2017-08-01

Improving Gamification by Leveraging Endogenous Value

Brennan Laurence Smith
Brigham Young University

Improving Gamification by Leveraging Endogenous Value

Brennan Laurence Smith

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Seth Holladay, Chair
Paris Egbert
Kent Seamons

Department of Computer Science

Brigham Young University

ABSTRACT

Improving Gamification by Leveraging Endogenous Value

Brennan Laurence Smith
Department of Computer Science, BYU
Master of Science

"Gamification" is the application of game design principles to non-game contexts, such as education, personal fitness, etc. Gamification's intent is to incentivize unpalatable tasks. Current gamification efforts in the industry use some features traditionally associated with games, but fail to use game design principles as defined by the games industry. One such principle is endogenous rewards for task completion. We propose that endogenous rewards will increase the efficacy of gamification by increasing user engagement and retention. To demonstrate, we create a gamification framework where the rewards for completing real-life tasks are items with high endogenous value in the game, incentivizing the completion of those tasks. We then conduct an experiment in which we compare our framework with a commercial framework lacking endogenous rewards. Our analysis shows that it is likely that these endogenous rewards contribute to user engagement and/or retention in a gamification framework.

Keywords: Gamification, Game Design, Endogenous Value

ACKNOWLEDGMENTS

Thanks goes out to Dr. Holladay and Dr. Egbert, who advised me during this process. Also to Dr. Clement, who first suggested graduate school might be a good fit. Also, and most importantly, to my wife, who had the patience and love to endure another four years of schooling with me.

Table of Contents

List of Figures

Chapter 1

Introduction

Gamification is defined as the use of game design elements in non-game contexts [10]. Gamification is currently used in industry, education, and parenting to make unpalatable tasks more palatable, to incentivize desired behavior, to facilitate skill teaching and learning, and to foster a sense of engagement and interaction. For instance, schools may have a month-long board game in which a completed homework assignment allows a student's team to move forward a space, a company may have a leaderboard where sharing the company's advertisements on a social media site improves your rank, and parents may narrate their child's chores as though they were fantasy adventures, describing crushing soda cans as stomping monstrous spiders and taking out the trash as hauling bags of treasure into a cart.

Some companies, seeing the growth of the internet and the games industry, have sought to harness the power of games to promote user involvement and cement brand loyalty. Thus, while simple uses of gamification have been used by parents and teachers for centuries, only within the past two decades has gamification been a serious field of study.

Gamification attempts to solve a problem people face at some points in their lives: how can they get themselves (or their children, or their students, or their customers) to do something nobody wants to do? Games have proven power in this regard: People will spend countless hours in online games doing boring, repetitive tasks. Why, then, do those same people struggle to do boring, repetitive tasks in real life? Effective gamification solves this problem by injecting game-like elements into other areas of a person's life to bridge the motivational gap between games and life.

While early studies in gamification produced exciting results, industry interest in gamification has started to plateau and some studies have begun to dispute its effectiveness. Many believe that gamification is a "passing fad" that is on its way out, an over-hyped and under-effective advertising tool that is not worth the development cost.

This thesis endeavors to help differentiate between effective and ineffective gamification by focusing on one component of game design (endogenous value) and examining its effect on the gamification framework's user engagement and retention.

In the following chapter, we discuss prior research related to gamification. We will show that there is a problem with gamification as it is currently implemented in the industry. In Chapter 3, we will describe our proposed solution. In Chapter 4, we will describe an implementation of our proposed solution, along with a similar currently available gamification framework that does not contain our proposed solution. In Chapter 5, we will discuss the experiment we use to test our framework, using the competing framework as a control. In Chapter 6, we will describe the methods we use to validate our experiment, go over the results, and give conclusions. Chapter 7 contains unanswered questions and possibilities for future work.

Chapter 2

Prior Work

The structure and design of video games motivate many people to play them compulsively. For instance, games such as World of Warcraft have such a powerful motivating force on those that play them that the American Psychological Association is currently debating adding "Internet gaming disorder" to the Diagnostic and Statistical Manual of Mental Disorders as an official mental health diagnosis [19]. While many people believe games have power simply because they are fun, games are even able to motivate people to do things most people would not consider fun; Rigby and Ryan point out that World of Warcraft motivates its players to perform exacting mathematical analysis of game features so that they can optimize their play performance [20]. Industry writer Greg Costikyan admits that "By and large, you can expect that a player will respond to the incentives a game provides." [7]

To try to harness the motivational power of games in a positive way, a field known as "gamification" has emerged. The literature has accepted Deterting et al's definition of gamification, namely that gamification is "the use of game design elements in non-game contexts." [10] Gamification has since become a multi-million dollar industry and is still growing [6]. The current body of gamification research shows that it has great potential. The National Foundation for Education Research, via meta-analysis, found that "The evidence suggests that game-based learning can improve engagement and motivation" in public schools [17]. For example, Knutas et al. found that gamification in the context of a programming course increased collaboration between students [13]. Ibez et al. found that the gamification of a computer science class positively affected student engagement and moderately increased

learning outcomes [11]. Gamification is not just used in education: Since before the field was formalized, the health and fitness industry has found videogames to be useful in promoting exercise [3] and good dietary habits [2].

However, there are serious holes in the body of literature on gamification. Rojas et al. have pointed out that formalized trials regarding certain aspects of gamification are few and far between [22]. This is important, because there is disputation regarding what the foundational elements of gamification are. While ostensibly, gamification is the use of game design principles in non-game settings, Lister et al. have found that there is a difference between game design elements as espoused by the games industry and "gamification" elements espoused by the health and fitness industry [16]. Thus, the study of which elements are effective in motivating behavioral change is both important and absent from the literature.

Such research is especially crucial given that not all gamification frameworks are successful. Primack et al. have found that while many studies in the health and fitness industry show gamification to be effective, a sizable portion do not, leading him to trepidatiously state in his conclusion only that video games "may have potential for improving health" [18]. There are many specific examples of failed gamification frameworks. Robson reports the failure of the *My Marriott Hotels* framework, which was unable to increase recruiting to a significant degree [21]. When they attempted to incorporate gamification into their software engineering class, Berkling and Thomas found that the students found it to be an unnecessary hindrance, that it had no significant effect on learning outcomes, and that it reduced end-of-term teacher evaluations [5]. Korolov reports the failure of a corporate gamification framework used with salesmen, in which only 5% of the users "cared about being at the top of the leaderboard" [14].

In light of these failures, we must find the differences between effective and ineffective gamification frameworks. One starting point may be the difference between true game design elements as understood by the games industry and the elements that have been adopted by the purveyors of gamification (hereafter referred to as "gamification elements.") Game industry

writer Ian Bogost has repeatedly decried the differences between game design elements and gamification elements, and goes so far as to say, "Game designers and developers have resisted gamification largely because they perceive it to mistake incidental properties of their medium - points and leaderboards and the like - for the more complex and fundamental activity of designing and playing real-time simulations of complex systems. ...Overall, gamification's relationship to games isn't just one of exploitation, but also one of total and complete indifference." [6]. More rigorously, a study of gamification-related iPhone apps by Lister et al. has shown that while there is a correlation between game design elements and gamification elements, one does not necessarily imply the presence of the other. They also found that while there was a correlation between game design elements and app popularity, no such correlation existed between app popularity and gamification elements. They go so far as to say, "If games are in fact more popular to the public, then the focus of developers should be to create in-depth, narrative gaming experiences and not apps that merely use convenient elements of games or gamification." [16]. Thus, to more fully understand the difference between effective and ineffective gamification frameworks, we must narrow our focus to test individual aspects of gamification frameworks.

Chapter 3

Proposed Solution

One of the features of true game design [23] that is missing from many gamification frameworks is the concept of "endogenous value". Endogenous value has a different meaning in games than it does in psychology or economics; In our context, endogenous value is value that is created within the game through the interaction of its components and carries no worth outside the game itself [7]. For example, the money in Monopoly is considered to be valuable in the game because it allows one to buy properties, protects one from unexpected costs, and can be used to deprive other players of items they want in auctions. Monopoly money is thus highly desired by the players even though it has no value whatsoever outside the context of the game.

Computer games, being more complex than traditional games, have entire ecosystems of endogenous value, in which a series of desirable items entices players along a journey of progression. For example, in online role-playing games (MMORPGs) gold is valuable because it can buy equipment, equipment is valuable because it helps a player defeat monsters, defeating monsters is valuable because doing so grants access to new areas, and those new areas are valuable because in them, a player can obtain gold more quickly. In games such as these, nearly every action a player takes provides some element of endogenous value. An act as trivial as picking up a stick may prove to be of value, as a player might be able to sell the stick to a merchant looking for wood or perhaps use it as the haft for a weapon. Because there are so many ways for a player to "progress" and gain value, these games are able to keep players interested and engaged. Endogenous value is widely known in the games

industry as an important part of games [23]. Some writers, such as Greg Costikyan, feel it is so important that they include the concept of endogenous value in their definition of games: "A game is an interactive structure of endogenous meaning that requires players to struggle toward goals" [7]. However, this sort of value is missing from many gamification frameworks.

For instance, *Habitica* is a framework that allows its players to get gold for completing self-assigned chores. However, the gold is mostly used to buy cosmetic items for characters (with no mechanical effect) and to buy out-of-game rewards (such as trading 50 gold for a real-life soda).[1] These items' primary value is exogenous, occurring outside the game. A cosmetic item's value is created either when a player likes the way their character looks or when they are able to show off a rare cosmetic item to a friend who values it. Once purchased, the item is literally valueless in the game itself and only has any value due to external factors. Trading gold for a soda is another example of exogenous value; the soda has no value inside the game given that the game does not have any ability to track the soda's state.

While it may seem strange that endogenous value is preferred by game designers over exogenous value, there are a few good reasons why this may be the case. First, if the value of an item is within the ecosystem of a game, it can be incorporated as part of a positive feedback cycle, where gaining an Item A is valuable because it assists in gaining an item B, which is valuable because it assists in gaining even more of item A. This sort of self-perpetuating value is common in games, and even very simple feedback loops (such as those found in Cookie Clicker [9]) can make for compelling gameplay.

Another reason why endogenous value may be preferred is because it is subject to controlled inflation. If the game sets an expectation that a player can gain 50 of item A per hour spent and then subverts that expectation by providing 100 of item A under some conditions (such as being in a certain area, playing at a certain time, etc.) then players will feel excited about those conditions. Because the rewards are entirely virtual, such bonuses cost the game company nothing. Carefully controlled inflation of endogenous rewards can keep

a player coming back by ensuring that their rewards-per-time-spent ratio always increases. This would be prohibitively expensive with most exogenous reward systems.

Finally, exogenous rewards are often competitive in nature, meaning players must be rewarded unequally. If a designer makes rare cosmetic items that are only available after playing the game well for an extended period of time, people without those items will be looked down on. Any online leaderboard that serves as a ranking system must have people on the bottom for the people on top to feel elite. "Badges" or "trophies" are valuable only as much as some other players don't have them; an accomplishment shared by all has little perceived value. This leads to an ecosystem in which the majority of players receive little reward. By contrast, endogenous systems allow designers to create gameplay experiences in which the rewards are based on an absolute standard of performance rather than on relative performance. The designer can then set the absolute standards such that anybody is able to complete them and thus gain the rewards, which encourages them to continue playing.

While endogenous value is greatly preferred over exogenous value in the games industry, the majority of corporate gamification frameworks do not include items of endogenous value, opting instead for the "quick fix" of points systems, leaderboards, badges, and trophies. Such systems are easy to implement, but provide only exogenous value, and therefore may not be as effective as the games-industry approved practice of also including endogenous value.

3.1 Thesis Statement

A gamification framework that includes items of high endogenous value will have better user engagement and retention than a gamification framework that does not.

To test this hypothesis, we built a gamification framework that includes items of high endogenous value and tested it against a framework with similar features, but no elements of endogenous value. Descriptions of the frameworks follow. The experimental design is found in Chapter 5. Validation and other results are found in Chapter 6.

Chapter 4

The Gamification Frameworks

In our experiment we compared two frameworks. The first is our thesis implementation: a framework we created by branching from the RPGMaker MV engine by Kadokawa games. We hereafter refer to this framework as *Hero's Life*.

The second framework, which we used as a control, is a pre-existing on-line web app called *Chore Wars*, previously known as *World of Chorecraft*. It is a commercial gamification framework that makes money from selling "gold accounts" with heightened privileges, and represents what a commercially viable gamification framework looked like in the era in which it was developed. (Development seems to have ceased about three years ago, and began at least ten years ago)

These frameworks were chosen because they have similar elements, but differ in their use of endogenous value. (A full comparison is given in Section 4.4.)

Both frameworks belong to a category of games known as RPGs.

4.1 An Overview of RPG Conventions

In this thesis we will be using terminology common to the RPG genre. For the benefit of those who have not played many games in this genre, an overview of terms and conventions follows.

RPG is an acronym for "Role Playing Game." This doesn't imply that RPGs involve role-playing in the theatrical sense; rather, modern RPGs stem from tabletop role-playing games such as Dungeons and Dragons and have taken their nomenclature with them [4]. When

electronic gaming first became available, many games either tried to reproduce Dungeons and Dragons as faithfully as possible or to create similar mechanics accomplishing the same end. Thus, RPGs became the term used for any game in which a player tries to accumulate wealth and power by going on adventures and defeating monsters, even if the "role-playing" aspect of the game is essentially non-extant.

In these games, you have a character that acts as the player's avatar and is directly controllable by the player. This character has a number of attributes, which are typically numeric representations of his/her capabilities. Depending on the game in question, attributes may include strength, speed, magical aptitude, resilience, willpower, and other such personal qualities.

The most important attribute, and the one present in virtually all RPGs, is Health Points, or Hp for short. This attribute is a representation of how much injury a character is capable of sustaining, and is reduced every time a character is hit by an attack. This reduction is called damage. The amount of damage a character suffers from an attack depends on the character's other attributes, such as his/her resilience, according to a predetermined mathematical formula that varies depending on the game. When a character's Hp reaches zero, he/she suffers some incapacitative negative consequence, such as unconsciousness or death.

Defeating foes in combat grants two primary rewards: in-game currency and experience. In-game currency is traditionally measured in Gold Pieces, or Gp, which is the standard accepted monetary system of RPGs. Gp is usually spent on items such as equipment, which provide mechanical effects to the characters. For instance, armor may increase a character's resilience, thereby reducing damage from attacks. Items that cannot be equipped are generally used as either consumables that provide temporary bonuses, as crafting materials for other items, or as plot devices used to control the player's progression through the story, such as a key to a locked door.

Experience (shortened to Exp, or in this paper, Xp) is a measure of the character's increasing mastery of combat techniques. When a character reaches a certain amount of Xp, his level increases. Levels are a more discretized form of measuring a character's combat capability. A character's attributes tend to be calculated based on his level; as a character's level increases, all his/her attribute scores also increase.

RPGs have a complex and interdependent ecosystem of endogenous value. High attribute scores permit victory over stronger foes, which grants greater Xp and Gp rewards. This increase in Xp and Gp permits higher attributes, because the player's level increases, and the player can purchase equipment items, respectively. The cycle of defeating foes to gain power to defeat greater foes is iconic to the RPG genre, so much so that many people consider this cycle of endogenous value to be an RPG's defining feature.

While these are only general guidelines and each RPG differs in its implementation of (and adherence to) these concepts, the preceding overview serves as a starting point to understand our descriptions of the experimental and control frameworks that follow.

4.2 Chore Wars

In Chore Wars, characters have six attributes. When a new game is begun, the game asks the player what chores are the user's favorites based on a predefined list of sample chores. The game then assigns the starting attributes to the player's character based on that list. (For instance, "paying bills" increases the character's starting intelligence, and "taking out the trash" increases that character's strength.)

After creating a character by selecting a name, an icon, and the aforementioned list of chores, a character can choose to join a party, or a group of other players' characters.

If the character does not join such a group, Chore Wars allows the player to create a list of chores. When so doing, users are able to tag a chore as belonging to one of six categories, each of which represents a different attribute. A user is able to also choose an amount of Xp and Gp the chore awards upon completion. If a character does join a group,

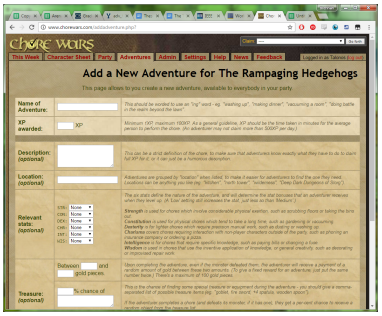

Figure 4.1: The Adventure Creation Dialog in Chore Wars

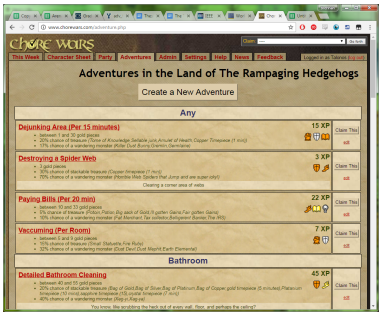

Figure 4.2: The Chore Completion Dialog in Chore Wars

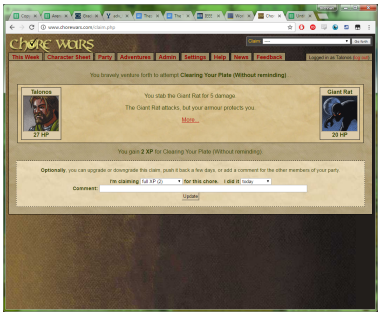

Figure 4.3: A Player Engaged in Combat in Chore Wars

the "Dungeon Master", which is a stylized name for the group administrator, is the one who creates the list of chores for all members of the party, using the same system as mentioned above. (This functionality can be used, for example, to allow a parent to create a party for his/her children, and set up the chores and rewards for the entire family at once.)

Chore Wars allows its players to self-report the completion of these chores. When so doing, the player gains the specified Gp and Xp reward, the attributes associated with the task increases, and all other attributes decrease slightly. The sum of the character's attributes remains the same, but doing tasks adjusts the ratio of those attributes in relation to each other.

Experience points cause a character to level up, as expected. The character's level does not seem to affect his or her statistics, but it is shown to other players in the player's party (providing exogenous but not endogenous rewards). An online leaderboard (consisting only of party members, therefore ensuring parity in task requirements) is present.

Gp cannot be used to purchase items in the game, which is a break from RPG tradition. Instead, Gp is used to purchase out-of-game rewards provided by either the users or the Dungeon Master. For instance, the Dungeon Master can manually deduct gold from a player's account to allow a child to "purchase" a toy (in real life) by "spending" in-game gold [8].

When a player logs completion of a task, two random events (whose probability can be defined by the creator of the task) can take place: the player's character may find a

user-defined item, and the player's character may encounter a monster. Both may happen at the same time, in which case the monster is presumed to be guarding the item.

If the player finds a monster, the player's character must fight it. This is accomplished by pressing the "more" button repeatedly. Each time the "more" button is pressed, the character attacks the monster and the monster attacks back. Each attack deals a highly variable amount of damage, which does not seem to scale with the character's level or be affected by the character's attributes. This lack of regard for a character's attributes is a striking break from RPG tradition, and prevents the attributes from having any mechanical effect on the game. In the end, players stand about an 80% chance of defeating any given monster, and the user has no control over the battle.

If the player defeats the monster (or if the monster was not present), the player receives an item. Unlike other RPGs, no items are equippable. By pressing a button in the player's inventory screen, a character can use an item, which removes it permanently from the character's inventory and sends a message to all players in the character's party that the item was used. The game documentation suggests that an item can be "used" for an out-of-game benefit, such as giving a "free drink" in real life to anybody whose character drinks a "fizzy potion" in the game [8]. Beyond this exogenous use, there is no use for items; they cannot be sold, equipped, reformed into other items, or used to advance any sort of progression.

Chore Wars was chosen as a control framework because on its surface, it shares many of the same characteristics as RPGs, but none of its components have endogenous value. This makes it ideal to test against; if endogenous value fuels user engagement and retention, as our thesis claims, then this framework should be revealed during testing to perform worse than a framework with endogenously valuable components.

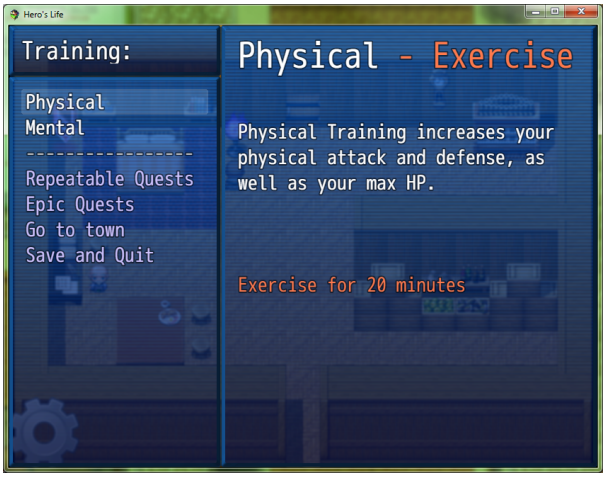

Figure 4.4: The Chore Completion Dialog in Hero's Life

Figure 4.5: The Equipment Purchase Dialog in Hero's Life

4.3 Hero's Life

Hero's life is a gamification framework we produced by branching the RPGMaker MV codebase [12] (descriptions of changes made are at the end of the chapter). When a player begins Hero's Life, they create a character by selecting a name and an icon. They are then given a brief 8-minute tutorial where they are taught how to engage in combat with foes (explained below) and how to report successful completion of chores. The tutorial helps players understand that their ability to defeat monsters is dependant on their chore completion by introducing a foe too powerful to defeat until the user completes a chore, after which the foe is defeatable. The game then instructs the user to create their own list of chores using a bundled editing program.

When creating chores, the user can select the Xp and Gp reward for the chore. Xp causes characters to level up, as normal to the genre, which provides a small increase to all attributes at once. Gp is used to purchase items which can be equipped to give further increases to specific attributes.

Some chores can be set to increase the character's attributes directly instead of increasing a character's Xp or Gp. These chores can only be completed once per day, and are intended to be used as "daily tasks" such as exercise. As many as seven such tasks can be

made, one for each attribute. If fewer than seven tasks are defined, some tasks will provide bonuses to multiple attributes at once. In the least case, a single task will boost all the character's attributes simultaneously upon completion.

The character, unlike in Chore Wars, is represented by a small icon that can move around an in-game world. The world consists of two areas: a town in which chores can be logged or items purchased, and a number of adventure sites. These sites are represented as maps that the avatar character can traverse, and the player is encouraged to get to the other side of the map to open a special blue chest on the other side. Along the way, the player fights monsters.

Combat in this game is more complex than in Chore Wars. Rather than clicking a "continue" button and watching a randomized, non-agentive combat take place, players instead are able to pick one of several actions from a list. The player's attributes are the most important factor in whether the player wins or loses, but a skillful player who selects optimal combat options for each situation can defeat a battle that an unskilled player could not. Not all monsters are of equal strength; some are more difficult than others, and defeating them requires higher attributes, better combat options, or a better strategy. The monsters become

Figure 4.6: A Player Navigating an Exploration Site in Hero's Life

Figure 4.7: A Player Engaged in Combat in Hero's Life

increasingly difficult the farther players progress in each adventure site, and each adventure site contains more powerful monsters than the previous.

Breaking from RPG tradition, defeating monsters grants neither Gp nor Xp, but a third currency called "skill powder" that is used to unlock new options in combat. Also breaking from RPG traditions, each exploration site only has a certain number of combats, which refresh at the beginning of each new real-life day. This prevents players from staying in an adventure site to get inordinate amounts of skill powder (Grinding.) Additionally, players may only enter one adventure site per day.

At the far end of each adventure site, in a special blue chest, is a magical item that permanently increases the gold gained by chores by +50%. These chests can each only be opened once per game per player. Opening such a chest unlocks an additional adventure site. There were four adventure sites in the framework during testing, which provided enough content that a player logging 1.5 hours of chores per day would require three weeks to exhaust the content. With more content developers, a framework like this could conceivably have as many as twenty to thirty such sites, providing about a year of content.

Players may play as a group if they desire. If multiple game states are saved on the same computer, players may load multiple games simultaneously, forming a party. While in a party, players may fight monsters together, meaning that when the players get a turn in combat, they may pick one action from each player's list instead of a single action, effectively multiplying the party's power (relative to a solitary player) by the number of players in the party. To counterbalance this advantage, each additional player beyond the first increases the number of monsters present by 50%. Thus, being in a party still provides a significant net advantage to the players while in adventure sites. Gameplay outside of adventure sites (logging chores, buying items, etc) is unchanged and can only be done while solitary.

Hero's life contains many endogenously valuable components. The items, Gp, Xp, skill powder, and blue chests all increase a player's ability to progress through game content. The root of all these items, and thus the root of all endogenous value, is chore completion.

Thus, the completion of chores becomes the most valuable action in the game. If our thesis holds true, then this framework will have greater success than the control framework.

4.4 A Comparison of the Control and Experimental Frameworks

The control framework has no items of endogenous value. Gold cannot be used to give the player an advantage in the game. Xp does nothing but increase one's place on a leaderboard, which is an exogenous, socially-based reward. Items have no mechanical effect within the system. Attributes seem not to affect combat in a meaningful way. No preparation or actions can affect a player's chances of winning a battle. Playing in a party provides no mechanical advantage in any way. Nothing obtainable in the game matters within the context of the game.

The experimental framework was designed to be similar, but it attaches high endogenous value to the gameplay elements. Gp is valuable because it allows you to buy items. Items are valuable because they increase your attributes. Xp is valuable because it also increases your attributes. Having high attributes is valuable because it allows you to defeat monsters. Being in a party is valuable because it increases your power more than it increases your opponents' power. Defeating monsters is valuable both because it gives the player skill powder (unlocking more combat actions), and because eventually, a player can get the special gem in the chest at the end of the exploration site. These gems are valuable because they inflate the gold value of completing chores, thus completing a self-reinforcing cycle. The blue chests are also valuable because they unlock additional adventure sites, where the monsters have more skill powder.

Defeating the increasingly difficult monsters requires higher attributes, which requires more Gp and/or Xp. The entire cycle is driven by Gp and Xp, but because monsters do not drop Gp and Xp as normal for this genre, the only way to obtain these resources is by doing chores. Thus, the entire ecosystem of endogenous value points players towards doing their chores, which becomes the basis of all endogenous value within the game.

Also of note is that the experimental framework provided no built-in support for exogenous rewards. Chore Wars' documentation actively encouraged the user to provide themselves exogenous benefits ("rewarding" themselves with sodas or other rewards). Chore Wars also provided a built-in mechanism to deduct items and/or Gp from a character to facilitate the conversion of in-game rewards to out-of-game rewards. Hero's Life had no such recommendations or functionality. All the rewards in Hero's Life are entirely virtual. This is valuable, because it allows us to test our theory that endogenous rewards are more indicative of user engagement than exogenous rewards.

The experimental and control frameworks are otherwise quite similar in structure. Both allow the user to create a list of self-defined tasks and have similar interfaces for logging chores completed. Both frameworks award Gp and Xp for completing tasks. Both frameworks involve combat against monsters. Both frameworks allow the character to gain items. Both frameworks allow players to form parties.

Due to the mechanical underpinnings of the games, there are some differences between the games unrelated to the endogenous value of their components. For instance, in Hero's Life, a more complex set of combat mechanics was necessary in order to encourage the purchase of equipment. Because the monsters must become stronger over time, a way was needed to showcase the player's progression, requiring multiple adventure sites. Given that there were multiple adventure sites, having the position of the player's character be represented in the game world was helpful to provide an interface by which a player could select an adventure site. This interface (in which the players moved around an avatar representing their character) was a logical fit to other additional interfaces required, such as going to a shop or completing a tutorial. Finally, the experimental framework used the stock artistic assets of the RPGMaker MV framework from which this project was branched, which are more copious and arguably of higher quality than the artistic assets of Chore Wars.

These differences are non-trivial and may have interfered with the results. If the experimental framework was higher quality in terms of graphics, combat mechanics, and

interface than Chore Wars, that may have impacted the way players felt about the game. On the other hand, if players found the experimental framework overwrought and over-complex, they may have avoided it. As the impact of the aforementioned differences is unclear, future researchers may improve upon this research by creating their own control framework or by trying to model the experimental framework to more closely resemble Chore Wars. However, given the importance that the games industry places on endogenous value, we believe our experimental results to be valid because the use of endogenous value is the most foundational difference between the two frameworks.

4.5 Changes Made to the RPGMaker MV Framework

While the RPGMaker Framework provided the majority of the code for the experimental framework, considerable researcher time went into modifying the code for our needs and developing content. In Javascript, we needed to create a new system by which players could create and store lists of chores and a way by which those chores could be loaded into an RPGMaker game. We needed to make a system by which players could claim rewards for completed chores. We needed to create a party system in which players could load multiple save games simultaneously. We needed to make a way for monsters to become more numerous dependent upon the number of players. In JSON notation, we needed to define the attributes and combat options for each of the twenty monster types in the game. We needed to create the maps of the town and adventure sites and write the tutorial and other dialogue. We needed to define the items a player could purchase, provide informative descriptions, and define their properties. In an XML-based scripting language called "Action Sequencing", we needed to script the effects that all combat options had on the combatants. This work process took five months of researcher time to complete.

Chapter 5

Experimental Design

Once the experimental framework was finished, we began an experiment to test its efficacy over the control framework.

After receiving IRB approval for our experiment, we went to undergraduate computer science classes in the 200 level or below and asked for volunteers to participate in our experiment. Those who expressed interest were given a flier with an e-mail address they could write to for further directions. Those writing to the e-mail address were given more information and a consent form. (A reproduction of this consent form is given in appendix A.) If the participant accepted the terms on the consent form, they were placed in either the control group or the experimental group (every second participant was placed in the control group; the rest were placed in the experimental group) and sent either a link to Chore Wars or a digital copy of Hero's Life.

Volunteers were asked to use the program as much or as little as they liked for three weeks. Each volunteer was given the opportunity to request help from a researcher if they ran into problems using the program. Three e-mails were sent to each volunteer: one around day three, one around day ten, and one around day 17. These e-mails reminded them of the existence of the framework and reminded them that they could leave the study at any time or request help at any time.

Those in the control group were instructed to assign one point of Xp to their tasks for each minute of work it was expected to take. Those in the experimental group had built into the chore-list creation tool a field for "minutes required" that asked them to specify

how many minutes of work they were reporting. Beyond that, no requirements were given to the participants concerning how they should structure their chore lists. Those in the control group were also not given any requirements or guidance on what external motivational items they could award themselves by expending in-game gold.

Participants were asked to take a survey either at the end of the experimental period or when they announced they wished to cease the experiment. The first question of this survey asked those in the experimental group to attach their save file and those in the control group to report how much Xp their character had accumulated.

The primary variable we measured was "minutes of work reported in the time period covered by the experiment." This metric is a combination of the amount of work done during the period and the subject's desire to continue using the program. We can derive this metric from the Xp of the control group, as the participants were instructed to give themselves one Xp per minute of work. We could measure reported work more directly in the experimental group, because we coded a function into the framework that added the "minutes required" field to a running tally every time the participants reported the completion of a task. These sums were included in the save file the participants sent, and we could manually extract this data.

Due to a smaller subject pool and a lower response rate than anticipated, a cash incentive of 10 USD was offered to all participants after obtaining approval from the IRB. This increased the number of responses we received by about 20%.

5.1 In Practice

The experimental design had a few flaws both in concept and in execution. First: the program logged work reported, not work completed. This means that the framework with endogenous value might have motivated participants to log work they may not have otherwise bothered to record, meaning it would appear that the experimental group did more work than the control group when they actually were just more motivated to record it. If such an effect

occurred, it would still imply greater user engagement with the program, which was a stated goal in our hypothesis.

Second, it was surprisingly difficult to convince people to fill out post-study surveys. Out of the 31 people who signed the consent form, 19 of them failed to return a survey, even though only 5 of those requested removal from the experiment. This could possibly lead to self-reporting biases. This abandonment of the survey also reduced our sample size to only six reporting participants per group, which made it difficult to extract statistical significance from the data. To counter this, we used an alternate form of significance testing involving Bayesian estimation instead of the standard t-test, as it proved better at extracting significance from small sample sizes.

Finally, as mentioned previously, five people requested removal from the experiment, disproportionately affecting the experimental group. This is probably because the experimental group had an additional step: they needed to install a program on their computer instead of using a web app. The disproportionate desertion of the experimental group over the control group may introduce a self-selection bias.

Chapter 6

Data and Validation

Our analysis indicated that the average minutes of work recorded by the experimental group was quite likely to be higher than the average minutes of work logged by the control group.

We used a standardly accepted Bayesian alternative to the t-test known as the "Monte-Carlo Markov Chain Bayesian Estimation that Supersedes the T-Test", or "BESTmcmc" [15]. This method takes the experimental group and the control group and models them as T distributions whose mean, standard deviation, and normality are represented not by concrete numbers, but by very wide probability distributions. It then uses Bayes rule to tighten said probability distributions by iterating over a Bayesian Markov chain. This method supersedes the classic t-test because it is able to do everything that the t-test can while providing additional capabilities. Prominent among these capabilities is the ability to provide a probability of a given outcome given the data rather than relying on rejecting a null hypothesis [15]. We therefore phrase our conclusions as a probability rather than as a p-value.

To conduct our Bayesian analysis, we used the BestMCMC package in R, an industry standard statistical computing tool. The results showed that despite our small sample size, our experimental group was 90.3% likely to have a higher mean number of minutes worked than the control group. This high likelihood implies with reasonable certainty that the endogenous framework encouraged participants to do and/or report more work.

Figure 6.1 shows the full probability distribution of the difference in reported mean minutes worked. All the weight to the right of the zero line represent the case in which the

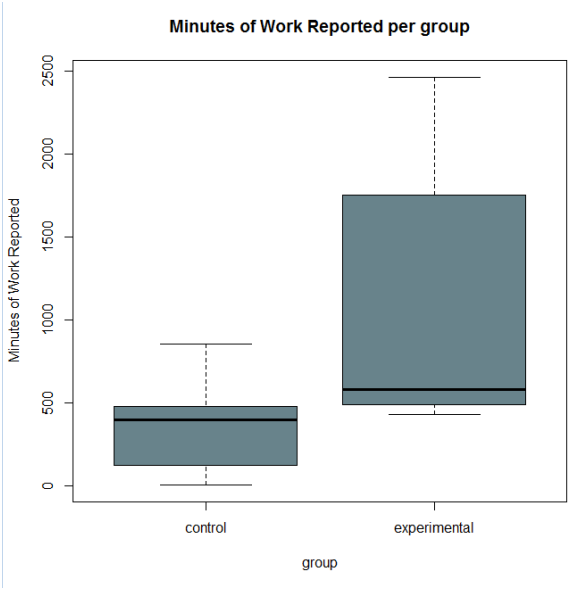

Figure 6.1: Box Plot of Reported Mean Minutes Worked over the Experimental Period

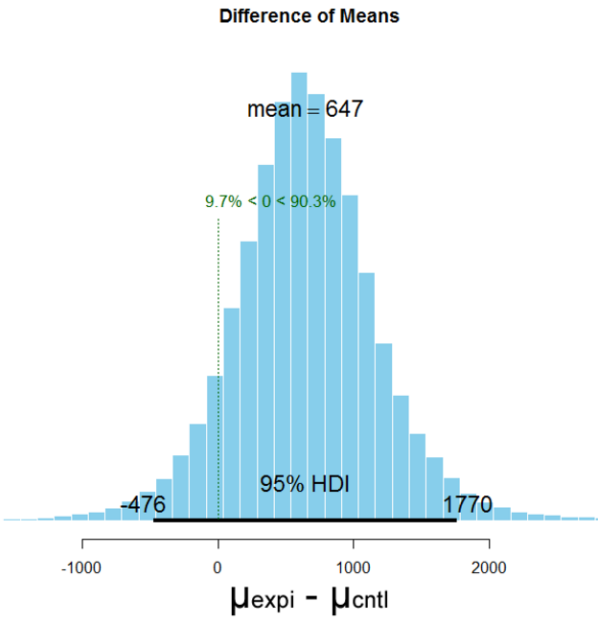

Figure 6.2: Difference in Reported Means

mean of the experimental group's source distribution is greater than the mean of the control group's source distribution, and the weight to the left represents the opposite. Because 90.3% of the probability lies to the right of the zero line, we can say that there is a 90.3% probability that the mean of the source distribution of the experimental group is higher than the mean of the source distribution of the control group.

Beyond our primary metric, the end-of-study survey included other metrics that we felt may prove relevant. We asked the respondents whether they agreed to given statements and coded their responses on a standard seven point Likert scale, with 1 being "Strongly disagree," 2 being "Disagree," 3 being "Somewhat disagree," 4 being "Neither agree nor disagree," 5 being "Somewhat agree," 6 being "Agree," and 7 being "Strongly agree." We then used Bayesian estimation, as above, to check the probability that a given hypothesis would turn out to be correct in a sufficiently large sample group.

The responses to each of the questions and the results of the subsequent Bayesian analysis are given in figure 6.3. The full descriptions of each question follow:

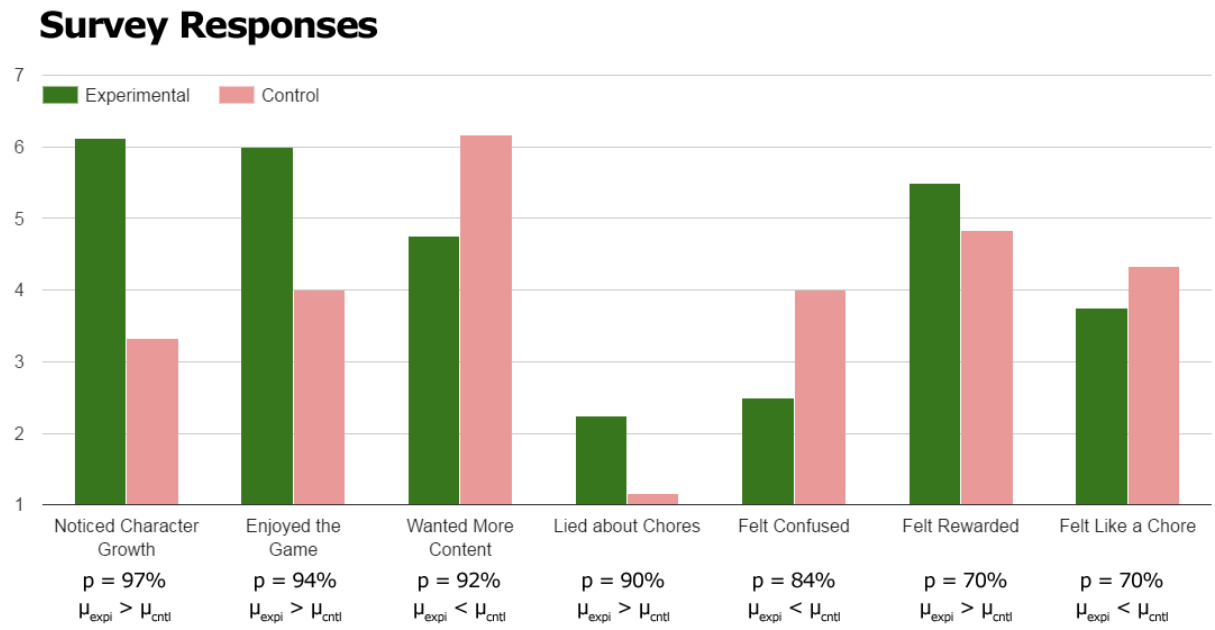

Figure 6.3: Survey Responses

When asked how much they agreed with the statement "I enjoyed the game portion of the framework (fighting monsters, etc.)", the experimental group signaled agreement far more strongly than the control group (a mean score of 6 compared to a mean score of 4). According to our Bayesian estimation analysis, it is 94.4% likely that the experimental group found the game-play portion of the framework more fun than the control group did (meaning the mean of the source distribution of the experimental group being higher than the mean of the source distribution of the control group).

Similarly, when asked how much they agreed with the statement "I noticed my character becoming more powerful when I did my chores," players in the experimental group expressed higher agreement than the control group (96.8% probability). Such is to be expected given that the chores in Chore Wars do nothing to increase the power of one's character.

Interestingly, when presented with the statement "I had no clue what I was doing," the control group seemed to agree more than the experimental group (84.4% probability). This is counterintuitive, because the experimental framework is a good deal more complex than the control framework. Optimistically, this could be because the experimental group was

more invested in the framework than the control group and therefore motivated to experiment with and master its features. Less optimistically, this could indicate a self-reporting bias, as those unable to comprehend the experimental framework may have left the study or refused to return a survey.

Also of potential concern, respondents in the experimental group signaled greater agreement with the statement "I often lied about work I had done in order to get the rewards" (89.7% probability). This may not be a problem, however; the modal response for the control group was "Strongly disagree" and the modal response for the experimental group was "disagree," indicating that both groups were mostly unwilling to lie when self-reporting chores. One comment from the free-response section of the experimental group may explain this: "Sometimes I would mark an extra activity I hadn't completed yet to get the next item or level and do the activity after, but I always did it." Participants may have marked "disagree" instead of "strongly disagree" because the endogenous rewards make it tempting to "pre-report" work that would be completed later in the day. Note that the participant was still encouraged to do work and eventually completed it, even after "lying" about it.

When presented with the statement "I wish this game had more content," those in the control group seemed starved for additional content (mean of 6.16) while the experimental group seemed less interested (mean of 4.75, 92.0% probability). If the control group wished for more content than the experimental group did, it would imply that the experimental group was more pleased with the content offered by the framework, assisting in the feeling of engagement.

The survey results suggest that players in the experimental group felt more rewarded when completing their tasks (70.4%) and less like tracking their work was a chore (61.4%) but the probabilities of differing means are not high enough to be conclusive.

Free response comments from both the control and experimental groups were encouraging. One person in the experimental group implied that the endogenous value motivated him: "It was nice being able to set my own tasks. I found myself wishing the game were a

traditional rpg (level up and gain money by fighting and exploring), but that would have defeated the point of the game." This comment shows that the player wanted to gain the endogenous value of leveling up and gaining money "for free" by playing the game, but recognized that the game was restricting these rewards contingent upon chore completion. The control group, on the other hand, seemed to lament the lack of endogenous rewards. One said, "I did not notice the 'playing' part of the game." Another said "...the game was very dry and minimalist for me." One actually requested endogenous content directly and asked for features already present in the experimental framework: "Adding some excitement like giving rewards to power-up a character that you control to fight monsters with could be more compelling for me."

Chapter 7

Conclusion

We created a framework with the express purpose of including as much endogenous value as we could and compared it against an industry standard framework that did not include endogenous value. We found that our framework incorporating endogenous value caused users to report a significantly higher amount of work over the experimental period. We also found our game was enjoyed more by the users.

Our research is valuable to the body of literature regarding gamification because it represents one of the few attempts made to determine what aspects of gamification lead to positive outcomes, when many studies instead try to evaluate the efficacy of gamification as a whole. With enough studies like this one that explore other facets of gamification, future gamification frameworks could be consistently designed to yield positive results by incorporating design features (such as endogenous value) that have been experimentally-proven to have a positive effect on user engagement.

Chapter 8

Future work

There are a number of further avenues of inquiry future research can pursue regarding the endogenous aspects of gamification.

Follow-up research could solve some of the unexpected problems that arose in this research, by (for example) making a control framework more similar to the experimental framework, working to increase the sample size, or breaking down some of the barriers to entry and streamlining the recruitment process.

A different experimental design could be used to try to isolate certain patterns of user behavior. For instance, instead of measuring minutes of work, one could measure "days the program was used" to see if endogenous value affects user retention, or shorten the trial period to one day to see if, in the absence of user retention, one of the competing frameworks results in more work done by the subject.

The experiment could be repeated with a different age group, such as children, or a different interest group, such as students recruited from American Heritage classes instead of intro-level computer science classes.

Instead of comparing a framework with endogenous rewards to a framework without endogenous rewards, one could compare it to a framework with standardized exogenous rewards. While we didn't specify what sort of out-of-game rewards the players in the control group could award themselves in this experiment, we could easily envision an experiment in which participants purchase candy from the researchers using in-game gold and test to see whether such an approach is more effective than the endogenous rewards within a game.

Research questions could more closely examine the user experience while playing the game. What items of endogenous value do players prefer to receive? Given any findings, would there be a more compelling gameplay experience upon which to set the foundation of endogenous value? Instead of fighting monsters, possible genres which could provide suitable ecosystems of endogenous value include base or city-building, puzzle-solving, or survival.

Finally, there are other aspects of game design besides endogenous value which could be examined. Other aspects of game design often not present in gamification frameworks include narrative and storytelling, three-dimensional environments and branching progression designs. Any of these may improve the effectiveness of gamification frameworks.

Appendices

.1 Consent Form

Consent to be a Research Subject

Introduction:
This research study is being conducted by Brennan Smith and Seth Holladay at Brigham Young University. The goal of the project is to improve understanding of effective gamification frameworks. "Gamification" is when one applies game design principles to an unrelated field to make the field more enticing. A "gamification framework" is a program or tool that makes a task more game-like. Prior and related studies have shown that properly designed gamification frameworks can improve a subjects motivation to complete tasks and can assist in behavioral change.

Procedures:

- You will be randomly placed into one of two groups, each of which will be testing a different gamification framework.
- You will be asked to either download a program from the internet or to create an account on a website that will let you access a web app.
- You will be asked to create a list of tasks that you wish to do but have a hard time convincing yourself to do. Examples of these tasks include homework, chores around the house, or religious study.
- After creating this list of tasks, you will be given the opportunity to use a gamification framework, which will give you in-game rewards when you report successful completion of these tasks.
- If the tutorials and instructions that come with the framework are insufficient, and you need additional help using the framework, you will be given the opportunity to meet with a member of the research staff one-on-one to receive further instruction on how to use the framework..
- You will be asked to continue using this tool for a period of three weeks.
- At the end of the three weeks, we will request that you send us *either* a file generated by the program *or* some information about your progress through the game. (Gold and XP collected, etc) Which information we ask for depends on which framework you tested.
- You will also be requested to describe your experience in using the tool by taking a survey. Your comments may be analyzed to assess the usability of the framework.

Risks/Discomforts:
There are no threatening or painful aspects to these experiments. The frameworks have elements of mild "fantasy" violence roughly on par with a Rated-E game. *If you find uncomfortable or offended by the content of the game, you can stop the experiment or excuse yourself from participation.*

Benefits:
There are no direct, proven benefits to this research, Gamification tools have been shown by other studies to help positively change behaviors in their users, but this is not guaranteed.

Confidentiality:
The research data will be kept on a password-protected computer and on a password protected Email account that will be deleted after the study is over. Only the researcher will have access to the data and passwords to the data. The only identifying information stored on the computer will be your name, which is the minimum necessary to store your consent statement. Your correspondence with us will be stored in the password-protected

Email and will include your Email Address and name. At the conclusion of the experiment, all data we gather will be anonymized. Three years after the study ends, all records that you agreed to this consent form will be deleted.

Compensation:

After the experiment is over, you will be allowed to keep the tool. If you found it useful, you may continue to use it for as long as you would like. In addition, participants who use the gamification framework on at least two days during the trial period, who provide the requested gameplay information (*either* a file generated by the program *or* some information about game progress, depending on the framework tested) and who complete the post-game survey will be provided $10 of cash compensation.

Participation:

Participation in this study is voluntary. *You may terminate your involvement at any time*. The experiment will last no more than 29 days. Participants must be 18 years of age or older.

Questions about the Research:

If you have any questions concerning this study you may contact one of the researchers Brennan L. Smith (435) 592-1947, Talonos@Gmail.com; or Dr. Seth Holladay, (801) 422-6490, seth_holladay@byu.edu.

Questions about Your Rights as a Research Participant:

If you have questions regarding your rights as a participant in a research project you may contact BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT 84602, (801) 422-1461,irb@byu.edu.

References

[1] How it works. `https://habitica.com/static/features`. Accessed: 2017-04-11.

[2] Tom Baranowski, Janice Baranowski, Karen W Cullen, Tara Marsh, Noemi Islam, Issa Zakeri, Lauren Honess-Morreale, et al. Squires quest!: Dietary outcome evaluation of a multimedia game. *American Journal of Preventive Medicine*, 24(1):52–61, 2003.

[3] Allison Barker. Study uses video games to fight obesity. *USA Today*, 2 April 2005.

[4] Matt Barton. *Dungeons and desktops: The history of computer role-playing games*. CRC Press, 2008.

[5] Kay Berkling and Christoph Thomas. Gamification of a software engineering course and a detailed analysis of the factors that lead to it's failure. In *Proceedings of the International Conference on Interactive Collaborative Learning (ICL)*, pages 525–530. IEEE, 2013.

[6] Ian Bogost. Gamification is bullshit. In Steffen P. Walz and Sebastian Deterding, editors, *The Gameful World: Approaches, Issues, Applications*, pages 65–80. MIT Press, 2015.

[7] Greg Costkyan. I have no words & i must design: toward a critical vocabulary for games. In *Proceedings of the Computer Games and Digital Cultures Conference, Finland*, pages 9–33, 2002.

[8] Kevan Davis. Chore wars help. `http://www.chorewars.com/help.php`, 2014. Accessed: 2017-04-11.

[9] Sebastian Deterding. Progress wars: Idle games and the demarcation of real games. In *Proceedings of the First International Joint Conference of DiGRA and FDG*, volume 13, pages 1–2, Dundee, Scotland, August 2016. Digital Games Research Association and Society for the Advancement of the Science of Digital Games. `http://www.digra.org/wp-content/uploads/digital-library/paper_267.pdf`, Accessed: 2017-05-19.

[10] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the Fifteenth Interna-*

tional Academic MindTrek Conference: Envisioning Future Media Environments, pages 9–15. ACM, 2011.

[11] María-Blanca Ibáñez, Angela Di-Serio, and Carlos Delgado-Kloos. Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on Learning Technologies*, 7(3):291–301, 2014.

[12] Kadokawa Games. Rpgmaker mv. `http://www.rpgmakerweb.com/`. Version: 1.3.4.

[13] Antti Knutas, Jouni Ikonen, Uolevi Nikula, and Jari Porras. Increasing collaborative communications in a programming course with gamification: a case study. In *Proceedings of the Fifteenth International Conference on Computer Systems and Technologies*, pages 370–377. ACM, 2014.

[14] Maria Korolov. Gamification of the enterprise. *Network World*, 9(2012):31–33, 2012.

[15] John K Kruschke. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2):573, 2013.

[16] Cameron Lister, Joshua H West, Ben Cannon, Tyler Sax, and David Brodegard. Just a fad? gamification in health and fitness apps. *JMIR Serious Games*, 2(2):e9, 2014.

[17] Carlo Perrotta, Gill Featherstone, Helen Aston, and Emily Houghton. Game-based learning: Latest evidence and future directions. *NFER Research Programme: Innovation in Education. Slough: NFER*, 2013. `http://www.nodo-observa.es/sites/default/files/GAME01.pdf`, Accessed: 2017-05-19.

[18] Brian A Primack, Mary V Carroll, Megan McNamara, Mary Lou Klem, Brandy King, Michael Rich, Chun W Chan, and Smita Nayak. Role of video games in improving health-related outcomes: a systematic review. *American Journal of Preventive Medicine*, 42(6):630–638, 2012.

[19] Andrew K Przybylski, Netta Weinstein, and Kou Murayama. Internet gaming disorder: investigating the clinical relevance of a new phenomenon. *American Journal of Psychiatry*, 174(3):230–236, 2016.

[20] Scott Rigby and Richard M Ryan. *Glued to games: How video games draw us in and hold us spellbound*. ABC-CLIO, 2011.

[21] Karen Robson, Kirk Plangger, Jan H Kietzmann, Ian McCarthy, and Leyland Pitt. Game on: Engaging customers and employees through gamification. *Business Horizons*, 59(1):29–36, 2016.

[22] David Rojas, Bill Kapralos, and Adam Dubrowski. The missing piece in the gamification puzzle. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, pages 135–138. ACM, 2013.

[23] Jesse Schell. *The Art of Game Design: A book of lenses*. CRC Press, 2014.

[24] Marko M Skoric, Linda Lay Ching Teo, and Rachel Lijie Neo. Children and video games: addiction, engagement, and scholastic achievement. *Cyberpsychology & Behavior*, 12(5): 567–572, 2009.