Science

Crowdsourcing is a term that encompasses a variety of different activities (Estellés-Arolas & González-Ladrón-de-Guevara, 2012). One prominent form of crowdsourcing used in research is Citizen Science. Citizen Science is a practice in which citizens are employed to actively assist and collaborate in scientific research by gathering data or performing analysis (Rotman et al., 2012). It is a practice that has been highly employed, especially in ecological sciences, where vast amounts of information are required from a disperse set of locations over a long period of time, making most other forms of data collection infeasible (Lee, 2006).

The use of Citizen Science has assisted in numerous research projects, providing data and manpower that otherwise would not have been available.

Kawrykow et al. (2012) utilized the problem-solving skills of citizens to assist in improving genetic sequence alignments through the use of a public internet-based game, Phylo. In their research they attracted over 12,000 participants, which increased the overall accuracy of up to 70% of the alignment problems they investigated (Kawrykow et al., 2012). To complete this same task using traditional analysis would have taken a considerable amount of computational power, which may not have been available to the researchers.

The advancement of mobile technologies and the pervasiveness of smart devices provide great opportunities for Citizen Science. The inclusion of sensors (GPS, Bluetooth, etc.) in smartphone devices has provided a great opportunity for not only data collection but also for gathering contextual information that can be used to inform the collected data (Kim et al., 2011, Hashemian et al., 2012).

2.1.2.2 Crowdsourcing User Studies

In the academic community, crowdsourcing has become increasingly adopted as an attractive alternative to laboratory-based user studies. Laboratory-based user studies involve recruiting and scheduling participants, which can be a time-consuming task. Also, dealing with scheduled participants is always susceptible to participants not showing up, requiring even more effort and time to find and schedule a replacement. The need for local participants also reduces the generalizability of the study (Mason & Suri, 2012) and can put a limit on the amount of participants that can feasibly be included. Laboratory-based user studies often require more effort and investment by the participants as they are asked to take time out of their day to physically be present at the lab during the study. This additional investment requires larger incentives to attract enough interest and motivate people to participate. Increased incentives and the time requirement of a person to handle scheduling and running the studies can lead to high operational expenses. The difficulty of recruitment, large time investment, and monetary costs often force researchers to make tradeoffs limiting the amount of factors and questions that can be investigated in their user studies.

Crowdsourcing alleviates much of these disadvantages, allowing researchers to reach larger and more diverse participant bases at reduced operational costs (Kraut et al., 2004) and streamlining the experimental process by reducing the administrative overhead of user studies (Kraut et al., 2004; Deterding et al., 2013). Recent research has also shown that crowdsourced participants, specifically those from the crowdsourcing platform Amazon's Mechanical Turk, show better focus than undergrads which are often recruited for laboratory based studies (Capaldi, 2015).

2.1.3 Crowdsourcing Platforms

As Crowdsourcing becomes a more popular practice software platforms have emerged to assist in delivering tasks to large and diverse populations. One of the largest and most well-known crowdsourcing platforms is Amazon's Mechanical Turk (AMT). This platform allows "requesters" to post tasks (referred to as HITs) for other users (known as Turkers) to accept and complete for a pre-determined monetary value. The value of each task is determined by the requesters and is often quite small, consisting of anywhere from a few cents to a few dollars depending on the effort of the task). The Turkers who successfully complete the tasks are only paid once the requesters have verified the quality of the Turker's work.

At the time this thesis was compiled, the AMT platform hosted more than 260,000 HITs posted by businesses and/or researchers (Amazon's Mechanical Turk, 2014). AMT has been found to be an attractive alternative to lab-based studies due to its 24/7 access to an extremely large and diverse participant pool (Mason & Suri, 2012; Heer & Bostock, 2010; Kraut et al., 2004), reduced time and monetary cost of experiment administration and execution (Heer & Bostock, 2010; Mason & Suri, 2012), and ease of experiment modification (Mason & Suri, 2012).

AMT has been successfully used in a number of studies, illustrating that crowdsourcing user-based research tasks is a plausible alternative. In 2010, Heer and Bostock showed that AMT could be used to successfully recreate the results of Cleveland and McGill's classical perception study (Cleveland & McGill, 1984) as well as provide new perceptual insights with modified experimental tasks. Komarov et al. (2013) investigated the feasibility of conducting performance-based research via AMT and found that there were no significant differences between their results obtained via laboratory experiments and those obtained from AMT, indicating that AMT can be reliable for conducting performance-based experiments. Shapiro et al. (2013) used AMT to conduct research on psychopathology and psychological health, and highlighted many advantages that crowdsourcing software like AMT can provide for conducting clinical research (Shapiro et al., 2013), such as fast data collection and the ability to complete complex research designs (longitudinal, experimental, intervention research).

Despite these successes, there are still some reservations regarding crowdsourcing (and the use of AMT) which prevent many researchers from adopting this approach (Komarov et al., 2013). The low rate of pay for completing tasks on the AMT platform raises questions of ethics (Mason & Suri, 2012) and the loss of experimental control and low monetary incentives (Downs et al., 2010) raise questions about the quality of data that is collected via crowdsourcing.

2.1.4 Non-Monetary Incentives and Data Quality

A traditional way of motivating participation in user studies is through monetary reimbursement for participants' time. Many crowdsourcing platforms follow this same incentive mechanism and incentivize participation through monetary rewards. Social science researchers have raised questions about the quality of work when externally motivated and, despite studies showing that small monetary incentives do not necessarily lower the quality of results in crowdsourcing (Mao et al., 2013; Harris, 2011; Mason & Watts, 2009), researchers have begun to investigate crowdsourcing platforms that motivate participation through non-monetary incentives.

One example is LabintheWild, developed by Reinecke and Gajos (2015). LabintheWild is an online experimental platform which incentivizes participation through a person's "innate desire"

(Reinecke & Gajos, 2015) to learn about themselves and compare their performances with others. Instead of monetary compensation, LabintheWild provides participants with personalized feedback and reports that they can use to gauge their performance, see how they compare with the performance of others, and share with their friends through social media platforms. In their evaluation, Reinecke et al. found that the LabintheWild platform was able to attract a large and diverse group of participants, and that non-monetary incentives did not compromise the quality of data gathered (Reinecke & Gajos, 2015).

Although not explicitly discussed in their study, LabintheWild uses a participant's intrinsic motivation (Ryan & Deci., 2000) to complete tasks rather than incentives, which are seen as extrinsic motivators. Research on the psychology of motivation has suggested that intrinsic motivation is a stronger pull and that if intrinsically-motivated people are rewarded for their participation (e.g., through payment), their motivation might be decreased (Lepper et al., 1973). One way to intrinsically motivate people to engage in experiments is to make the experience be an activity that people want to do anyways, such as playing video games.

2.2 LEVERAGING THE MOTIVATIONAL DRAW OF VIDEO GAMES

Video games are an extremely popular entertainment medium attracting millions of players to spend hours immersed in gameplay. Psychological research on the exposure to video games has shown them to have a strong motivational effect on their players (Przybylski et al., 2010). Due to this motivational effect, researchers have attempted to utilize games and their motivational appeal to enhance non-game related areas, such as education (Decker & Lawley, 2013; Watson et al., 2013; Muntean, 2011) and health (McCallum, 2012).

2.2.1 Gamification

Researchers have been attempting to identify what exactly it is about video games that hold such motivational value and appeal to players. While some researchers have focused on investigating the motivational appeal of actions and behaviours that video games allow players to act out

(Bartle, 1996; Yee, 2006), others have looked at specific game elements and mechanics such as points (Mekler et al., 2013), badges (Easley & Ghosh, 2013), or achievements (Hamari & Eranti, 2011), to investigate which the players find most enjoyable and which are most effective at motivating player participation. A recent development has seen the use of these motivating game elements being applied to non-game related activities. This use of game elements in non-game related contexts is referred to as *Gamification* (Deterding et al., 2011), and is a practice that has seen quite a bit of popularity in academia and in commercial industries.

Businesses, such as Samsung (with Samsung Nation), have applied the concept of gamification to their company websites in order to attract and retain a large amount of customer traffic. These companies use gamification to provide incentives to visitors to engage in activities that are beneficial to the business such as reviewing products, leaving comments, or even just spending time on the website.

Figure 1. Gamification features from Samsung's *Samsung Nation* [4].

2.2.2 Gamification in Research

Researchers have investigated using gamification to motivate work-based tasks as well. Recent research into the effectiveness of adding game elements to work-based tasks shows that gamification can motivate participation (Cechanowicz et al., 2013), enhance performance, and increase participants' level of enjoyment (Flatla et al., 2011). With these advantages gamification has been used in a variety of projects intended to facilitate research.

Cechanowicz et al. used gamification to motivate consumer participation in a market research survey. In their study they created two gamified online market research surveys (one partially-

[4] Image retrieved from https://jaymanalotoibm.wordpress.com/2013/12/15/gambling-on-gamification-gimmicks/gambling-12-samsung/

gamified, and one fully-gamified) and compared participation and results obtained with those obtained from a traditional market research survey. Their study gathered over 600 participants and showed that task motivation increased with the inclusion of game elements.

In 2011 Flatla et al. showed how adding game elements to calibration tasks can increase participants' perception of fun and enjoyment and motivate participants to perform better. In their study the authors created three calibration games to replace the typical calibration tasks for determining color perceptibility, optimal C-D (control-display) ratios for use with mice, and input ranges for a physiological sensor. Comparing the calibration games with typical calibrations tasks, participants reported a significant difference in fun and enjoyment. Results for two of the calibration comparisons did show significant differences; however this was attributed to the game elements motivating the participants to expend more effort in the game versions than in the normal calibration tasks.

2.2.3 Crowdsourcing through Games

Due to the motivational draw of gamification, researchers have also experimented with the use of games and game elements in crowdsourcing tasks as a way to motivate participation. This was largely introduced in von Ahn and Dabbish's successful GWAPs (Games With a Purpose) that have participants, as a side effect of playing, perform basic tasks that are unable to be automated by current technology (von Ahn & Dabbish, 2008). The most well-known GWAP is the ESP game (von Ahn & Dabbish, 2004), which has participants paired up in a game where they must try to guess what word the other participant is thinking after being shown an image obtained from Google. The goal of this GWAP was to have participants provide accurate and descriptive labels for images on the internet. ESP managed to attract over 13,000 participants in four months and resulted in the accurate labeling of nearly 300,000 images. Crowdsourced games have been used to examine many research problems such as studying protein structures (Cooper et al., 2010), improving web accessibility (von Ahn et al., 2006), attracting participation in relevance assessment tasks (Eickhoff et al., 2012), and generating social network graphs (Guy et al., 2011).

Despite the benefits gamification has shown to provide, it is still largely unknown what effects adding game elements can have on the quality of the data received from such tasks. Also the development of gamified systems can be largely time consuming, requiring that highly customized games be developed for each research problem being studied (Flatla et al., 2011; von Ahn & Dabbish, 2008).

2.3 FREEMIUM MODEL

The Freemium business model is a model whereby an initial service is provided to all users for free, but includes the option of receiving additional services or features for a fee. This model relies on the existence of a population of users that are willing to pay the fee in order to gain access to the additional content, which in turn covers the cost of the service for the users who consume it for free (de la Iglesia & Gayo, 2009).

2.3.1 Freemium in the Mobile Market

The Freemium business model has seen a large adoption in the mobile app market, offering developers an opportunity to reach a large user base while still covering the costs of development or generating a profit. This model has been adapted and used in various ways in the mobile market. Below we introduce some common ways games have implemented the Freemium model.

Unlocking Additional Content: Some games will provide an initial set of content for free and require payment in order to unlock additional content. An example of this would be games that provide demo levels or demo gameplay and require players to pay a fee in order to access the full game (Cut the Rope[5] by Chillingo).

Game-Based Extras: There are many games that will offer basic gameplay for free but will provide game-based extras such as power-ups, lives, costumes, characters, or in-game currency

---

[5] play.google.com/store/apps/details?id=com.zeptolab.ctr.ads

for a fee (Clumsy Ninja[6] by Natural Motion, Puzzle and Dragons[7] by GungHoOnlineEntertainment).

Removing Negative Experiences: Games containing negative experiences often provide players with the option to pay in order to bypass or reduce these negative experiences. For example, in games containing ads, players are often able to pay in order to have the ads removed (e.g., Bombcats[8] by Chillingo).

Replenishing Depleted In-Game Resources: In games where the player's actions are enabled through limited resources that replenish after a set period of time, such as "energy" (Gods Rush by IGG[9]) or "lives" (Candy Crush Saga[10] by King), players can be offered the opportunity to instantly replenish the resource for a fee so that they are able to continue playing instantly.

---

[6] play.google.com/store/apps/details?id=com.naturalmotion.clumsyninja
[7] play.google.com/store/apps/details?id=jp.gungho.padEN
[8] itunes.apple.com/ca/app/id595968088
[9] play.google.com/store/apps/details?id=com.igg.android.godsrush_en
[10] play.google.com/store/apps/details?id=com.king.candycrushsaga

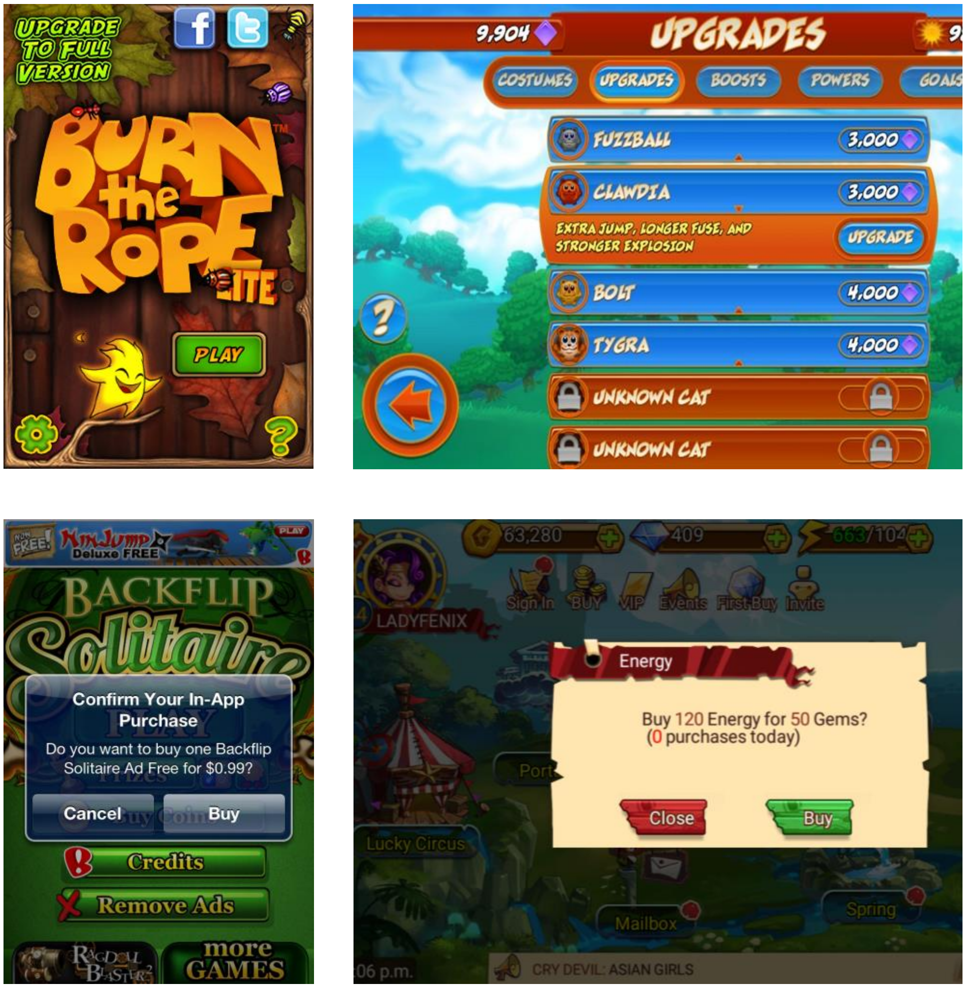

Figure 2. Top Left: Burn The Rope Lite highlighting *unlocking additional content*[11]; Top Right: Bombcats highlighting *game-based extras*[12]; Bottom Left: Backflip Solitaire highlighting *removing negative experiences* [13]; Bottom Right: Gods Rush highlighting *replenishing depleted resources* [14].

---

[11] Image retrieved from http://mobilegamepatterns.com/post/11685516393/burn-the-rope-lite-main-menu.
[12] Image retrieved from http://puzzle536.rssing.com/chan-12853795/all_p3.html
[13] Image retrieved from http://www.placeplay.com/3-simple-in-app-purchase-items-increase-app-revenue/
[14] Image taken by author.

2.3.2 Motivating Actions through Freemium

Mobile developers have begun to experiment leveraging the Freemium Model as a way of motivating players to perform other actions beyond payment. For example, Candy Crush Saga by King rewards players with additional lives for reaching out to friends via their social network. In this example, players are rewarded with their game-based extra for performing the desired action, and Candy Crush is rewarded with increased advertisements and the ability to access a much larger potential player-base.

Given the success of the Freemium model at attracting players and motivating behaviour, we are interested in whether we can leverage this model and ask players to do research-benefiting work in order to play a mobile game for free and gain in-game rewards.

2.4 MEASURING PLAYER EXPERIENCE AND MOTIVATION

In order for us to know that the Freemium model is motivating participants to do research-benefitting work, we need to determine whether or not participants are motivated to play our game. Below we introduce and discuss two validated scales used to measure player experience and motivation – Player Experience of Needs Satisfaction (PENS) and the Intrinsic Motivation Inventory (IMI).

2.4.1 Players Experience of Need Satisfaction

The Player Experience of Need Satisfaction (PENS) is a measure of player experience based on Self Determination Theory (SDT), which posits that player experience is derived when players "seek to satisfy psychological needs within the context of play" (Ryan et al., 2006).

The main needs outlined by Self Determination Theory consist of competence, autonomy, and relatedness. Immersion (or presence) is also highlighted by SDT to be associated with motivation in gaming (Ryan et al., 2006).

Competence: According to SDT, the feeling of competence is associated with facing and overcoming challenges (Deci, 1975). Within games this can be facilitated with mechanics such as the gradual increases in task difficulty (Birk & Mandryk, 2013). The competence subscale consists of 5 items. Items include statements such as "I felt very capable and effective".

Autonomy: Feelings of autonomy are garnered through a sense of control over one's actions and a willingness to participate in a task (Ryan et al., 2006). The subscale measuring autonomy also consists of 5 items, including "I did things in the game because they interested me".

Relatedness: The experience of relatedness derives from feeling connected to others (Ryan et al., 2006). In the context of games, this feeling of connectedness does not necessarily have to be experienced with other human players though that is the situation in which it is most studied (Ryan et al., 2006). The relatedness subscale contains 5 items, such as "I find the relationships I form in this game important".

Immersion: Immersion is experienced through feelings of actually being within the game world and experiencing it as though it were real, rather than feeling like a separate person on the outside looking in. Immersion is measured with a 3-item subscale and includes such items as "when moving through the game world I feel as though I am actually there".

The PENS scale has been employed in prior studies to show that a player's experience of competence, autonomy, relatedness, and immersion within a game can affect player motivation (Birk & Mandryk, 2013; Watson et al., 2013; Vicencio-Moreira et al., 2015).

2.4.2 Intrinsic Motivation Inventory

The Intrinsic Motivation Inventory (IMI) is a measure that assesses a participant's subjective experience of an activity along multiple dimensions (Ryan, 1982) and has been previously used to examine player experience in video games (Ryan et al., 2006; Birk & Mandryk, 2013). The core dimensions measured consist of interest-enjoyment, perceived competence, effort-importance, and tension-pressure (Intrinsic Motivation Inventory, 2015; McAuley et al., 1989).

Each dimension acts as a subscale that measures the participant's agreement to given statements on a 5-point scale, with options ranging from "strongly agree" to "strongly disagree".

Interest-Enjoyment: This subscale acts as a self-reported measure of intrinsic motivation, as this measure gauges how much the player was self-interested in playing the game. Items for this measure include statements such as "Playing the game was fun" and "I would describe this game as very interesting".

Perceived Competence: This subscale measures a participant's perceptions on how competent they were at the task being studied and it is thought to be a positive predictor of intrinsic motivation (Intrinsic Motivation Inventory, 2015). Items for this measure include statements such as "I am pretty skilled at [the task]" and "I tried very hard [at the task]".

Tension-Pressure: Measures of tension and pressure are negative predictors of intrinsic motivation (Intrinsic Motivation Inventory, 2015) and include statements such as "I felt tense while [completing the task]" and "I was anxious while [completing the task]".

Effort-Task Importance: This measure determines how much investment a player has in completing the task. It is not always considered relevant to certain motivation questions, but it has been validated along with the other three dimensions (McAuley et al., 1989). Items for this measure include statements such as "It was important to me to do well [at the task]" and "I tried very hard while [completing the task]".

The IMI has been employed in prior studies on player experience (Birk & Mandryk, 2013; Watson et al., 2013; Vicencio-Moreira et al., 2015).

CHAPTER 3

3 SYSTEM DESIGN AND IMPLEMENTATION

This chapter discusses the design and implementation of our prototype framework as well as our prototype game, Sugar Rush. First, we outline the architecture and technical implementation of our framework as well as the responsibilities of each component. Second, we introduce Sugar Rush and discuss the decisions behind its design and implementation.

3.1 OVERVIEW

Our goal in this thesis was to develop a framework that combines crowdsourcing and gamification into a flexible platform for delivering experimental tasks via the mobile game market. Taking a modular component-based approach to our design, we separated and isolated the experimental tasks that were included in the framework from the game with which they were deployed. Our framework used a prototype Android game, Sugar Rush, which we built around a popular game mechanic and which allowed for quick game sessions, making it an ideal option for those brief moments of downtime during a person's day. We used the Freemium model of micro-transactions to integrate and deliver our microexperiments. Players earned credits through the completion of microexperiments, which they were then able to spend on power-ups to enhance their gameplay. To investigate performance between different types of tasks, two classic human computer interaction tasks, one motor task and one cognitive task, were implemented and included within our framework.

3.2 REQUIREMENTS

To ensure that we improved upon the flexibility of traditional gamification and crowdsourced game approaches, we developed a set of requirements to be taken into account when designing our platform.

3.2.1 Handle Multiple Experimental Tasks

Crowdsourcing platforms house multiple experimental tasks whereas gamified experiments typically consist of a single custom game developed to answer a specific question. We wanted our gamified platform to be able to deliver multiple experimental tasks, as is done by crowdsourcing platforms. We also wanted to make sure that we developed a platform that could handle completely unrelated tasks within the same game at the same time.

3.2.2 Easily Add, Modify, or Remove Tasks

The current integrated approach to gamified experiments means that changes to the underlying experimental task would need to be reflected in the game mechanics surrounding them, and if you wanted to study an entirely different experimental task, the gamified system would need to be mostly – if not entirely – re-developed. We wanted our platform to be flexible and re-usable without requiring heavy modifications and developmental effort every time we needed to modify or change the tasks that we had originally included.

3.2.3 Easily Modify or Replace the Game

In the same way that we wanted to make sure changes to the underlying tasks would not affect the game surrounding them, we also wanted to ensure that changes and modifications could be made to the game without requiring modifications to the tasks. We wanted this modularity to be taken far enough to even allow for the game to be completely replaced without the tasks being aware or affected.

3.2.4 Able to Use Any Freemium Approach to Deliver Tasks

In our system, the Freemium model determines when during the game the experimental tasks will be delivered. Not all Freemium models make sense with all genres of games, and there may be certain Freemium models that work better for a given game type or even for a specific game. Having the flexibility to modify the model that is used would increase the game types that could be used within our system. Not having the tasks bound to any particular Freemium model could also open up the possibility of using multiple Freemium models within the same game (e.g., micro-transactions to unlock content or to enable progress).

3.3 SOFTWARE ARCHITECTURE

We took a modular component-based approach to designing our framework, ensuring separation between the mobile game that is used to deliver the tasks and the experimental tasks that are being delivered. This was done to maximize the flexibility of our framework and to allow researchers to include multiple experimental tasks into a single instance of the framework, as well as easily and seamlessly add, edit, or remove experimental tasks at any time without having to modify the game being used to deliver the tasks. This separation also allows for researchers to use multiple mobile games to deliver a set of experimental tasks. Using multiple mobile games to deliver experiments can increase the potential participant base, target different demographics, and sustain interest in the experimental tasks.

Our game, Sugar Rush, was developed to run on Android devices (running Android 2.3+) using the libGDX[15] game development framework. LibGDX runs on Java and so the task manager and task components were developed in Java and simply plug in to the libGDX framework.

---

[15] libgdx.badlogicgames.com/

3.3.1 Components

Our modular framework consists of three main components – the game, the task manager, and the task. We introduce these components and outline their responsibilities below.

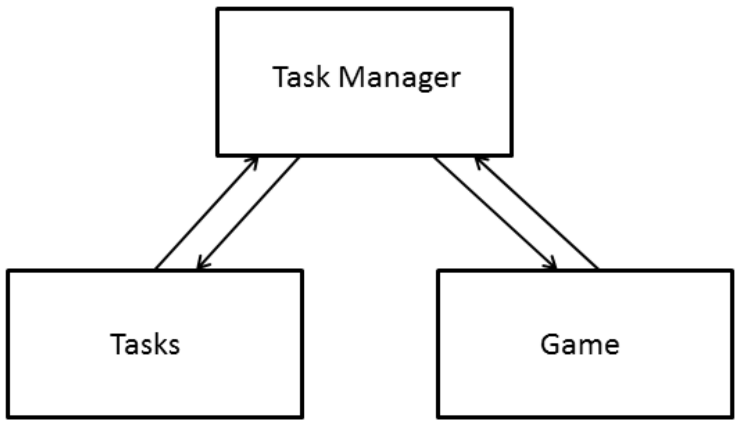

Figure 3. The conceptual layout of the three main components and the interaction between them.

3.3.1.1 Task Component

The task component consists of the individual research tasks that have been developed and included within the framework. Each task implementation is self-contained and is responsible for managing its own display and logic. The task component communicates with the task manager component only and is completely unaware of the existence of the game.

There can be multiple individual task modules within the framework at a time. The task manager component keeps track of the different task implementations that are included in the framework and notifies the appropriate task when it is its turn to be executed.

3.3.1.2 Game Component

The game component consists of the mobile game with which our framework has been included. The game is a completely self-contained implementation as well and communicates only with the task manager. The game component has no knowledge of the existence of the task component, or the individual task implementations that have been included in the framework.

The game component communicates with the task manager and notifies the task manager when the player has interacted with the Freemium implementation (signifying that a task should be executed). In our prototype, this interaction would occur when the player presses the “earn more credits” button from the main menu.

3.3.1.3 Task Manager Component

The task manager component is central to the system and is responsible for choosing and executing the experimental tasks.

All experimental task logic is completely isolated from the task manager and contained within individual Task objects. These objects conform to a common Task interface, which allows the manager to notify the chosen task when it is supposed to run its programmed experiment. The task manager acts solely as a container and controller from the experimental tasks and task objects can be modified, added, or removed quickly, requiring only the development of a single object and an update pushed to the Android market.

The task manager itself is isolated from the gameplay aspect of the framework. The game has no knowledge of the tasks included in the task manager and the task manager has no knowledge of the game the player is playing or the rewards the player will receive for completing the tasks. When the player taps the “add credit” button to earn more credits, the manager receives a call from the game, chooses a task, and notifies the chosen task to run its programmed experiment. Once the player has completed the experiment the game receives notification from the manager that the task was completed; it is up to the game itself to determine what type of reward the player then receives.

3.3.2 Task Integration

Integrating the tasks in this way decouples the experiments from the actual gameplay itself, while still providing the motivational framework. This is important because we wanted to ensure that the experimental tasks could be modified or substituted without the need to modify the underlying game mechanics. In this way, a single game can accommodate a variety of experimental tasks, as opposed to the previous integrated approach of creating a new game for each task. Likewise this allows our task manager to potentially be integrated into many different game types utilizing different variations of the Freemium model in order to target multiple demographics or sustain interest and participation in the platform.

3.3.3 Flow of Control

To demonstrate how our system functions as a whole, we walk through the standard usage of our prototype game to highlight the flow of control and interaction between the components in our framework.

1. From Sugar Rush, the participant clicks on the “earn more credits” button.
2. The game component notifies the task manager that the player has clicked on the button.
3. The task manager chooses a task from the task component and notifies it that it is to become active.
4. The participant completes the displayed task.
5. The task notifies the task manager that it has been successfully completed.
6. The task manager relays the message to the game and notifies the game that it is to reward the player for clicking on the “earn more credits” button.
7. The game component gives the participant a credit as their reward.

3.4 GAME DESIGN

In this section we describe the design of our prototype game, Sugar Rush.

3.4.1 Gameplay

Our game follows a genre referred to as an infinite vertical platformer, which has seen recent popularity in the mobile market (e.g., Doodle Jump[16] by Lima Sky LLC, Happy Jump[17] by Noodlecake Studios Inc., Mega Jump[18] by Get Set Games Inc.). In our game, Sugar Rush (see Figure 4), players control a continuously bouncing avatar that they must guide through a series of pseudo-randomly generated platforms in order to collect as many candies as possible. The player controls the avatar by tilting the device left or right (which moves the avatar left or right), navigating their way to bounce on top of platforms. Players also try to gather candy and avoid enemies in an attempt to attain a new high score. If the avatar collides with an enemy or if it misses a platform, the player “dies”, falls off the bottom of the screen, and the game session ends.

We chose this style of game because it allows players to quickly get into or leave game sessions so players do not need to invest a large amount of time in order to play the game. This makes our game a good choice for those moments we are trying to leverage for participation in our experiment, i.e., when someone needs a quick distraction.

Another advantage of using this particular game type is that it lends itself well to auto-generation, allowing the game world to be dynamically created during each game session. This means that the player will get a different gameplay experience each time they play, which increases replayability.

---

[16] play.google.com/store/apps/details?id=com.lima.doodlejump

[17] play.google.com/store/apps/details?id=com.noodlecake.happyjump

[18] play.google.com/store/apps/details?id=com.getsetgames.megajump

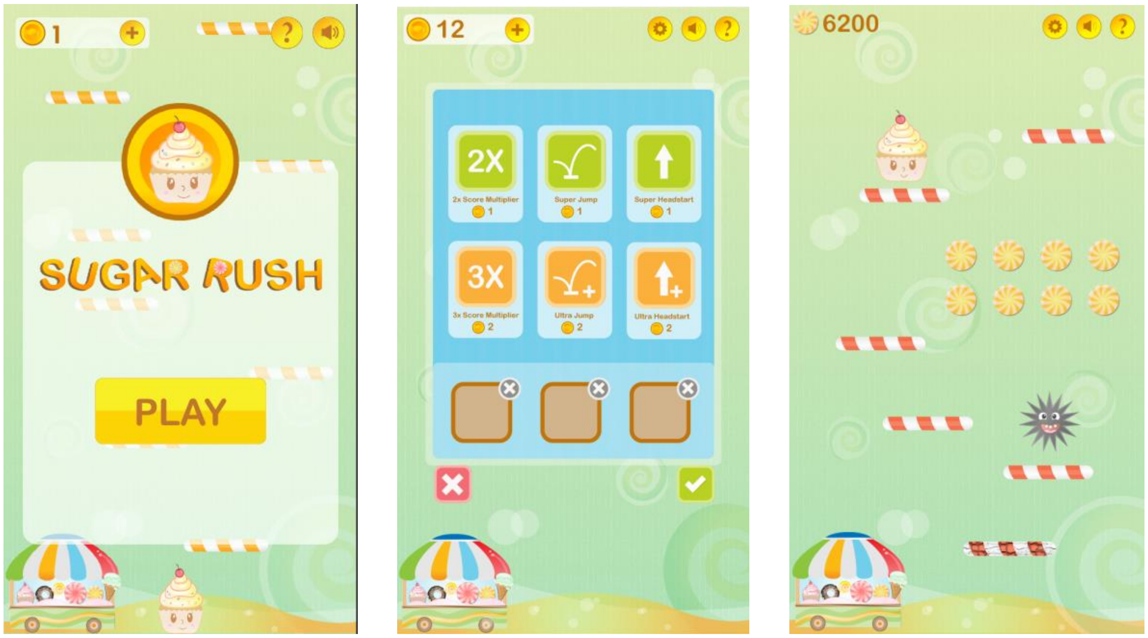

Figure 4. Sugar Rush: Title screen (left), Power-up menu (middle), Gameplay (right).

3.4.2 Power-ups

In Sugar Rush, players have the opportunity to purchase power-up items to assist them in their gameplay. These power-ups each contain an effect that alters the gameplay in an advantageous way for the player. There are three types of power-ups to choose from each with two levels of strength to their effect (see Figure 4).

Multiplier: The multiplier power-up increases the amount of points that player receives from collecting candy. At the first level it multiplies the individual candy points by 2, and at the second level it multiplies the points by 3.

Headstart: The headstart power-up modifies the player's velocity and blasts the player flying up into the sky at a tremendous speed, flying them past the starting sections of the game. While the player is boosted in the air, their avatar collects all candies that it passes along the way, providing the player with a substantial starting score. At the first level, the headstart boost lasts for approximately 2 seconds, while at the second level the duration is approximately 3 seconds.

Jump: The jump power-up modifies the avatar's jumping velocity, allowing the player to bounce higher every time they jump on a platform. At the first level, the avatar's jump velocity is increased by a factor of 1.2, and at the second level it is increased by a factor of 1.6.

These power-ups are purchased using credits earned through the completion of experimental tasks that have been deployed with the game. The first-level power-ups cost 1 credit each and the second-level power-ups cost 2 credits each. We wanted to offer multiple tiers of power-ups to gives players some incentive to do multiple tasks at a time, because the higher-level power-ups require the player to have multiple credits at a time. The player can purchase and use three power-ups within a single game session, but they are only able to purchase a single power-up of each kind (i.e., they cannot purchase both a level 1 and a level 2 multiplier). We wanted to allow players to use multiple power-ups within a single game session in order to promote the idea of players purchasing multiple power-ups at a time, again requiring that the player has earned multiple credits.

3.4.3 Local Scoreboard

The game maintains a local scoreboard, which is displayed after every session and contains the top five scores achieved on that device. A *game session* is defined as one round of play from the beginning of the game until the avatar "died". We chose a local scoreboard over a global scoreboard due to the technical difficulties of implementing and maintaining global stats; however, in future work it would be interesting to see if global leaderboards or score sharing capabilities like posting to social media platforms, provide an increased motivational effect.

3.5 TASKS

To evaluate the system, we chose to investigate two common experimental tasks that differ in their requirements. As such, we implemented and integrated the following tasks within our system: Fitts' reciprocal tapping task (MacKenzie, 1992), and Cleveland and McGill's perception task (Cleveland & McGill, 1984).

We chose a motor task and cognitive task to help determine if the system was better suited to one type of experimental task. We might expect to see degradations in the cognitive task that requires attentional resources; however, it is possible that the mobile nature of the participant as well as the small screen real estate provided by a smartphone or tablet will result in degradations in a motor task.

3.5.1 Motor Task

The Fitts' reciprocal tapping task was created to understand human motor processing in physical environments (Fitts, 1954) and has been extensively appropriated in computer-based aiming research (MacKenzie, 1992). In this task, players tap on two horizontal bars in an alternating pattern (see Figure 5).

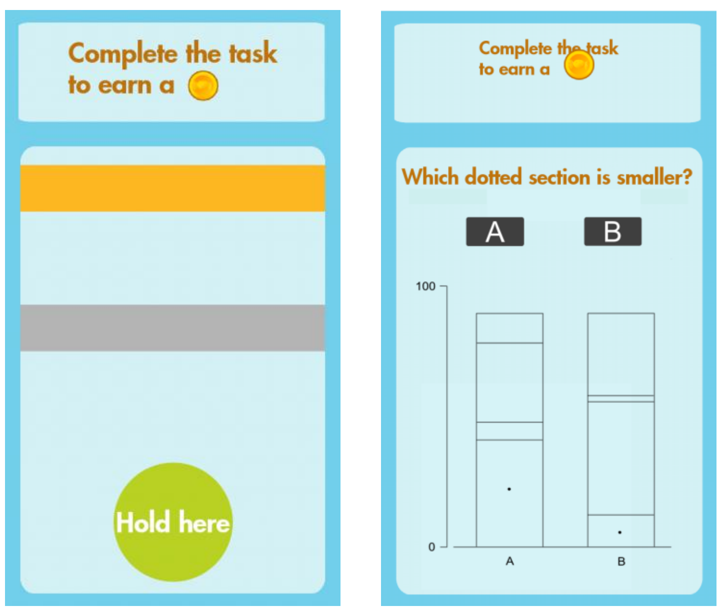

Figure 5. In-game motor (Fitts) task (left), in-game cognitive (McGill) task (right).

First, players are shown an instruction screen that explains the goal of the task, and contains what we refer to as an anchor – a button on the screen where players are asked to place their thumb in order to start the task. Once the player places their thumb on the anchor, the two horizontal bars are displayed and the participants are instructed to tap the bars with their opposite thumb. The anchor was included to try and avoid the situation where a player uses both thumbs when tapping the bars, one thumb over each bar. This behaviour would "game" the system by producing fast completion times for the task that do not conform to the expectations of a reciprocal tapping task.

There are three sizes of horizontal bars (W) as well as three distances (D) between bars that can be displayed in the task, resulting in seven unique indices of difficulty, according to the following equation: $ID = \log_2\left(\frac{2D}{W}\right)$, where ID = index of difficulty (MacKenzie, 1992).

A single combination is randomly chosen and used for a single task session. The system logs player performance metrics, including the time between successive taps, the number of errors made, as well as how long it took to complete the task.

We chose the Fitts task because it has been well studied in literature in human-computer interaction (see Balakrishnan, 2004) and represents a standard motor task. We assumed that the low attentional resources required to complete this task would make it a good candidate for data collection through crowdsourcing; however, we were concerned with whether or not the mobile aspect of the data collection would affect participant response.

3.5.2 Cognitive Task

The Cleveland and McGill perception task (Cleveland & McGill, 1984) was created to understand how humans perceive and respond to various ways of visually presenting quantitative data, and has been used in research in information visualization (Heer & Bostock, 2010). In this task, data is presented to a user in different chart formats and users are asked to judge which category is smaller and by how much.

In this task, players are first shown a bar graph (as seen in Figure 5) and asked to identify which data category, indicated by the dots, is smaller ("Which dotted section is smaller?"). The players indicate their choice by pressing on a corresponding A or B button. They are then presented with a text field and asked to estimate the percentage that the smaller bar takes up of the larger bar ("How much of the large section (in percentage) would the small section fill? (ex. 30.5)"). Five different chart types were represented in the system. The system logs player responses, the task characteristics, as well as task completion time.

We chose this task because it represents a well-studied task in the area of visual perception and has been previously used in evaluation of crowdsourced data collection on the AMT platform (Heer & Bostock, 2010). Our assumption was that the attentional resources required to complete the task might result in a reduction in the quality of data collected through crowdsourcing; however, that the task lent itself well to completion in a mobile environment.

3.6 DATA COLLECTION

For the evaluation of our framework, we logged both task and game performance data. The data was logged using an SQLite database, and stored locally on participants' devices before being opportunistically uploaded to a remote NoSQL database hosted on Cloudant[19].

Data gathered from completion of tasks and game sessions was saved in memory after completion and would be written to the local SQLite database when the application was interrupted either by the player closing the application, the player hitting the home button sending the application to the background, or an incoming call taking focus away from the application. Upon startup, the application would look for an active network connection and when one was found all new data would be pushed to the Cloudant server using JSON and HTTP requests.

3.6.1 Assigning Unique Participant IDs

Participants' devices were assigned a unique randomly-generated ID on the first launch of the game. This is the ID that was associated with all of their data in the logging records and could in no way be used to personally identify the participant's device. This introduced a drawback where

---

[19] https://cloudant.com/

in the event that a participant uninstalled our game and re-installed it at some point in the future, we would not be able to identify that their device was associated with previous participant records, as they would receive a new ID. We felt that the increased concern for privacy outweighed the damage that would be cause by an occurrence of this scenario.

CHAPTER 4

4 METHODOLOGY

This chapter will cover the factors isolated by our study as well as the game variations used to gather experimental data under our three groups. We will also discuss the surveys used to gather participants' reactions to the inclusion of experimental tasks within a free-to-play game.

4.1 FACTORS

To fully evaluate our framework, we need to isolate and account for three main factors in order to see the effect they have on the quality of the data gathered: the mobile environment, the gamified environment, and the crowdsourcing of data collection. The fully counterbalanced matrix is presented in Figure 6.

For our initial investigation, presented in this thesis, we chose to isolate the factors of crowdsourcing and gamification. The effects of the mobile environment will be investigated in future work.

	Mobile	Gamified	Crowdsourced
Crowdsourced Game	✓	✓	✓
Laboratory Game	✓	✓	
Mobile Crowdsourced Tasks	✓		✓
Laboratory No Game	✓		
AMT Desktop Game		✓	✓
Desktop Laboratory Game		✓	
AMT Tasks			✓
Traditional Laboratory Tasks			

Figure 6. Conditions required to fully counterbalance the factors present within our framework.

4.2 GROUPS

We used multiple variations of the Sugar Rush game in order to observe the motivational draw of power-ups and the quality of experimental data gathered under differing experimental situations. Three groups, each using their own version of the Sugar Rush game, were used in order to isolate and observe the effects of gamification and crowdsourcing separately – Crowdsourced, Laboratory, and Control.

4.2.1 Crowdsourced

The crowdsourced group consisted of the complete version of the system that was released on the Android market. Data from the game was gathered over 5 weeks and included data from 841 unique downloads. This group investigates the factors of gamification and crowdsourcing in combination.

4.2.2 Laboratory

The laboratory group investigates the gamification factor in isolation, and was conducted within a lab setting overseen by an experimenter. The game used in this group is identical to the crowdsourced game, except that it includes an additional starting screen that prompts the experimenter to enter an assigned participant id, rather than have one generated by the game itself.

4.2.3 Control

The control group investigated performance on the tasks in isolation from both crowdsourcing and gamification factors. The version of the system that was used removed all elements of the game and instead organized the tasks into a series of 7 blocks with 14 tasks in each block – one for each unique combination of bar height and bar width in the Fitts task (9 in total) and one for each graph in the McGill task (5 in total). The tasks were identical to the ones used in the crowdsourced and laboratory games except any mention of the game or earning credits was removed. At the beginning of each block of tasks and in between each task the participant was presented with a screen that prompted them to take a break if needed. The order of the tasks within the blocks was randomized. This version falls under neither the gamification nor the crowdsourcing conditions and acts as the control condition for data analysis.

4.3 SURVEYS

Surveys were issued to participants in all three groups to gather data regarding demographics, prior experience with mobile games, player motivation, player experience, and subjective responses to statements concerning the deployment of tasks within a mobile game.

In both the control and laboratory groups, participants filled out demographic and mobile game surveys after completion of the study. Participants in the laboratory group were also asked to complete surveys gauging player motivation, player experience, and statements concerning players' attitudes and thoughts towards being asked to complete experimental tasks within a mobile game setting. These final surveys were not included in the control group, as participants were never exposed to the game.

Surveys for the crowdsourced group were delivered within the game and participation in the surveys was optional. After two weeks from the time of installation, so as to give the players adequate exposure to the task framework, the game presented a dialogue box giving players the option to participate in the survey. Players were offered a reward of 25 in-game credits as incentive. If the players chose to participate, the browser on their mobile device was opened and they were directed to the surveys hosted on the University of Saskatchewan's FluidSurveys[20] platform. If participants indicated that they did not want to participate, they were allowed to continue on to the game but would be asked again after two days. If at that time, they again indicated that they did not wish to participate in the survey, the system would suppress itself and they would not be prompted again.

[20] https://fluidsurveys.usask.ca

We made the survey optional and instead decided to provide incentive for completion so as to not deter further participation in the game by forcing players to complete the survey before being allowed to continue on with their gameplay. We felt that gathering more usage data by having players play the game was more valuable than have players fill out a one-time survey. However making participation voluntary, and simply providing incentive, came with a tradeoff as it is known that gathering subjective data voluntarily through uncontrolled crowdsourcing is a challenging task (Kittur et al., 2008), which was reflected in our data.

There were very few participants who elected to fill out the survey when prompted. A total of 8 participants were recorded to have started the survey. Out of those 8 only 6 of the participants were shown to have completed the entire survey. There was not enough data gathered from these surveys to perform any form of analysis.

4.4 SURVEY QUESTIONS

Below we discuss the questions used within our studies. Full instruments have been included in the appendices.

4.4.1 Demographics and Mobile Games

Basic demographic information, i.e., sex and age range, was gathered. In addition, participants were asked questions regarding their prior experience with mobile games and the typical usage of their devices. We asked questions determining how often they use their device to play mobile games, typical duration of play, and also if any other people that typically use their device to play games. Questions like these were important because it gave us an idea of how effective data collection would be through mobile games, and it also gave us an idea of how common it is for multiple people to operating the same mobile device. Since our framework assigns ids on a per-device basis, knowing this is important to determine the effect it would have when we are gathering performance data from multiple people under the same participant id.

4.4.2 Player Experience

Participants were asked to rate their agreement with statements from the validated scale *Player Experience of Needs Satisfaction (PENS)* to measure player experience. The statements that were included covered the measures of perceived competence (i.e., “I feel very capable and effective when playing”), perceived autonomy (i.e., “Sugar Rush provides me with interesting options and choices”), experienced relatedness (i.e., “I find the relationships I form in Sugar Rush fulfilling”), and experienced immersion (i.e., “Sugar Rush was emotionally engaging”). This part of the survey consisted of 18 statements total (3 for competence, 3 for autonomy, 3 for relatedness, and 9 for immersion), and all statements were rated on a 5-point scale of agreement from “strongly disagree” to “strongly agree” (see appendices).

4.4.3 Player Motivation

Player motivation was gathered subjectively, with participants rating their level of agreement with statements taken from the validated scale *Intrinsic Motivation Inventory (IMI).* Statements included covered the measures of interest-enjoyment (i.e., “I enjoyed Sugar Rush very much”), perceived-competence (i.e., “I am pretty skilled at Sugar Rush”), invested effort (i.e., “It was important to me to do well at this game”) and experienced tension (i.e., “I was anxious while playing Sugar Rush”). This survey also consisted of 18 statements total, and as with the player experience scale, all statements were rated on a 5-point scale of agreement from “strongly disagree” to “strongly agree” (see appendices).

4.4.4 Inclusion of Tasks in Gaming

We included 10 statements to gauge the participant’s frequency of play, use of power-ups, and thoughts regarding the inclusion of tasks within a free-to-play game. Participants were asked to indicate their level of agreement using a 5-point scale from “strongly agree” to “strongly disagree” (see appendices).

CHAPTER 5

5 EVALUATION

To evaluate our framework we were interested in examining two main questions. First, we needed to find out whether the quality of the work done by participants meets the standards set by similar experiments conducted using more traditional approaches, such as in the lab. Second, we needed to know whether participants were motived to play our game and whether they found enough value in the power-ups to “work” for them by doing tasks. To investigate these questions, we developed three versions of the Sugar Rush game to study the effects of gamification and crowdsourcing separately – Crowdsourced game, Laboratory game, and Control.

5.1 PARTICIPANTS AND SETTING

Participants were unique to each group, i.e., there was no overlap between the three versions. The crowdsourced game was uploaded to the Google Play Store and was available to anyone that downloaded the game and agreed to the ethical terms of service that explained participation in the study. The crowdsourced game was available to all Android devices running Android 2.3 or higher – excluding the Motorola Droid 2X due to technical difficulties. Twenty participants were recruited via the university mailing list for each of the Laboratory study and the crowdsourced study individually. Participants were placed in whichever study their availability allowed for. The laboratory and control groups were evaluated using a Samsung Nexus S device running Android 3.4.

5.1.1 Crowdsourced Game

Data for the crowdsourced game was collected over a period of 5 weeks and was gathered from 841 unique downloads. Upon launching the game for the first time, the players were presented with a form consenting to the collection of their gameplay data (to comply with ethical approval at the University of Saskatchewan, the only information that was logged that was not directly related to their performance in the game was the width, height, and density of the screen for their device. No other device or personal information was gathered).

5.1.2 Laboratory Game

Data for the lab game were collected in a university laboratory. Twenty people (8 female, ages between 18 and 35 years) participated. In terms of gaming experience, 90% of the participants indicated that they play games at least a few times per month; 75% indicated that they play games on a mobile device at least a few times per month; 7 reported being the only people who used their mobile device to play games; 11 reported that their friends, spouse, or children also use their device for playing games.

Participants completed a consent form (to comply with ethical approval) and were introduced to the game objects, controls, screens, as well as walked through the two task types. Participants were then left to play the game for 50 minutes. The only requirement placed on the participants during their gameplay was that between each game session they were required to complete at least one task. If the participant wanted to complete multiple tasks to be able to purchase and use multiple power-ups within their game, they were free to do so. At the end of the 50 minutes, the participants were asked to fill out a questionnaire recording demographic information and gathering thoughts regarding the inclusion of experimental tasks within the game (see appendices).

5.1.3 Control

Data for the control system were collected in a university laboratory. Twenty people (11 female, aged 18 to 35, with one participant being 36 to 45) participated. In terms of gaming experience, 90% indicated that they play games on a mobile device at least a few times per month; 13 reported being the only person to play games on their mobile device; 7 reported that their friends, spouse, children, or parents used their device for playing games.

Participants completed a consent form and were then introduced to the system and walked through the two task types. Once the participants had finished the series of tasks they then filled out a brief demographic questionnaire (see appendices).

5.2 DATA ANALYSIS

Data collected by the crowdsourced game were temporarily stored on the participants' devices and opportunistically uploaded to a document-based NoSQL server when the game detected a network connection. A python script pulled the data from the server and parsed it into a useable format. Data collected for both the lab game group and the control group were stored locally on the mobile device used in the studies. A similar python script was used to pull the data from the device and parse it into the same format used for the crowdsourced data.

We removed participants who only played one game, i.e., opened the application and played once, but never returned to the game for a second play session (260 people), and players who had an average score of zero in their game sessions, i.e., opened the game more than once but did not appear to play it (80 people) from further analyses.

5.2.1 Motor Task

In the motor task, our dependent measure was the mean movement time (MT) for a single index of difficulty (ID). Participants completed ten reciprocal taps within a trial. After removing taps that contained an error we aggregated the remaining taps, creating a measure of mean MT per trial. Outlier trials in which the MT was 3 standard deviations below the mean (52/14971 trials, 0.3%) were also removed. We then aggregated across ID to create a measure of MT for each unique ID for each participant. Although the screen sizes differed in the Crowdsourced Game, ID is calculated as a ratio of distance over width, so the seven IDs were the same across all participants in all groups. We also conducted a linear regression of ID on MT for each of the three experimental groups separately to determine the differences in quality of the model fits.

5.2.2 Cognitive Task

In the cognitive task, our dependent measures were the mean number of errors in judgment, i.e., the number of times the participant guessed that the bar was smaller, when in fact it was larger, and vice-versa (selection errors), and the logarithm (base 2) of the absolute error in the judgment of the estimation of the difference in size (log estimation errors) (Cleveland & McGill, 1984). Participants completed several repeated estimates of the same chart (there were five charts presented); data were aggregated over repeated trials of a single chart for each participant to get an indicator of central tendency.

For the Crowdsourced Game, we also created a spurious trial filter to identify trials with input that clearly was not an accurate estimate. Our goal with the spurious filter was to implement heuristics that a machine could use to remove unauthentic data. Our heuristics included: repeated entry of a single value across more than 3 successive trials, values outside the range of possibility (i.e., values greater than 300% - over 100% is not a possible result; we allowed up to 300% because it would be a possible value if the participant had misinterpreted the direction of the estimation), and values that were less than 10%. This resulted in the classification of 40% of trials as spurious (one participant was removed from further analysis for having almost entirely spurious data). We address the problem of spurious data further in the discussion.

5.2.3 Statistical Analysis

We conducted linear mixed models to investigate differences between groups and for the different levels of a task within a group, i.e., ID in the motor task and graph number in the cognitive task. We used a linear mixed model instead of an RM-ANOVA to account for the lack of full coverage in the data set for each participant in both the Crowdsourced Game and Lab Game. Because the type of task and the level of the task were presented randomly in the Crowdsourced and Lab Games, not all participants completed all levels of all tasks. The linear mixed model accounts for this missing data. Alpha was set to 0.05 and pairwise comparisons used the Bonferroni method of adjusting for multiple pairwise comparisons.

CHAPTER 6

6 RESULTS

This chapter presents the results of our evaluation organized into three sections. First, we present results related to the quality of data, i.e., how the results for the motor and cognitive task differ across the three experimental groups. Second, we present the results for data motivation, i.e., how often did players earn and access the power-ups by doing tasks in the Crowdsourced Game. Finally, we present results from the surveys on player experience.

6.1 DATA QUALITY

We present the results for the motor (Fitts) task and cognitive (Cleveland and McGill) task separately.

6.1.1 Fitts' Reciprocal Tapping Task

A linear mixed model with Index of Difficulty (seven IDs) and Group (Crowdsourced, Lab, Control) as fixed effects and participant modeled as a random factor revealed a main effect of ID on MT ($F_{6,31.6}$=38.5, p<.001). As Figure 7 shows, MT increases with ID; pairwise differences were significant, except for each ID with its closest neighbours (1.32<2.00 and up, 1.58<2.33 and up, 1.70<2.47 and up, 2.00<2.47 and up, 2.33<2.81 and up). There was, however, no significant effect of group on MT ($F_{2,56.5}$=1.85, p>0.5) or interaction of group and ID on MT ($F_{12,31.2}$=1.6, p>0.5). Although Figure 7 shows that MTs were slightly faster in the Control group, these differences were not significant.

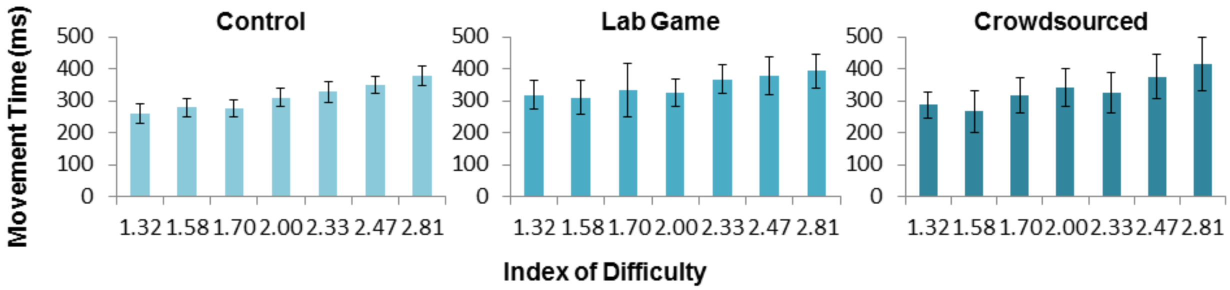

Figure 7. Mean movement time for the motor (Fitts) task across all three groups.

Figure 7 also shows how there is less variance in the mean MT for the Control group. This could be due to the full coverage of that data set (more points were used in the aggregation) or because there is greater consistency in the controlled, non-gamified task. We investigated how this variability affects the fit of a model regressing ID on MT for each group separately. Results showed that the prediction of MT by ID was significant for each group (Crowdsourced: R^2=.710, β_{ID}.84, t=3.5, p=.017; Lab: R^2=.880, β_{ID}=.94, t=6.1, p=.002; Control: R^2=.985, β_{ID}=.99, t=17.8, p<.001). However, the fit of the model was better for the Control group than the Lab Game, which was in turn better than the fit for the Crowdsourced Game. In addition, the standardized Beta values exhibit the same relationship. As previously noted, this could be due in part to the full and repeated coverage of all IDs for all participants in the Control group, but could also reflect the variability in performance that results from the game and from crowdsourcing the data collection.

These results show that there was no significant reduction in the performance of participants in the motor task when the task was gamified or crowdsourced; however, the regressions suggest that there is more variability in the data as a result of gamification and crowdsourcing. Whether or not this variability is due to the environment of the experiment or is an artifact of data coverage is unclear.

6.1.2 McGill Graph Task

A linear mixed model with graph number (Type 1 through Type 5) (see Figure 8) and experiment groups as fixed factors and participant modeled as a random factor revealed a main effect of group on the mean number of selection errors ($F_{2,9.2}$=6.15, p=.020), but no effect of graph type ($F_{4,4}$=5.8, p>.05) or interaction of graph type with group ($F_{8,5.2}$=4.6, p>.05).

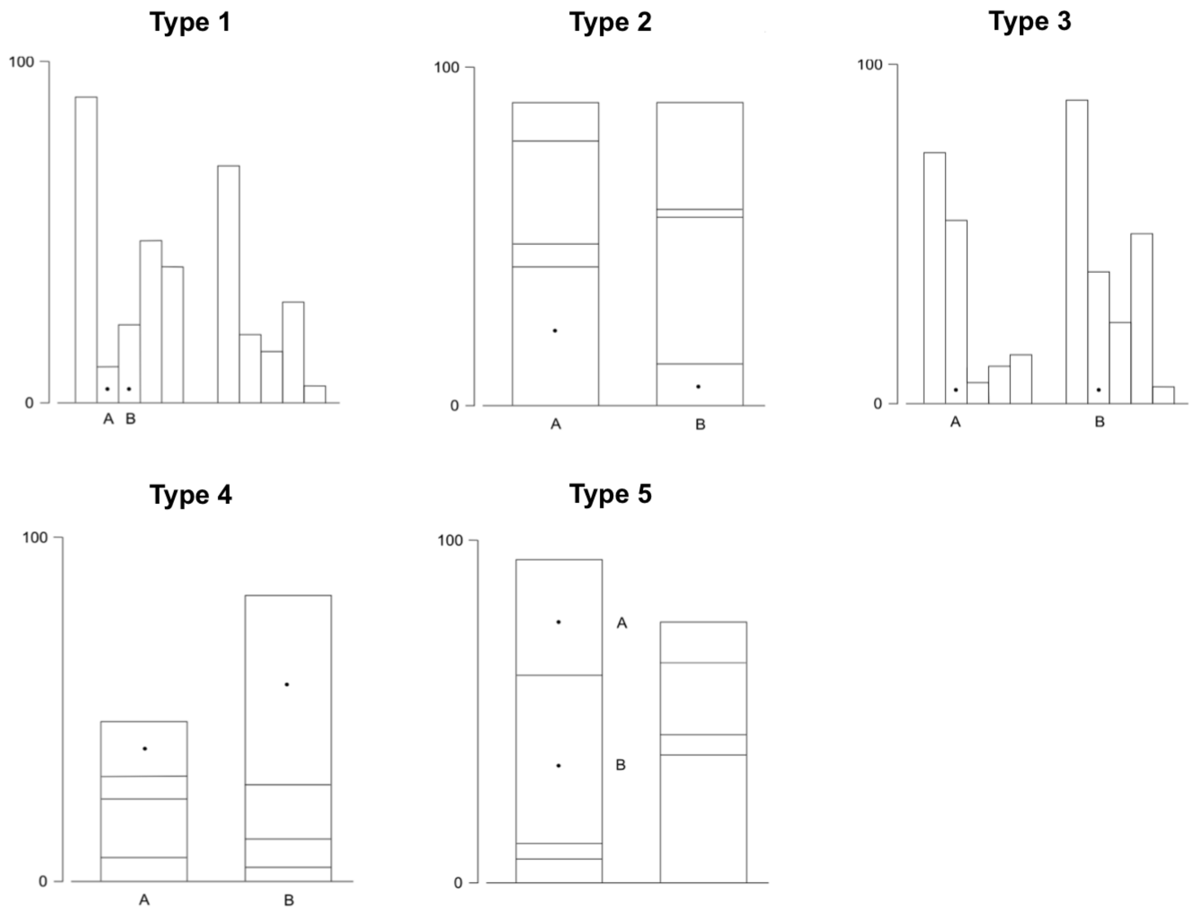

Figure 8. Graphs used in the cognitive task.

The pairwise comparisons revealed that participants made more errors in the Crowdsourced Game than the Lab Game or Control group (see Figure 9).

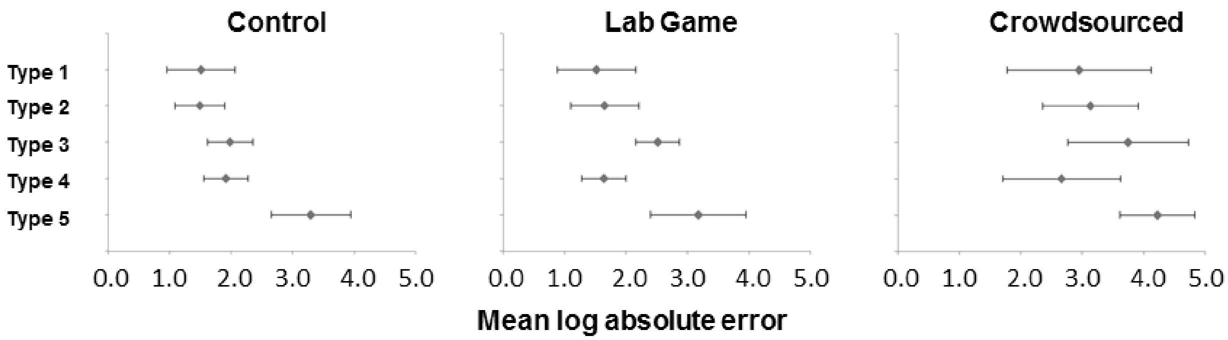

Figure 9. Mean log absolute error for the cognitive (McGill) task across all three groups.

A similar linear mixed model revealed a main effect of group on the $\log_2$ estimate error ($F_{2,44.1}$=10.1, p<.001). The pairwise comparisons revealed that participants had higher estimation error in the Crowdsourced Game than the Lab Game (p=.027) or Control group (p<.001), (see Figure 9). There was also a significant effect of graph type ($F_{4,28.2}$=18.7, p<.001), with participants performing worse with the Type 3 and Type 5 charts than the Types 1, 2, and 4 charts (see Figure 8 and Figure 9). There was a marginal interaction of group and graph type on estimation error ($F_{8,31.1=}$2.2, p=.052). As Figure 9 suggests, the poor performance of the Type 3 chart was driven by the Lab Game and Crowdsourced groups.

These results show that even with the spurious data filter applied, participants still performed worse (and with more variability) in the Crowdsourced Game than the Lab Game or Control group on the cognitive task.

6.2 MOTIVATION FOR COMPLETING TASKS

We first present survey results from the Lab Game study and follow up with power-up usage data from the Crowdsourced Game.

6.2.1 Lab Game Survey Results

In the Lab Game group, we asked participants for feedback on a variety of questions related to crowdsourcing experimental data collection through mobile gamification. Statements were posed for which participants rated their agreement on a 5-point scale from 1 (Strongly Disagree) to 5 (Strongly Agree). One-sample Wilcoxon Signed Ranks Tests compared the data to the hypothesized median result.

Participants significantly differed from the hypothesized median response (3.0) in their *agreement* with the following statements: I would play other games with embedded tasks if they were fun and free (median=4.0, χ^2=3.49, $p<.001$); Having to complete the tasks did not negatively affect my play experience (median=3.85, χ^2=3.85, $p<.001$); For a free game, I would rather complete tasks such as these than be forced to view advertisements (median=4.5, χ^2=3.09, $p=.002$); I would prefer to complete multiple tasks at a time to earn lots of credits (median=5.0, χ^2=3.70, $p<.001$); and, The tasks were easy to complete (median=5.0, χ^2=3.76, $p<.001$).

Participants significantly differed from the hypothesized median response in their *disagreement* with the following statements: I would rather not use power-ups than have to complete the tasks (median=1.0, χ^2=3.01, $p=.002$); and Having to complete the tasks ruined the game play (median=1.0, χ^2=3.15, $p=.002$).

There was no difference from the hypothesized median response for the following statements: Playing Sugar Rush with the tasks made me feel like I was contributing to science (median=3.0, χ^2=1.34, $p=.182$); and I would pay $4.99 to play Sugar Rush without the tasks (median=1.0, χ^2=1.10, $p=.272$).

These results suggest that participants are open to the idea of leveraging the Freemium model for crowdsourcing data collection through gamification, that doing so does not negatively affect their play experience, and that the tasks were easy and are preferred to viewing advertisements. In this sample of participants, we asked them to complete one task for each game, and results might differ for those participants who completed the experiment in the wild.

6.2.2 Crowdsourced Game Use Statistics

We calculated how many players used power-ups in the game. The results showed that power-ups were only used in 1.1% of gameplay sessions (SD=5.9), which suggests that players were not motivated to use power-ups. However, we also discovered that only 5.3% of players ever used a power-up at all. Conversely, those players who tried even a single power-up used power-ups in 21% of gameplay sessions (SD=16.4), which is a reasonable rate of power-up usage. Our game did not expose players explicitly to the power-ups, but those players who discovered power-ups used them in around $1/5^{th}$ of game sessions. Because players needed to earn the power-ups by completing experimental tasks, our results suggest that very few experimental tasks were ever completed. However, the 40 players who used power-ups in the game completed 1101 tasks, an average of 27.5 per person, suggesting that players who find the power-ups motivating in the game are willing to complete the tasks to earn them.

Data from the Lab Game group suggest that exposing players to the power-ups will improve their use. In that group, players were explicitly asked to complete tasks and thus had credits available to purchase power-ups for use in their game sessions. Nineteen of the twenty participants used power-ups in game, and those that did used them in 69% of game sessions (SD=24). Although we would not expect to see this rate of power-up usage in the wild, there is an indication that exposing the players to the power-ups might increase their appeal in the game. In the discussion, we outline how this could be addressed through small changes to game design.

6.3 PLAYER EXPERIENCE

In the Lab Game group participants were asked to fill out questions from the IMI and PENS validated scales for player experience. The average subscale scores (with standard deviation) for

PENS and IMI are summarized separately in Figure 10. Since it was not part of our experiment to have participants of the Lab Game group play the game without tasks, we are not able to isolate the effect that the inclusion of tasks has on players' experience of the game.

Although we are not able to draw any concrete conclusions from the player experience data, looking at the distributions we can see that the overall average score for the Interest-Enjoyment scale of IMI is just below the median score for the scale (3.0). This is typical of the scores we often see with research prototypes of casual games and would seem to suggest that the inclusion of tasks within Sugar Rush did not provide a negative experience for the participants, which is in agreement with our hypothesis.

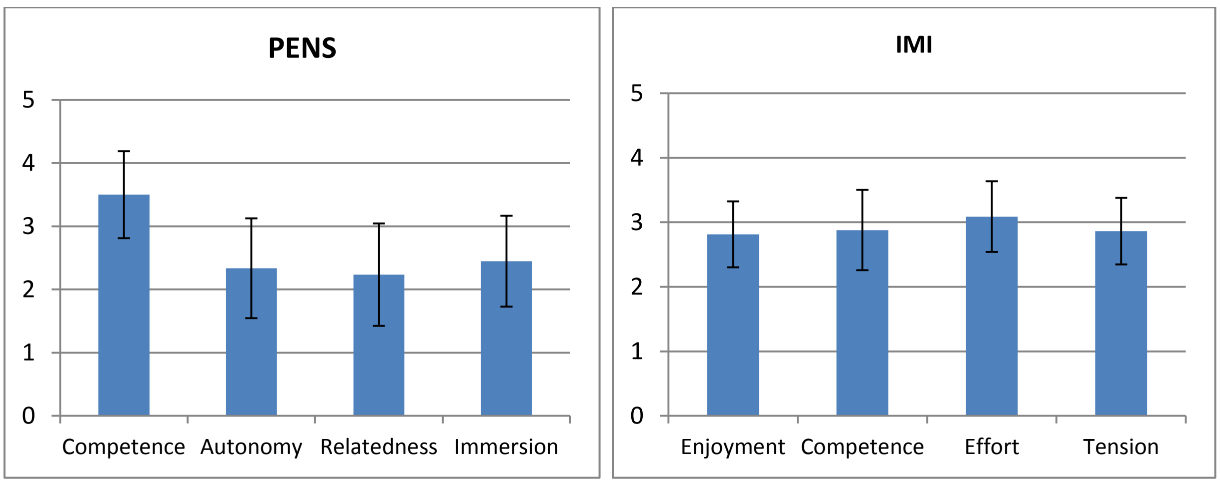

Figure 10. Average sub-scale scores for the PENS scale (left), Average sub-scale scores for the IMI scale (right).

CHAPTER 7

7 DISCUSSION

The results of our evaluation showed that data quality did not appear to be affected by our gamified framework for the motor task; however, crowdsourcing significantly reduced data quality for the cognitive task. Also, despite the overall usage of power-ups being low, those participants who were exposed to the power-ups in their gameplay used them often and, for the Lab Game group, players reported being open to the idea of more free-to-play games that include experimental tasks for in game rewards.

In this chapter we discuss the implications of our results on the design and use of our framework.

7.1 SEPARATION OF GAME AND TASKS

When designing our framework we focused on meeting requirements for flexibility and robustness, resulting in a design the separates and modularizes the tasks from the game. Because our framework separates out the game from the tasks, we are able to substitute any game for Sugar Rush and any task for the Fitts and McGill tasks that we chose. This allows us to maintain interest in the system through the release of new games and to ask any number of research questions – as long as the task can be deployed on a mobile phone.

7.1.1 Task Network and Delivery Platform

Mobile ad networks (such as AdMob[21]) consist of an application or website where advertisers can submit advertisements and specify an amount that they would be willing to pay developers in order for their advertisements to be integrated with mobile games. Mobile game developers then use the task network's application programming interface (API) to include advertisements within their game, and the task network application is in charge of collecting the fee from the advertisers and distributing it to the developers. This network is possible due to the separation between the game and the ads that are being displayed, only ever needing to interact with the ad network API to manage the display of advertisements.

Thanks to the modular approach that we took to our design and the separation between the tasks and the game with which they have been integrated, it would be possible to develop a task delivery network similar to current ad networks. A task manager API, similar to the ad network APIs, could be developed as a way to communicate between mobile games and the task network in order to easily add experimental tasks to developers' games without requiring prior cooperation and coordination between the researchers and developers like the direct integrated approach taken in this thesis. This could open up the possibility for research tasks to be included easily in multiple games at once increasing the exposure of the tasks, increasing the target market of participants, and increasing the sustainability of the experimental tasks.

7.1.1.1 Economic Comparison between Mobile Ads and Amazon's Mechanical Turk

The main benefit of crowdsourcing on Amazon's Mechanical Turk is that it provides researchers with the opportunity of gathering large amounts of data at a lower cost than laboratory-based studies.

In the previous section we propose the possibility of combining our framework with the model used by ad-networks as a way to develop a mobile-based crowdsourcing platform. Following the same task structure used in our laboratory-based control group, we can estimate and compare

[21] http://www.google.com.my/ads/admob/

how much it could cost to crowdsource our tasks via Amazon's Mechanical Turk with crowdsourcing via a mobile platform that follows an ad-network based model as described in the previous section.

In the control group, participants were asked to complete 7 blocks consisting of 14 tasks each: 5 McGill tasks, and 9 Fitts tasks. If this study were to be completed using an ad-based task network like the one described in the previous section, a participant would be given a single task in place of an ad view.

The cost of advertising in mobile apps can be determined by the following formula;

$$Cost = Impressions\ x\ eCPM$$ [22]

In the above formula, Impressions is the number of ads that are served and eCPM is the Effective Cost per Mille, or roughly how much it costs to pay the developer for every 1000 impressions served by their app[23]. The effective cost per mille can depend on factors such as what type of ads are being used and how effective the different types of ads are on different mobile networks. Since the research tasks that we utilized in our study were fullscreen interactive tasks, we can use the calculated eCPMS for interstitial mobile ads (fullscreen ads that interrupt the app just like our tasks).

Following the setup used in our control group, we can gather the cost to run a single participant using the following modified formula:

$$Cost\ per\ participant = (Tasks\ x\ Blocks) x\ (\frac{eCPM}{1000})$$

In 2014 the average eCPM for Android Interstitial ads recorded by MonetizePros[24] was between \$2.00 and \$4.00. Plugging the number of tasks, blocks, and the average eCPM into the formula

---

[22] http://appflood.com/blog/revenue-equation-for-mobile-app-ads
[23] https://gigaom.com/2011/02/06/best-practices-for-maximizing-mobile-app-revenue/
[24] http://monetizepros.com/cpm-rate-guide/mobile/

above we get a cost range of $0.20 and $0.40 per participant on our ad-based task delivery network.

On Amazon's Mechanical Turk we would post our entire study setup in a single job rather than breaking it up by tasks like we did above. The payout per task on AMT is set by the poster and can vary wildly based on the duration and effort required. Using the minimum reservation amount (minimum amount for which participants are willing to participate in tasks) of $1.38 per hour (Chilton et al., 2010) we can determine the minimum acceptable cost for our control study. The duration of our control study was between 15 – 20 minutes on average so at a minimum cost of $1.38 per hour, or $0.023 per minute, the cost of our task on AMT could range between $0.30 - $0.46. However, in our lab, we pay more ethical rates to AMT workers, ensuring a minimum of $6 per hour, and so our experiment would cost us around $2 per participant on AMT.

7.2 MOBILE CAPABILITIES AND POSSIBILITIES FOR CONTEXT-AWARE BEHAVIOURAL RESEARCH

Most mobile devices on the market contain numerous sensors and recording capabilities – Camera, GPS, accelerometer, Bluetooth, etc. – that can be used to gather contextual data. There has been significant research conducted that makes use of the data gathered by these sensors (Hashemian et al., 2012; Barrington et al., 2011). A mobile crowdsourcing platform could utilize the contextual sensors available on mobile devices to conduct context-aware behavioural studies, giving researchers insight into how people behave, think, and perform differently under different situations. This additional contextual information has not been readily available to researchers when performing laboratory-based studies or using crowdsourcing platforms that require participants to be seated at a computer, often in the privacy of their own home or work place.

7.3 VARIABLE REWARD STRUCTURE

Our current implementation associates all tasks with a single reward – players get one credit when completing any of the tasks. In implementations that contain multiple experiments with different difficulties or required effort, participation in the tasks may vary. To account for this or to motivate players to participate in more time-consuming or difficult tasks, a variable reward structure could be used. To do this, a standardized ranking could be developed within the task manager and used by the tasks and the game. The game would decide what type of rewards the individual ranks are to receive, and whenever a task is completed the task manager would notify the game which rank of reward is to be assigned. For example, since the cognitive task required more effort in our implementation, the framework could notify the game that this task should receive a higher reward and the game could respond accordingly, in our case by providing more credits for task completion.

7.4 SUPPORT FOR CITIZEN SCIENCE TASKS

In previous sections we mentioned the opportunities that mobile devices can provide for citizen science. Due to its mobile nature, our framework can offer these same opportunities. In addition, our framework utilizes a well-established distribution network by leveraging the Android Play Store, or other mobile game markets. Games delivered on these markets have the potential to reach a wide distribution of players, providing an opportunity for more widely distributed data collection for citizen science projects.

7.5 INCREASING TASK PARTICIPATION

In this section we discuss the effectiveness of the power-ups at motivating participation and introduce alternate design strategies for increasing task participation.

7.5.1 Increase Exposure to the Benefits of Power-ups

While the results showed that there was a small uptake in the number of players who used power-ups, those players who used them did so frequently. This would suggest that once players are exposed to the benefit in using the power-ups, they are more likely to want to use them more. There are design techniques that could be employed to increase the players' exposure to the benefits of the power-ups.

Games such as Candy Crush Saga[25] and Frozen: Free Fall[26] employ a technique whereby the player starts off with a limited number of power-ups of various types. This gives the player an opportunity to try them out in their game and experience the benefits the power-ups provide. The players also become used to having the power-ups to use in their gameplay, increasing the likelihood that they will pay, or in our case complete tasks, in order to obtain more. This same technique could be employed in Sugar Rush by giving the participants free uses of the power-ups for a short while, or by starting the player off with a number of credits so they can purchase the power-ups they wish to try out. Frozen: Free Fall also chooses to give players a free power-up each day they return to the game, encouraging repeated play and getting players used to having power-ups at their disposal so in the event they run out they will be more inclined to want more.

Another option is to include a tutorial session that walks the player through earning a credit by completing a task, and then using a power-up in a game session. Not only would this introduce players to the power-ups and highlight the benefits of using them in their game, but it would also introduce the player to the quick process of earning the credits, possibly assuaging any concerns or misconceptions they may have about what all is involved to receive a credit (for example, assuming that monetary payment is required instead of simply completing a task).

---

[25] play.google.com/store/apps/details?id=com.king.candycrushsaga
[26] play.google.com/store/apps/details?id=com.disney.frozensaga_goo

7.5.2 Non-Voluntary Task Integration

The current integration of our task framework makes the completion of tasks voluntary. Players are left to decide whether or not they participate in the tasks in order to receive game power-ups. We chose to keep participation voluntary in order to not deter interest in our game by forcing players to perform non-game related actions, and we also hoped that by choosing to participate in the tasks players would be more likely to put effort into their performance (Ryan & Deci, 2000). There are game designs and Freemium models that could be employed that would force the player to participate in tasks instead.

Some games employ mechanics that hinder the progress of the player until a requirement has been met. For example in some free-to-play games, players are required to watch an advertisement between game sessions. Other free-to-play games contain 'locked' content that will only become available to the players after they have purchased the full game or paid money to have them unlocked. Both of these designs could be utilized in order to increase participation in experimental tasks.

7.5.3 Alternate Freemium Models

The Freemium model of microtransactions elicits a similar pattern of participation as what we found with our power-ups– a small part of the player base pays for the monetary extras so the larger player base may consume the game for free (de la Iglesia & Gayo, 2009). Utilizing this particular Freemium model may have played a large part in the participation pattern that we saw in our results and our framework may benefit from exploring alternative Freemium models.

As mentioned elsewhere in this thesis, one possible model that could be considered is to deliver a task between each game session as is done with ads. As our current prototype game is meant to be played as a quick paced game with relatively short game sessions, this would mean quite a bit of participation but as it would be frequent and involuntary, it is possible that some players would be discouraged from playing the game.

7.6 DATA QUALITY

In this section, we discuss our findings regarding the quality of the task data, the implications they have on the usage of our framework, and suggest ways to improve performance.

7.6.1 Effect of Task Type on Data Quality

Analysis of our results showed that data received from the motor task was able to predict movement time with relative accuracy across each group. However, analysis of the data from the cognitive task was found to be significantly worse in the Crowdsourced Game group than the data received from either the Lab Game group or the Control group. One possible explanation for this difference could be the uncontrolled environment in which participants are not asked to put effort into the tasks by an experimenter. Rather, the participants are more concerned with simply completing the task in order to quickly earn their credit. Because this concern of earning the credit quickly is in line with the performance-based nature of the motor task, it does not have a negative effect on the quality of the results. However, because the cognitive task requires that participants take the time to make a judgment regarding the values they input, this goal of finishing as quickly as possible is detrimental to quality.

7.6.2 Effect of Game Type on Performance

In our study it was found that the performance-based motor task resulted in a higher level of data quality than the cognitive task. This may have been due to the goal of the performance-based task – tap the bars as quickly as you can - being in line with the goal of the player wanting to quickly earn a credit. Another possibility may be that the performance-based nature of the game - requiring quick motions and reflexes - is in line with the performance requirements of the task and place the players in the proper mindset to complete the task. This possible effect could be investigated in future studies and could possibly explain why the cognitive task performed so poorly with this game. It would be interesting to see if cognitive tasks would perform just as poorly when paired with a more cognitive-based game such as a puzzle game.

7.6.3 Rewarding Effort Instead of Participation

One way to motivate greater effort and accuracy by the participants in their responses or performance is to make the rewards contingent on the apparent effort put into the task. If the data that is entered is easily identifiable as being false data – such as the extremely large sequential numbers observed in the graph task or unrealistic response times – the credit reward can be diminished or withheld. Although this would only deter players from providing obviously false data, our results showed that when participants took a reasonable time to enter a value, the entered value itself was not unreasonable. This hints at the fact that if the participants invest time in providing the answer, they are more likely to provide a response that seems worth the effort. In addition, the value of the reward could be tied to the accuracy of the data, motivating participants to input quality data. Withholding the rewards for obvious lack of effort or tying the value of the reward to the quality of the answer or performance would encourage participants to spend more time in completing the task and likely result in better quality data.

7.6.4 Increase Microexperiment Length

The larger amount of data in the Control group – when compared with the game versions – is an artifact of our experimental design. However, we could increase the amount of data gathered in the game versions by increasing the number of trials that are included in a single instance of the microexperiments. In our current study, each task consisted of only a single trial - the cognitive graph task only asked the player to judge a single graph, and the motor tapping task only required the participant to complete the series of taps for a single ID. Increasing the number of trials included in a single task would increase the amount of data we collect, however care would need to be taken so that the duration of the microexperiments do not become so long that it deters participation. Future investigations would be required to determine the optimal duration of microexperimental tasks.

7.7 LIMITATIONS

We released our prototype game and experimental framework on the Android market and gathered data from two lab studies. We began an initial evaluation of our framework by isolating and investigating the factors of gamification and crowdsourcing on the quality of data gathered. To fully evaluate our framework, we will also need to isolate and investigate the effect that the mobile platform has on the quality of the results.

We chose to use the Freemium model of microtransactions in our prototype to integrate our research tasks within our game. We saw from our data that this resulted in task participation following a similar trend as in the business model where a small population of users is responsible for the majority of the data gathered. Future research using this platform should investigate different Freemium models and the effect they have on the amount of data gathered.

Another possible limitation to our evaluation is the design decision to choose a Freemium model in which participation is voluntary. We chose to make participation voluntary so as to not discourage players from choosing to play our game; however, if we had chosen a Freemium model that forced participation (such as displaying a task between each game session in place of an ad) we would have received more data per player.

In our evaluation, we asked participants to fill out questions from the IMI and PENS validated scales for player motivation and experience. Our results showed that the overall score for player experience showed no change from the median neutral value and that the overall score for player motivation was marginally lower than the median. Due to our methodological design, we are unable to distinguish whether these ratings were due to the game itself or if they were due to the inclusion of tasks within the game, and if the scores would be any different had the tasks not been integrated. Future work should try to isolate and determine the actual effect that the inclusion of tasks has on player motivation and experience by comparing survey results from having participants play the game in isolation and survey results from participants playing the game with tasks included.

CHAPTER 8

8 CONCLUSION

Gathering data on human behaviour is a limiting factor in behavioural research. Current techniques are time-consuming, costly, and subject to limited participant pools that often consist of students from the local university. These factors result in researchers making tradeoffs between the type and amount of questions they can study and the amount of resources they need to put into running the experiments. Crowdsourcing offers an attractive alternative as it provides access to a large participant base and has been shown to result in a quicker experiment developmental cycle. Some issues regarding crowdsourcing have been raised with regard to the low monetary incentive and data quality. Gamification is another alternative that is gaining in popularity. Research involving the gamification of work-based tasks has shown that adding game elements to studies can motivate participation and increase enjoyment. Though some research has concluded that gamification can increase participant effort and performance, the effect that game elements can have on the resulting data is largely still unknown. Also gamification requires that highly-personalized games be developed for the specific research questions being addressed.

8.1 SUMMARY OF RESEARCH

In this thesis, we presented a novel technique that combines both gamification and crowdsourcing techniques into a smartphone-based platform to motivate voluntary participation and provide researchers with a flexible framework that can be used to investigate multiple research questions without the need to develop costly specialized games. Results from our initial evaluation showed that the quality of the motor task data did not suffer; however, the data from the cognitive task was of lower quality. We feel that tying the reward to the quality of the data could improve data quality for attention-based tasks. Despite a low adoption of power-up usage, participants that were exposed to the experimental tasks were supportive of participating in return for in-game benefits.

8.2 FUTURE WORK

The concept of leveraging the Freemium model in mobile games as a technique to crowdsource research is new and has been previously unexplored. We showed that a framework designed around this concept is plausible and, as highlighted by our discussions in Chapter 7, has great potential for further experimentation and development. Here we highlight some of the more important and interesting future directions.

8.2.1 Full Factor Comparison

In Chapter 4 we introduced a matrix containing all of the different factors to be isolated in order to perform a full evaluation of our framework. In our initial evaluations, presented in this thesis, we chose to isolate and investigate the effect that the crowdsourced environment of our platform and the addition of a game environment had on the quality of data we gathered from our research tasks. Future research should complete the factor evaluation by isolating the effect that the mobile environment has on the gathered data. We intend on conducting a comparative study between the data gathered by our framework and data that is gathered from a non-mobile environment (i.e., desktop game in a laboratory setting, and a crowdsourced game on the AMT platform). In addition, it would be valuable to consider how our approach could be moved to a desktop environment to leverage the large computer game market.

8.2.2 Identification of Optimal Task Parameters

The game and prototype framework that was constructed for the work in this thesis was very preliminary and the decisions regarding task duration and task effort were not based on any previous research. Future research could look at investigating and trying to determine what the optimal task duration would be for this type of model and also what the optimal amount of effort is that players are willing to expend for the rewards that they are being offered.

8.2.3 Integration of Device Sensors for Contextual Behavioural Research

In Chapter 7 we discussed the use of mobile device sensors in research. Future work could investigate the integration of these contextual sensors with the data gathered from research tasks included within the framework. As has been outlined elsewhere in this thesis, there are recent research studies showing the importance of contextual data gathered from mobile sensors and how they can be utilized in research.

8.2.4 Effect of Game Type on Task Performance

In our study it was found that the performance-based motor task resulted in a higher level of data quality than the cognitive task. One explanation for this may have been due to the goal of the performance-based task being in line with the goal of the player wanting to quickly earn a credit and return to the game. However, it is also interesting to consider if it may have been that the performance-based nature of the game had placed the players in a particular mindset that was conducive to performing well on the performance-based task. It would be interesting to investigate whether the type of game that is used to frame the context of the tasks has an effect on the performance of the tasks that are included.

8.3 CONTRIBUTIONS

The main contributions made by this research are twofold.

First, leveraging the Freemium model within mobile games as a way to crowdsource research is a previously unexplored concept. In investigating this concept we designed a modular and flexible framework that separates the research tasks from the games with which they are deployed, allowing multiple unrelated research tasks to be included within a single game and the same task to be included in multiple games increasing exposure and the potential target audience.

Second, we developed a prototype framework and game to provide a proof of concept and evaluate the feasibility of our design on the Android Play Store. Using two classic experimental tasks, a Fitts motor task and a Cleveland and McGill cognitive task, our results showed that, while some design decisions needed to be taken into account, our framework was successfully used to crowdsource behavioural research tasks.

9 LIST OF REFERENCES

1. *Amazon's Mechanical Turk*. (Accessed 2014, January). Retrieved from Amazon's Mechanical Turk: https://www.mturk.com/mturk/
2. Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., & Werthimer, D. (2002). SETI@home: An Experiment in Public-Resource Computing. *Communcations of the ACM*, 56-61.
3. Balakrishnan, R. (2004). “Beating" Fitts' Law: virtual enhancements for pointing facilitation. IJHCS, 61(6), 857-874.
4. Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., . . . Huyck, C. (2011). Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics*, 680-687.
5. Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of Online Environments*.
6. Birk, M., & Mandryk, R. L. (2013). Control Your Game-Self: Effects of Controller Type on Enjoyment, Motivation, and Personality in Game. *CHI 2013*. Paris, France: ACM.
7. Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 3-5.
8. Capaldi, C. A. (2015). Graduating from undergrads: Are MTurk workers more attentive than undergraduate participants? Retrieved from https://osf.io/d2zxw/.
9. Cechanowicz, J., Gutwin, C., Brownell, B., & Goodfellow, L. (2013). Effects of Gamification on Participation and Data Quality in a Real-World Market Research Domain. *Gamification 2013*, (pp. 58-65). Stratford, Ontario.
10. Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2010). *Task search in a human computation market. In Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 1-9), New York: ACM.
11. Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Statistical Association, 79*, 531-554.
12. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., . . . Popvić, Z. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 756 - 760.
13. de la Iglesia, J. L., & Gayo, J. E. (2009). Doing Business by Selling Free Services. In *Web 2.0* (pp. 1-14).

14. Deci, E. L. (1975). *Intrinsic Motivation.* New York.

15. Decker, A., & Lawley, E. L. (2013). Life's a Game and the Game of Life: How Making a Game Out of it Can Change Student Behavior. *SIGCSE'13* (pp. 233-237). Denver, Colorado, USA: ACM.

16. Deterding, S., Bjork, S., Nacke, L. E., Dixon, D., & Lawley, E. (2013). Designing Gamification: Creating Gameful and Playful Experiences. *CHI 2013: Changing Perspectives* (pp. 3263-3266). Paris, France: ACM.

17. Deterding, S., Sicart, M., Nacke, L. E., O'Hara, K., & Dixon, D. (2011). Gamification: Using Game-Design Elements in Non-Gaming Contexts. *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425-2428). Vancouver, BC, Canada: ACM.

18. Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system?: Screening Mechanical Turk workers. *In Proc. CHI '10* (pp. 2399-2402). ACM.

19. Easley, D., & Ghosh, A. (2013). Incentives, Gamification, and Game Theory: An Economic Approach to Badge Design. *EC' 13* (pp. 359-376). Philadelphia, USA: ACM.

20. Eickhoff, C., Harris, C., & de Vries, A. P. (2012). Quality through Flow and Immersion: Gamifying Crowdsourced Relevance Assessments. *SIGIR'12* (pp. 871 - 880). ACM.

21. Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 189-200.

22. Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. J.Exp. Psych. ,47, 381-391.

23. Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., & Mandryk, R. L. (2011). Calibration Games: Making Calibration Tasks Enjoyable by Adding Motivating Game Elements. *UIST `11* (pp. 403-411). Santa Barabara, CA, USA: ACM.

24. Franklin, M. J., Kossmann, D., Kraska, T., Ramesh, S., & Xin, R. (2011). CrowdDB: Answering Queries with Crowdsourcing. *SIGMOD '11*, (pp. 61-72). Athens, Greece.

25. Freitas, J., Calado, A., Braga, D., Silva, P., & Dias, M. S. (2010). Crowd-sourcing platform for large-scale speech data collection. *FALA 2010*, (pp. 183-186).

26. Guy, I., Perer, A., Daniel, T., Greenshpan, O., & Turbahn, I. (2011). Guess Who? Enriching the Social Graph through a Crowdsourcing Game. *CHI 2011* (pp. 1373 - 1382). Vancouver, BC: ACM.

27. Hamari, J., & Eranti, V. (2011). Framework for Designing and Evaluating Game Achievements. *Proceedings of DiGRA*, (pp. 1-20).

28. Harris, C. G. (2011). You're Hired! An Examination of Crowdsourcing Incentive Models in Human Resource Tasks. *WSMD 2011, Workshop on Crowdsourcing for Search and Data Mining*, (pp. 15-18). Hong Kong, China.

29. Hashemian, M., Stanley, K. G., Knowles, D. L., Calver, J., & Osgood, N. D. (2012). Human Network Data Collection in the Wild: The Epidemiological Utility of Micro-contact and Location Data. *IHI '12* (pp. 255-264). Miami, Florida, USA: ACM.

30. Heer, J., & Bostock, M. (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Asses Visualization Design. *CHI 2010: Visualization* (pp. 203-212). Atlanta, GA, USA: ACM.

31. Howe, J. (2008). Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business.

32. *Intrinsic Motivation Inventory (IMI)*. (Accessed 2015). Retrieved from Self Determination Theory: http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/

33. Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., . . . Waldispühl, J. (2012). Phylo: A Citizen Science Approach for Improving Multiple Sequence Alignment. *PLoS ONE*, 1-9.

34. Kim, S., Robson, C., Zimmerman, T., Pierce, J., & Haber, E. M. (2011). Creek Watch: Pairing Usefulness and Usability for Successful Citizen Science. *CHI 2011* (pp. 2125-2134). Vancouver, BC, Canada: ACM.

35. Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsoucing User Studies With Mechanical Turk. *CHI 2008.* Florence, Italy: ACM.

36. Komarov, S., Reinecke, K., & Gajos, K. Z. (2013). Crowdsourcing Performance Evaluations of User Interfaces. *CHI 2013: Changing Perspectives.* Paris, France: ACM.

37. Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological Research Online: Opportunities and Challenges. *American Psychologist*, 105-117.

38. Lee, T., Quinn, M. S., & Duke, D. (2006). Citizen Science, Highways, and Wildlife: Using a Web-based GIS to Engage Citizens in Collecting Wildlife Information. *Ecology and Society*.

39. Lepper, M.P; Greene, D.; Nisbett, R. E (1973). "Undermining children's Intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis" (PDF). *Journal of personality and Social Psychology* **28** (1): 129–137.

40. MacKenzie, S. I. (1992). Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction, 7*, 91-139.

41. Mao, A., Kamar, E., Chen, Y., Horvitz, E., Schwamb, M. E., Lintott, C. J., & Smith, A. M. (2013). Volunteering Versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing. *Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing* (pp. 94-102). Association for the Advancement of Artificial Intelligence.

42. Mason, W., & Suri, S. (2012). Conducting Behavioural Research on Amazon's Mechanical Turk. *Behavior Research Methods*, 1-23.

43. Mason, W., & Watts, D. J. (2009). Financial Incentives and the "Performance of Crowds". *KDD-HCOMP '09* (pp. 77-85). Paris, France: ACM.

44. 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*, 48-58.

45. McCallum, S. (2012). Gamification and Serious Games for Personalized Health. *Stud Health Technol Inform*, (pp. 85-96).

46. Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Disassembling Gamification: The Effects of Points and Meaning on User Motivation in Performance. *CHI 2013: Changing Perspectives* (pp. 1137-1142). Paris, France: ACM.

47. Muntean, C. I. (2011). Raising engagement in e-learning through gamification. *The 6th International Conference on Virtual Learning*, (pp. 323-329).

48. Przybylski, A. K., Ryan, R. M., & Rigby, C. S. (2010). A Motivational Model of Video Game Engagement. *Review of General Psychology*, 154-166.

49. Reinecke, K., & Gajos, K. (n.d.). LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. *To appear in CSCW'15.*

50. Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., . . . Jacobs, D. (2012). Dynamic Changes in Motivation in Collaborative Citizen-Science Projects. *CSCW 2012* (pp. 217-226). Seattle, Washington, USA: ACM.

51. Ryan, R. M. (1982). Control and Information in the Interpersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology*, 450-461.

52. Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology 25*, 54-67.

53. Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motiv Emot*, 347-363.

54. Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to Study Clinical Populations. *Association for Psychological Science*, 213-220.

55. Vicencio-Moreira, R., Mandryk, R. L., & Gutwin, C. (2015). Now You Can Compete With Anyone: Balancing Players of Different Skill Levels in a First-Person Shooter Game. *CHI 2015* (pp. 2255-2264).

56. von Ahn, L., & Dabbish, L. (2004). Labeling Images with a Computer Game. *CHI 2004* (pp. 319-326). Vienna, Austria: ACM.

57. von Ahn, L., & Dabbish, L. (2008). Designing Games With A Purpose. *Communications of the ACM*, 58-67.

58. von Ahn, L., Ginosar, S., Kedia, M., Liu, R., & Blum, M. (2006). Improving Accessibility of the Web with a Computer Game. *CHI 2006* (pp. 79 - 82). ACM.

59. Vonikakis, V., Subramanian, R., Arnfred, J., & Winkler, S. (2014). Modeling Image Appeal Based on Crowd Preferences for Automated Person-Centric Collage Creation. *CrowdMM '14* (pp. 9-15). Orlando, FL, USA: ACM.

60. Watson, D., Hancock, M., & Mandryk, R. (2013). Gamifying Behaviour that Leads to Learning. *Gamification '13*, (pp. 87 - 90).

61. Watson, D., Hancock, M., Mandryk, R., & Birk, M. (2013). Deconstructing the Touch Experience. *ITS '13*, (pp. 199-208)

62. Yee, N. (2006). The Demographics, Motivations and Derived Experiences of Users of Massively Multiuser Online Graphical Environments. *PRESENCE: Teleoperators and Virtual Environments*, 309-329.

APPENDIX A

10 LAB NO-GAME GROUP CONSENT FORM

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Mobile Gamification for the Collection of Experimental Data**

Investigators: **Dr. Regan Mandryk**, Department of Computer Science (966-4888)
Kristen Dergousoff, Department of Computer Science

Purpose(s) and Objective(s) of the Research: The purpose of this project is to determine if mobile gamification is a viable method for crowdsourcing data collection.

Procedures:

- In this study, you will first be given a demographic questionnaire that asks you questions about your mobile game usage. Following the demographic survey you will be asked to perform a series of mobile-based experimental tasks.
- **Funded by:** The Natural Sciences and Engineering Research Council of Canada (NSERC).
- **Potential Risks and Benfits:** There are no known or anticipated risks to you by participating in this research. Your participation will help us to design new experimental techniques.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- **Storage of Data**
    - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
    - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

- **Right to Withdraw:**

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.

- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

- **Follow up:** To obtain results from the study, please contact Kristen Dergousoff (kristen.dergousoff@usask.ca).
- **Questions or Concerns:**

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Participant's signature:____

Date:______________________

Investigator's signature:____

Date:______________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Research Ethics Office at the University of Saskatchewan.

11 LAB GAME GROUP CONSENT FORM

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Mobile Gamification for the Collection of Experimental Data**

Investigators: **Dr. Regan Mandryk**, Department of Computer Science (966-4888)
Kristen Dergousoff, Department of Computer Science

Purpose(s) and Objective(s) of the Research: The purpose of this project is to determine if mobile gamification is a viable method for crowdsourcing data collection.

Procedures:

- In this study, you will be asked to play a semi-structured game that include some experimental tasks. Following gameplay and tasks, you will be asked to complete a questionnaire that asks you questions about your experience.
- **Funded by:** The Natural Sciences and Engineering Research Council of Canada (NSERC).
- **Potential Risks and Benfits:** There are no known or anticipated risks to you by participating in this research. Your participation will help us to design new experimental techniques.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- **Storage of Data**
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

- **Right to Withdraw:**

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

- **Follow up:** To obtain results from the study, please contact Kristen Dergousoff (kristen.dergousoff@usask.ca).
- **Questions or Concerns:**

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Participant's signature:____

Date:______________________

Investigator's signature:____

Date:______________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Research Ethics Office at the University of Saskatchewan.

12 CROWDSOURCED GAME GROUP CONSENT FORM

The Interaction Lab is a research lab at the University of Saskatchewan that deals with Human Computer Interaction research. Part of the research conducted at the lab revolves around many different aspects of games including, but not limited to, game balancing, health benefits, data gathering.

If you are interested in learning more about the Interaction Lab you can visit the lab website at *hci.usask.ca*.

This consent form is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve.

If you would like more detail about something mentioned here, please contact the researchers at *sugarrush@cs.usask.ca*.

Please take the time to read this form carefully and to understand any accompanying information.

This study is concerned with gathering data on standard psychophysics tasks in a mobile context. The goal of the research is to understand how we can employ gamification in a crowdsourcing context to gather data for standard psychophysics experiments by emulating the freemium model seen in many mobile games. Before each gameplay session you may earn in-game credits by completing tasks; these tasks will involve making visual judgements between two images or tapping on a series of targets. The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of the study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations.

If you would like to receive a copy of this summary, please contact the researchers at *sugarrush@cs.usask.ca*.

All personal and identifying data will be kept **confidential**. The informed consent form all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

Dr. Regan Mandryk, Assistant Professor, Dept. of Computer Science, (306) 966-4888, regan@cs.usask.ca.

By selecting 'Accept' you are indicating that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities.

APPENDIX B

13 DEMOGRAPHIC QUESTIONNAIRE

Sugar Rush

Please answer all questions to the best of your ability. At the end of the survey you will be given a code to input into the game and receive 25 credits.

Sex? Male ____ Female____

Age Range

a) 18 – 25
b) 26 – 35
c) 36 – 45
d) 46 – 55
e) Over 55

How often do you play games on any device? (e.g., computer, console, smartphone)

a) Every day
b) A few times a week
c) A few times a month
d) A few time a year
e) Never

How often do you play games on a portable device? (e.g., smartphone, iPod/tablet, Nintendo DS, Sony PSP/PS Vita)

a) Every day
b) A few times a week
c) A few times a month
d) A few time a year
e) Never

How often do you download new games for your smartphone or tablet?

a) Every day
b) A few times a week
c) A few times a month
d) A few time a year
e) Never

Who plays games on your portable device? (Check all that apply)

Me____ My friends____ My partner or spouse____

My kids____ My parents____ Strangers____

14 INCLUSION OF TASKS QUESTIONNAIRE

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
I would prefer to play Sugar Rush without the tasks					
I would pay $4.99 to play Sugar Rush without the tasks					
For a free game, I would rather complete tasks such as these than be forced to view advertisements					
Having to complete the tasks ruined the game play					
The tasks were easy to complete					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
I would rather not use powerups then have to complete the tasks to earn credits					
Having to complete the tasks did not negatively affect my play experience					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
I would prefer to complete multiple tasks at a time to earn lots of credits					
Playing Sugar Rush with the tasks made me feel like I was contributing to science					
I would play other games with embedded tasks if they were fun and free					
I played at least once a week since I installed Sugar Rush					

15 PLAYER EXPERIENCE QUESTIONNAIRE

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
I feel competent at Sugar Rush					
I feel very capable and effective when playing					
My ability to play Sugar Rush is well matched with the game's challenges					
When playing Sugar Rush, I feel transported to another time and place					
Exploring the game world feels like taking an actual trip to a new place					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
When moving through the game world I feel as if I am actually there					
Sugar Rush provides me with interesting options and choices					
Sugar Rush lets you do interesting things					
I experienced a lot of freedom in Sugar Rush					
I am not impacted emotionally by events in Sugar Rush					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
Sugar Rush was emotionally engaging					
I experience feelings as deeply in Sugar Rush as I have in real life					
I find the relationships I form in Sugar Rush fulfilling					
I find the relationships I form in Sugar Rush important					

I don't feel close to other players					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
When playing Sugar Rush I feel as if I was a part of the story					
When I accomplished something in Sugar Rush I experienced genuine pride					
I had reactions to events and characters in Sugar Rush as if they were real					
I enjoyed Sugar Rush very much					
Playing Sugar Rush was fun					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
I would describe Sugar Rush as very interesting					
While playing I was thinking about how much I enjoyed it					
Sugar Rush did not hold my attention					
I think I am pretty good at Sugar Rush					

I am satisfied with my performance at Sugar Rush					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
After playing Sugar Rush for a while, I felt pretty competent					
I am pretty skilled at Sugar Rush					
I couldn't play Sugar Rush very well					
I put a lot of effort into Sugar Rush					
It was important for me to do well at Sugar Rush					

Rate how much you agree with the following statements

	1 – Strongly disagree	2	3 – Neither disagree nor agree	4	5 – Strongly agree
I tried very hard while playing Sugar Rush					
I didn't try very hard at playing Sugar Rush					
I felt tense while playing Sugar Rush					

I felt pressured while playing Sugar Rush					
I was anxious while playing Sugar Rush					
I was very relaxed while playing Sugar Rush					