Critical Acclaim and Commercial Success in Mobile Free-to-Play Games

Kati Alha, Elina Koskinen, Janne Paavilainen, Juho Hamari
University of Tampere
FI-33014 Tampere
Finland
+358 40 190 4070
kati.alha@uta.fi, koskinen.m.elina@student.uta.fi, janne.paavilainen@uta.fi,
juho.hamari@uta.fi

ABSTRACT
Critical acclaim is considered to be one of the main predictors of profitability of game products. Major game publishers face tremendous hurdles in order to fare well in different forums that review and rate their products. However, little evidence exists on the relationship between ratings and profitability beyond anecdotal assumptions. In this study we investigate the relationship between critical acclaim and commercial success in mobile free-to-play games via a mixed-method study. First we look at the correlation of reviews and profitability, and then present an exploratory qualitative inquiry, analyzing games with high Metascores and games with high grossing. The results reveal that the relationship between review ratings and profitability is even more problematic in mobile free-to-play games than in many other game categories. Games with high Metascores differ substantially from the top-grossing games, being closer to traditional single-player games than typical free-to-play games, with little emphasis on monetization mechanics.

Keywords
Free-to-play, mobile games, Metascore, reviews, grossing, sales

INTRODUCTION
The game industry has grown significantly during the last years and continued to spread to various new platforms (Siwek 2014). To better inform the gaming audience in choosing games from an increasingly large selection, a large variety of different media outlets provide game reviews and ratings. Most of the online marketplaces, such as Steam, Amazon, and App Store have their own consumer rating systems, and services like Metacritic and GameRankings aggregate the increasing number of critic reviews into simple numbers.

Free-to-play (F2P) games, too, have been affected by the increased role of ratings. F2P games have gained unforeseen popularity and commercial success, and this has led to a "gold rush" towards converting traditional game business models into the F2P model (Alha et al. 2014; Brockmann et al. 2015; Hamari 2015). This inevitably means there are abundance of games of varying quality, and to help to find the best of the whole, reviews can help.

However, equating critical acclaim with the overall quality of a product and its commercial success is problematic. As witnessed in areas of other media content, even

Proceedings of 1st International Joint Conference of DiGRA and FDG

though review scores can in some cases influence or be used to predict the sales, the relations are not always straightforward (Dellarocas et al. 2007; Chen & Xie 2005; Bailey 2005). Despite this, reviews have gained an increasing position of power in game publishing (Nieborg & Sihvonen 2009; Totilo 2008; Rose 2012).

In this paper we investigate the relation between critical acclaim and commercial success in mobile F2P games. First we look into the correlation of critic reviews and grossing rank in iOS games in general and in F2P iOS games specifically. We continue with an exploratory qualitative inquiry into F2P iOS games by analyzing two groups of games: one with high Metascores and one with top-grossing ranks.

BACKGROUND

Ratings and Game Profitability

Although there is a dearth of academic research on the connection of review scores to consumer behavior in the area of games, the connection has been studied with other media content, especially in film studies. According to research, reviews do correlate with box office income in movies, although it is not clear whether they influence the sales or merely predict them (Eliashberg & Shugan 1997; Basuroy et al. 2003). These relations are not always straightforward and their appropriate use is challenging (Dellacoras et al. 2007; Chen & Xie 2005; Bailey 2005). In game products the correlation of Metascore and sales has been studied with Xbox 360 and PS3 games (Greenwood-Ericksen et al. 2013) and Steam games (Orland 2014), both studies finding a strong positive connection between Metascores and sales. According to a study by the video game research company EEDAR (2010), the reviews players read before playing strongly influence the way they themselves rate and value games, suggesting that the reviews would not only predict sales, but could also influence them.

At least when it comes to mobile games, however, critic reviews are not usually the first choice of information for the player when deciding which mobile game to play. According to EEDAR (2014), at least friend recommendations, consumer reviews, and top charts are more influential when choosing to download a mobile game. Game criticism is changing as the power from the press is spreading to consumer reviews and new forms of consumer criticism, which have partly replaced the role of critic reviews. Especially Let's Play videos have gained popularity, with most popular and influential streamers gaining billions of views (Martin 2015; Time 2015).

Despite the change, critic review scores remain in a critical role and are still used as the measure of critical acclaim. Especially Metacritic, one of the biggest game review aggregator sites, has gained increasing popularity and importance. Platforms like Steam and Amazon have included Metascores on their stores and Metacritic has gained power as a tool for measuring games. It has even been criticized for damaging the game industry, as publishers have been reported of giving out bonuses dependent on the Metascore instead of offering sales-based royalties (Totilo 2008; Rose 2012). Reaching a high Metascore has even been used as a job requirement in the industry (Matulef 2012).

Metacritic aggregates critic reviews of games and other entertainment products, and the aggregated reviews are combined into a numerical form called a Metascore. Having a Metascore requires a game to have a minimum of four critic reviews among all of the over 100 review publications Metacritic has selected as their review providers. Instead of simply extracting the means, Metacritic puts more weight on "some critics and

publications than others based on their quality and overall stature" but have not released the details of the procedure[1]. Metacritic excludes consumer reviews from Metascore and instead has a separate User Score, which is collected from Metacritic's users.

Free-to-Play

Regardless of the success of the F2P industry, only few games end up making it to the top while a vast amount of games receives only little or no revenue (Pinchefsky 2013). As F2P games have no entry cost and the revenue is mainly generated through selling in-game content and virtual goods, commercial success cannot be measured by sales or downloads of the game (Hamari 2015; Hamari & Keronen 2016; Lehdonvirta 2009). The profitability is measured by grossing, the total income of the game. F2P games dominate the top-grossing lists compared to premium games; at the time of writing, the top ten list of top-grossing iOS games comprised solely of F2P games. In addition, the top of this list is relatively stationary, as the same games tend to stay in the peak positions for years, while new additions rarely make it to the top or stay there. For instance, *Clash of Clans* (Supercell 2012) has placed in the top ten since shortly after its publication in 2012.

As the F2P model changes the way games are designed and played, it inevitably affects the game experience (Alha et al. 2014; Hamari 2015; Lin & Sun 2011; Paavilainen et al. 2013). The developers have to take the revenue model into account while designing the game as the purchases are designed as a part of the gameplay (Hamari & Lehdonvirta 2010; Hamari & Järvinen 2011). These real-money transactions are known as in-app purchases and can range from less than a dollar to hundreds of dollars. Bringing spending inside the game creates new challenges and new types of game experiences. The design of these games is usually strongly based on metrics, deducing from statistics which gameplay features or items are the most popular (Hamari & Järvinen 2011; Paavilainen et al. 2013). This way, however, some developers may go for faster revenue instead of trying to create better experiences and lasting interest.

F2P model has raised ethical issues, especially in cases where the spenders are underage or the game is producing its main income through "whales", players who spend substantial amounts of money on the game (Alha et al. 2014). F2P games have received a high amount of critique for weakening the game experience and even taking advantage of players. However, not all F2P games have a negative reputation. For instance, *League of Legends* (Riot Games 2009), *World of Tanks* (Wargaming 2010), and *Hearthstone* (Blizzard Entertainment 2014) have been positively received by critics. These are also games that are commercial successes. Surprisingly, this is not a common combination. For instance, the current leader of the top-grossing list of iOS, *Game of War - Fire Age* (Machine Zone 2013), has a Metascore of 67, having mixed or average reviews by Metacritic's standards. At the time of the writing, it had been reviewed only by four critics on Metacritic, which is quite a low number for a game that has enjoyed immense commercial success.

METHODS AND DATA: QUANTITATIVE PHASE

To examine the relation between critical and commercial success, we used both quantitative and qualitative methods to gather and analyze data. For the quantitative analysis, two datasets were collected and combined in August 2014. The first data was retrieved from Metacritic by including all iOS games with a Metascore. The list included 2596 games. The second dataset was collected from Sensor Tower, a service providing data about mobile games. We retrieved the top 1000 iPad games from the US top-grossing list.

These two datasets were then merged into one list to allow comparison between the two, resulting in a list with 3360 games. The list included games with different monetization models, which were sorted into four categories: F2P (free with in-app-purchases (IAP)), free (free, no IAP), paid (purchase cost, no IAP) and paid with IAP (purchase cost and IAP). Unfortunately, we had this information only from the games that were part of the Sensor Tower dataset, as Metacritic does not provide such information. We then used regression and correlation analyses to investigate the relationships between Metascore, grossing rank, and the business model.

RESULTS: QUANTITATIVE PHASE

The first observation when looking at the merged dataset was that only 236 of the games on the 1000 top-grossing games (23.6%) had a Metascore. It can therefore be concluded that from the games that make the most money, only a small portion ends up with reviews to begin with.

One of the explanations for this is derived from the next observations: as many as 736 (73.6%) of the top-grossing games are F2P (see Figure 1). When we look at games that utilize some other monetization model, 41.5% have enough reviews to have a Metascore, while for F2P games the percentage is only 18.2%. It therefore seems that mobile F2P games are relatively rarely reviewed by the press. F2P games also receive lower scores by the reviewers than games with other monetization models. The average Metascore for F2P games on the list was 71.8 (n = 158), while for games with other monetization models it was 79.8 (n = 156).

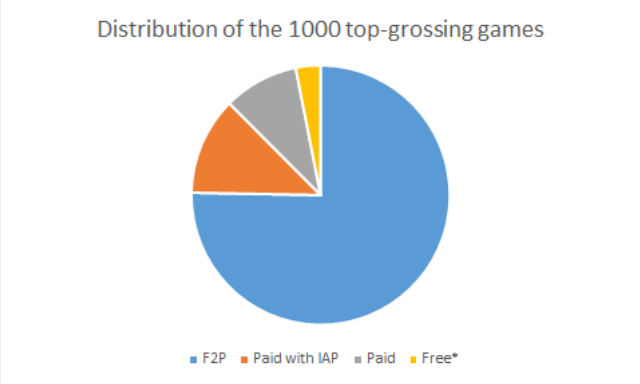

Figure 1: The distribution of the top-grossing games. *Free games are most likely paid games that have removed their price either temporally or permanently.

The scatter plot (Figure 2) shows the distribution of games along Metascore and grossing rank. While there are no remarkably clear visible patterns, the Pearson bivariate correlation test shows a statistically significantly negative correlation ($r = -0.169^{***}$, $p = 0.009$, $n = 236$), which suggests that those games that are ranked highly in the grossing list receive lower Metascores. These results indicate that games that receive better reviews do not necessarily make more money.

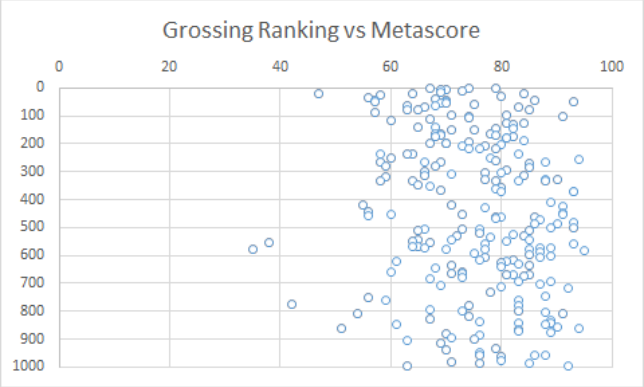

Figure 2: Distribution of games by their grossing rank and Metascore.

We conducted a regression analysis in order to investigate the effects of both Metascore and the business model on grossing rank simultaneously. The results reveal that the selected business model of the game may have more explanatory power on how much the game makes money. In terms of commercial success, it seems to matter more how the business model is integrated into the game rather than the actual quality of the game. A game being F2P predicts grossing rank (Beta = 0.338***, p = 0.000) whereas the effect of Metascore is lowered to small insignificant effect (Beta = -0.047, p = 0.472). The raw bivariate Pearson correlations between variables can be seen in Table 1.

	Metascore	**Grossing rank**	**Business model**
Metascore	1		
Grossing rank	-0.169***	1	
Business model	0.049**	0.203***	1

- * = p < 0.1, ** = p < 0.05, *** = p < 0.01
- Business model: 0 = not F2P, 1 = F2P
- Rank is reversed: higher rank, higher value

Table 1: Correlations.

METHODS AND DATA: QUALITATIVE PHASE

In game studies, playing games has been noted as an important method to understand games (e.g. Aarseth 2003; Mäyrä 2008; Karppi & Sotamaa 2012). Analytical play has been recognized as something different from leisure play as such a utilitarian approach requires the player to take notes and understand the wider context of gaming culture (Mäyrä 2008). Mäyrä (2008) distinguishes two approaches to analytical play: structural and thematic. While structural gameplay analysis focuses on rules, interactions, and games as artifacts with attributes and properties, thematic gameplay analysis highlights the symbols and messages that are experienced by the player as a cultural medium. Aarseth (2003) provides several strata to analytical play: superficial play, light play, partial completion, total completion, repeated play, expert play, and innovative play. For instance, the goal of partial completion is to reach certain goals or sub-goals in the game while total completion includes playing the game through.

Lankoski and Björk (2015) have introduced the formal analysis of gameplay, where game elements and their interactions are examined closely. Compared to the playing research approach, the formal analysis of gameplay can be seen as an isolationist approach

focusing on the game – not so much on the playing context. According to Lankoski and Björk (2015), this approach has been used by Myers (2010) to study the aesthetics of games while Björk and Holopainen (2005) have used it to develop game design patterns. In the formal analysis of gameplay, the game is understood through primitives (components, player actions, and goals) that form the game state. To understand gameplay, the researcher must play the game several times and build understanding by first recognizing the primitives and their relations, then the principles of design, and lastly the role of primitives and principles of design in the game.

Our gameplay analysis is based on approaches suggested by Aarseth (2003), Mäyrä (2008), and Lankoski and Björk (2015). The premise was to gain understanding of the gameplay through playing the games while taking notes, focusing on the formal aspects of gameplay. Partial completion is a sensible approach as the F2P games are usually never-ending and constantly updated.

We chose to examine five games with high Metascores and five top-grossing games more closely by playing them analytically. The lists used in the quantitative analysis were used to find games for both categories by selecting the games among the tops of the lists. We tried to select as representative spectrum of genres as possible. Genre refers here to the genres shown for the games in App Store. After the selection, some games were disqualified from the analysis as a closer examination revealed them not being actual F2P games, for instance in a case where the IAP included only unlocking the full game. New games were selected to replace these games. The final selection with their Metascores and grossing ranks are listed in Table 2.

	Game	Publisher	Published	Genre	Type	Metascore	Grossing
HIGH METASCORE	Hearthstone: Heroes of Warcraft	Blizzard Entertainment	2014	Card, strategy	Collectible Card Game	93	49
	Punch Quest	Rocketcat Games	2012	Action, arcade	Endless Runner	93	>1000
	Galaxy On Fire 2	Fishlabs	2010	Adventure, role-playing	Role-Playing Game	90	>1000
	Elf Defense Eng	Jellyoasis	2012	Board, strategy	Tower Defense	89	>1000
	Angry Birds Rio HD	Rovio Entertainment	2011	Arcade, puzzle	Physics Puzzle	88	332
TOP-GROSSING	Game of War - Fire Age	Machine Zone	2013	Role-playing, strategy	Combat Builder	67	3
	The Simpsons: Tapped Out	Electronic Arts	2012	Adventure, simulation	City Building	69	7
	Clash of Clans	Supercell	2012	Action, strategy	Combat Builder	74	1
	Candy Crush Saga	King	2011	Arcade, puzzle	Match-Three	79	2
	Hay Day	Supercell	2012	Family, simulation	Farm Simulation	TBD	4

Table 2: Analyzed games and their details as they were in August 2014.

As mobile F2P games have a special nature due to their monetization model and platform, no existing game analysis template was used. A specific template was created

by a group of researchers, based on the design literature of games (Fullerton et al. 2004; Schell 2008; Fields & Cotton 2012) and on studies focusing on F2P games (Hamari & Lehdonvirta 2010; Hamari & Järvinen 2011; Paavilainen et al. 2013). The template was tested and modified with two test rounds, both including several mobile F2P games.

The final template had eight categories with specific concentration points: 1) First-time experience, 2) Game mechanics, 3) Audiovisuals, 4) Narrative, 5) Sociability, 6) Monetization, 7) Playability and bugs, and 8) Returning to the game. In addition, a simple gaming log was kept by taking notes while playing.

Two researchers analyzed each of the selected games using the template. The high Metascore games and the top-grossing games were analyzed in turns to minimize the effect of time on the analysis process. Each game was played for a minimum of one hour, but as much time as necessary was used until the researchers were confident they understood how the game works and all the important aspects were covered. This usually varied from a couple of hours to a few weeks. After the analysis round was completed, the data was analyzed by the two researchers, first separately, then together. For each category from the template and for each game analyzed, the findings were discussed and compared to find the relevant issues to be taken under a more detailed inspection.

RESULTS: QUALITATIVE PHASE

First Experience

The first experiences in the high Metascore category varied substantially, most visibly in their tutorials. *Punch Quest* had no tutorial, only having the very simplistic instructions visible each time a new game was started. One could start the game right away and know what to do. *Hearthstone*, then again, had approximately a 30-minute long tutorial, which is by far the longest of all of the games analyzed. However, neither of the researchers found the tutorial tedious or prolonged, as it felt well-paced and well integrated to gameplay, feeling like playing the game properly.

In both examples – as in all cases in the high Metascore category – the beginning of the game and the tutorial or the lack of it felt purposeful and fitted the game. When the mechanics were simple, there was a very short and simple tutorial or no tutorial at all. When there was more complexity in the game, the tutorial was longer and more comprehensive. The player had freedom of choice in some parts of the tutorial and was not forced to follow instructions strictly. Each of these games could be without in-app purchases played as long as the player wanted to.

In the top-grossing category the tutorials were experienced as more limiting, using arrows as indicators where to click and having no control on what to do next. These types of tutorials were not experienced as that useful either, as the player usually clicked the indicated points without thinking what she was doing, and thus did not actually learn how to play the game. In *Game of War - Fire Age* the whole screen was darkened during the tutorial except for the point where you needed to click. This way, the clicking was, in a sense, missing its context. While in *Candy Crush Saga* the tutorial was as limited, it offered the option to skip it (see Figure 3). In some games, like in *The Simpsons: Tapped Out*, the game halted quite soon after the beginning, and the player had to either wait or turn to in-app purchases.

Figure 3: Tutorial phases in *Game of War - Fire Age* and *Candy Crush Saga*.

Game Mechanics

The games in the top-grossing group typically had a lot of simple mechanics, such as constructing buildings, growing crops, or ordering a character to carry out missions. Usually these actions then triggered a waiting time, which in turn resulted in rewards after completion. These types of mechanics were then repeated often, sometimes in cycles, creating the core loops of the games. For instance, in *Hay Day*, one would plant the seeds, wait for crops to grow, harvest the crops, create fodder, feed the animals, harvest the animal products, product foodstuffs, sell those on the market, and improve the farm with these resources so it could produce even more products. The core loops are not usually straightforward and might include several parallel, yet intertwining loops. For instance, in *Hay Day* the growing of crops could be followed by selling the crops, feeding them to animals, creating foodstuffs from them, harvesting them for more seeds, or completing specific missions that required the products. These games created their seeming complexity through a vast number of simple click-based mechanics rather than introducing more interesting mechanics. *The Simpsons: Tapped Out*, however, also kept the number of possible tasks that the player was able to do low, and usually the sessions only included collecting the cumulated resources and setting the characters to do the next tasks. These games lacked in difficulty, and the challenge was more in having the patience to slowly evolve the farm, city, or barracks – or use money to skip the waiting.

The feeling of progression was important in these games, but they executed it at a different pace. Where *The Simpsons: Tapped Out* progressed slowly and kept sessions short unless money was used, *Game of War - Fire Age* included a lot of small things to do, and although the overall progress was not that fast, the feel of progression was strong in every session. For instance, each of the buildings could not only be built, but also upgraded many times. At the same time, this brought more repetitiveness into the game.

These games usually included quests or missions to guide the progress of the player. In some games, these quests had to be done in a certain order to get the rewards, and if for instance the player had already constructed a certain building before the quest became

active, the player had to construct a second building of the same type to get the reward. *Game of War - Fire Age* was an exception, as the quests were more alike an achievement system, and even though the player might not have been aware of all of the quests, the player was rewarded if she, sometimes unintentionally, completed any of them.

Candy Crush Saga, as expected as a puzzle game, differed from these games. As a match-three game, the main mechanics were switching the places of candies to create candy combinations of three or more. As the game advanced, the game introduced additional gameplay elements and varied the game experience this way. *Candy Crush Saga* also offered quite challenging levels. The game started as relatively easy, so the player could play long enough to get hooked before the first difficult level, which tried to monetize the player into buying enhancements to help to pass the challenge.

The high Metascore games varied quite a lot in their game mechanics. *Punch Quest* was the simplest of the games, as jumping, hitting, and blocking were the only mechanics used during the actual gameplay – the character even moved on its own. The players played a similar, procedurally generated level over and over again, but kept evolving the character between the levels, therefore making it more probable to advance further on each round. In addition, the player could influence the progression by selecting the path when the level branched into different directions, giving a greater feeling of control to the game. On the other end of complexity, *Galaxy on Fire 2* included almost full-fledged role-playing game elements with story missions, side missions, talking to non-player characters, character and spaceship development, moving around the galaxy, mining resources, and battling enemies. All of these mechanics still stayed reasonably simple, and while the game had a lot of content for a mobile game, a lot of it remained similar.

Hearthstone, again as an exception, offered the feeling of deepest gameplay experience with rather simple mechanics. As a collectible card game, the game offered a starter set of cards, which the player could then use to play with other players online. More cards could be acquired by playing or buying with in-game currency or money, and building an optimal deck or decks became a part of the game. The existence of other players created more depth than AI-based opponents could have, and as the players were matched based on their performance in previous games, the opponents were usually well within the same skill level. Hearthstone has also included single-player campaigns in the game, which can be unlocked with in-game currency or money. They include challenges against different AI decks, and as a reward the player gains exclusive cards.

Audiovisuals

Both the high Metascore and top-grossing games had typically high-quality graphics and audio, although in some cases the audio loops were rather short. Many of the high Metascore games used simple graphics, such as *Punch Quest's* pixel graphics or *Elf Defense's* and *Angry Birds Rio's* cartoonish graphics. *Hearthstone's* cartoonish graphics were more detailed and had high production values. Attention had been paid to details for instance in the opening animation of a new card pack. When the player had no possible moves left, the game played a "Job's done!" audio clip imported from the *Warcraft III: Reign of Chaos* game (Blizzard Entertainment 2002). The graphics of *Galaxy on Fire 2* were more realistic, and although a bit outdated, it should be mentioned that the game has an HD version with updated graphics.

The Simpsons: Tapped Out stood out among the games as its graphics and animations and voice acting resembled the animation series. However, some portions of gameplay had no

music at all, and not all dialog was voice acted. As an exception, *Game of War - Fire Age* featured more medium-quality graphics, with some of the visuals pixelated or blurry, and the choice of fonts not feeling finalized. The audio did not stand out either, and the game used one notification sound for various occasions inside the game. The graphics mattered not only in the actual game, but also in the advertisements shown in the game, which influenced the attitudes towards the game. This was especially true when the advertisements were of poor quality or even suspicious. Although not directly connected to visuals, *Game of War - Fire Age* featured grammar errors, further negatively influencing the feel of quality of the game.

Narrative

The high Metascore games varied in their storytelling. *Hearthstone* and *Elf Defense* both had a story setting, which set the basis or motivation for the gameplay. The gameplay itself did not include many story elements. In *Punch Quest*, no reason for the game was explained when the game started. Only if the player stayed for a longer while in the menu, the game showed a short animation explaining the starting point of the game. As an exception in this group, *Galaxy of Fire 2* had both a setting and a story, which unraveled through playing and had an ending.

All the top-grossing games featured a background story with small story elements in the missions of the game. For instance, in *The Simpsons: Tapped Out*, the player was rebuilding Springfield after Homer had accidentally blown it up. This was used as the reason why player had to clear destroyed blocks and build new buildings and roads. The characters could be set to do missions, which somehow drew their inspiration from the animation series. All the stories in the top-grossing games were never-ending in a sense, as the games keep updating and adding content. In the high Metascore games, Hearthstone and Punch Quest are never-ending as well, but the story does not progress similarly.

Sociability

The top-grossing games included several social features. The more competitive games, *Clash of Clans* and *Game of War - Fire Age* were the most social, featuring guilds, chats and messaging systems, and making it possible to attack other players and wage war against other guilds. *Game of War - Fire Age* featured a global map, which had a location for each player on that server. Usually the players in a guild moved their locations next to each other, making it easier to both send group attacks and help to defend other guild members. After joining a guild, the members could easily ask help and assist others to construct buildings a bit faster. This feature lowered the threshold to help others, as it did not cost anything, requiring only a simple click. It was possible to donate the less valuable soft currencies to other players and both soft and hard currencies to the guild. *The Simpsons: Tapped Out* and *Hay Day* made it possible to visit other players and see how they were doing. *Candy Crush Saga* was the most single-player experience in the top-grossing group, but it too featured social elements such as rankings of other players and sending and receiving extra lives and moves.

The high Metascore games featured very few social aspects. Even though in *Hearthstone* the matches were played with other players, there was very little communication allowed between the players, as the only possibilities were a few pre-selected expressions. All the other games in the category were solely single-player games, from which only *Punch Quest* featured social media shares and high score lists.

Monetization

F2P games typically use two kinds of currencies, soft and hard. Soft currency can be gained easily through gameplay, while hard currency is more rarely rewarded, if at all. The game then sells the hard currency for real money. It is noteworthy that each of the top-grossing games taught how to use hard currency in their tutorial, while none of the high Metascore category games did this. The offline waiting time grew considerably long relatively early in the top-grossing games. For instance, in *The Simpsons: Tapped Out* the waiting times grew to 24 hours at the longest during the first 1.5 hours of play, and the sessions quickly became extremely short if hard currency was not used.

In all of the analyzed games the player could gain some kind of advantage by paying, including the directly competitive games. However, all the games that had at least one in-game currency also had the possibility to earn that currency by playing, and therefore it was theoretically possible to gain everything in the game without using any real money. In most games, however, gaining everything without paying would have been impossible simply due to the sheer amount of time it would have taken.

The only games that had exclusive purchases with real money were in the high Metascore category. These included new episodes in *Galaxy on Fire 2*, permanently doubling the earning of the in-game currency in *Punch Quest*, and an item in *Angry Birds Rio* that would then allow infinite amounts of uses to skip levels. All of these exclusive purchases were permanent items, and all consumable items could be acquired by playing. *Punch Quest* featured an interesting "donate a buck" feature, where the player could give a bit of money for the developers without gaining any advantage or content for the game.

The categories differed quite clearly in in-game currencies. Most games with high Metascore had only one currency, and this currency could usually be earned in abundance within the game, but could also be bought with real money. *Hearthstone* was an exception, as its single currency resembled hard currency by being slowly earned through gameplay, but more gold could not be bought. However, everything that could be bought with money could be bought with gold, giving gold monetary value. All the top-grossing games used at least two in-game currencies: one hard currency, which was earned only in small amounts by playing, and one or more soft currencies, which could be gathered easily by playing. For instance, *Game of War - Fire Age* had gold as the hard currency and several soft currencies with different values: silver, stone, wood, ore, and food.

Interestingly, none of the games in the high Metascore group had offline progress mechanics, such as waiting times for buildings to finish or energy to gather, while all the games in the top-grossing category used some version of these mechanics. The player had to wait for something put in motion to complete, or as in *Candy Crush Saga*, wait for lost lives to regenerate. In all the games in the top-grossing category, the waiting could be skipped with hard currency or money. Only *Candy Crush Saga* included an option to ask help from friends as an alternative way to skip the waiting.

Some of the top-grossing games felt aggressive in their monetization. In *The Simpsons: Tapped Out* aggressive monetization emerged through limited times for purchasing specific items, and forcing to use the scarce hard currency during the tutorial. In *Game of War - Fire Age* the user interface was crammed with different limited time offers that repeated themselves and other buttons that led to the shop. Sometimes a limited offer ad popped up and filled the whole screen. In addition, naming the hero, the city, or the player herself, choosing a profile picture or changing the hero's avatar were all items that

could be bought and used. This meant that if the player wanted to change any of these after using the item given in the beginning of the game, she had to buy a new one.

However, when the game notified in the beginning that it involved in-app purchases as in *Clash of Clans*, the game already felt a bit fairer. In *Hay Day*, the purchases that cost hard currency needed a double click, preventing the player spending hard currency unintentionally, again making the game feel fairer and more trustworthy.

Playability and Bugs

All the analyzed games had at least some problems with playability, but typically these problems were scarce and minor. The top-grossing and the high Metascore games did not differ substantially as groups, and the differences were more distinct between individual games. *Game of War - Fire Age* had several problems due to a crowded user interface which caused miss-clicking and user interface elements blocking other elements or game view. The game had to be reloaded every time the iPad went into a screensaver mode, and as reloading took relatively long, this hindered the experience.

Galaxy on Fire 2 had more problems as well, mostly caused by the complicated nature of the game, making it hard to find the right menus or know what to do in the game. The icons on the user interface were not self-evident, and the player had to learn where they directed. While most of the games did have some bugs in them, usually connection failures or freezing, there were no apparent differences between the analyzed games.

Returning to the Game

Interestingly, none of the high Metascore games used push notifications. Push notifications are notifications that appear on the screen of the device when the game is not running, and usually let the player know that something has happened or been completed in the game. However, there are other ways to lure the player back. *Hearthstone* included tournaments during which the player had to be active to fare well. *Punch Quest's* Facebook posts might remind a friend of the player to play again. All in all, however, these games were quite passive in persuading the player to return.

On the opposite, all the games in the top-grossing categories had push notifications. Usually the notifications included sounds by default, and especially *The Simpsons: Tapped Out* had quite loud and long notification sounds. The frequency of the notifications varied from one game to another, and was usually highest soon after the game sessions. When the player had not played for a longer while, the notifications stopped or became scarce, notifying mostly of special events happening in the game.

For the researchers, the best way to lure the player back was not the push notifications, but the gameplay: when it was possible to keep the sessions short but still progress in each session, the visits in the game were most frequent. The push notifications worked best when their frequency was not too high, and notified about something essential, while too high frequency led into ignoring most of the notifications. In some games, such as *Game of War - Fire Age*, the strong sociability hooked the player and created commitment.

Summary: Comparison

The high Metascore and top-grossing games differed in most of the analysis categories. Table 3 presents a summary of some of the properties of the games, visualizing the distinct differences between the analyzed groups.

	Hearth-stone	Punch Quest	Galaxy on Fire 2	Elf Defense	Angry Birds Rio	Game of War - Fire Age	The Simpsons : Tapped Out	Clash of Clans	Candy Crush Saga	Hay Day
Teaches to use hard currency	No	No	No	No	No	No	Yes	Yes	N/A	Yes
Currency	Single	Single	Single	Single	No	Several	Double	Several	Single	Double
Everything achievable without money	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Offline progress mechanics	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Push notifications	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of social mechanics	Low	Low	None	None	None	High	Low	High	Low	Low
Direct competition	Yes	No	No	No	No	Yes	No	Yes	No	No

Table 3: Comparing game properties.

DISCUSSION

Whereas a strong positive correlation has been found between higher Metascores and better sales with console (Greenwood-Ericksen et al. 2013) and Steam games (Orland 2014), with mobile games the raw correlation between review scores and revenue was found negative, implying that the games higher on the grossing list receive lower Metascores on average. When we controlled the effect of the business model, we found that instead of the review scores the revenue model explains higher or lower revenue, with F2P games bringing more income. Mobile F2P games received lower critic scores on average, but dominated the top ranks of the grossing list.

One reason for this could be the current game press being conservative and not yet accustomed to the new F2P game type. Bringing paying inside the game might feel wrong for the generation that is used to gameplay that concentrates on creating the best possible game experience, free of asking for money (Alha et al. 2014). Mobile F2P games were rarely reviewed to begin with, which can be at least partly explained by how review copies and codes are sent out to game magazines to acquire the game for free for review purposes. As F2P games have no purchase cost, such codes are not needed, and the published F2P games might sometimes left without reviews purely because of this. This is sometimes remedied by sending codes with some virtual currency or other purchasable content for the game for the reviewer to test out.

The game analysis revealed that the high Metascore games were in many ways distinctive from the top-grossing games. Even when considering the fact that the analysis included only ten games, the differences are clear as visualized in Table 3. The top-grossing games were more active in both monetization and retention, both of which are important factors for a F2P game (Luton 2013). The games guided and encouraged to use hard currency and real money and even required it if the player did not want to wait between the game sessions. All of the top-grossing games were active in sending push notifications, trying to keep the player's interest on the game and get her to always return for the next game session. The higher number of social mechanics is also a factor that can strengthen retention, as the social pressure and will to keep playing with friends can be a strong motivation to return in the game. A good group of friends can even make up for the poor game design, stressing the importance of sociability (Paavilainen et al. 2013). Strong

sociability is also a good incentive for acquisition – another important factor for F2P games (Luton 2013) – as when players want to play with their friends for instance in the same clan, they might try to persuade them to start playing to join in.

The high Metascore games seem close to traditional single-player games, which further stresses why the game press might be more favorable to these games. On the other hand, high Metascore games are not aggressive in their monetization, which might explain their low placement in the grossing charts on its part. In a way, the high Metascore games might even be too good to motivate players into spending money. This is supported by earlier findings where enjoyment has decreased purchase intentions (Hamari 2015).

Naturally the division between critically acclaimed and commercially successful games is not exact. For instance, *Hearthstone* builds a bridge between these categories by having both a high Metascore and being relatively high on the grossing list. This is an interesting case, and these types of games could teach us more about the special nature of mobile F2P games and would need further research.

There are limitations to this study, which should be taken into account. First, as the game analysis covered ten cases, it does not allow broad generalizations. Second, game analysis is inevitably a subjective process. This was partly remedied by two researchers analyzing each game. Third, a game's critical acclaim and especially commercial success are not dependent only on the in-game experience, as for instance marketing is a crucial part of the bigger picture. Despite these limitations, the results and their implications are interesting and worthy of further examination.

CONCLUSIONS

In this paper, we examined the relation between critical acclaim and commercial success in mobile F2P games. We used both quantitative and qualitative methods and data, and found clear distinctions between these two ways to measure a game's success. The correlation between mobile game reviews and their position on the top-grossing chart was negative. F2P games were reviewed with lower scores on average than games with other monetization models, but on the other hand, F2P games were significantly more successful commercially. From the analyzed games, the highly reviewed games differed substantially from the top-grossing games by resembling more traditional games than typical F2P games, while the top-grossing games featured more monetization, acquisition, and retention mechanics. The combination of critical acclaim and commercial success is relatively rare in mobile F2P games, and more research should be conducted with the games that reach both of these goals.

ENDNOTES

1 http://www.metacritic.com/about-metascores

BIBLIOGRAPHY

Aarseth, E. (2003) "Playing research: Methodological approaches to game analysis," Paper presented at DAC 2003. Melbourne, Australia.

Alha, K., Koskinen, E., Paavilainen, J., Hamari, J. & Kinnunen, J. (2014) "Free-to-play games: Professionals' perspectives," in Proceedings of Nordic Digra.

Bailey, A. A. (2005) "Consumer awareness and use of product review websites," in Journal of Interactive Advertising, 6(1), pp. 68-81.

Basuroy, S., Chatterjee, S. & Ravid, S.A. (2003) "How critical are critical reviews? The box office effects of film critics, star power, and budgets," in Journal of Marketing, 67(4), pp. 103-117.

Blizzard Entertainment (2002) *Warcraft III: Reign of Chaos* [PC Computer] Blizzard Entertainment.

Blizzard Entertainment (2014) *Hearthstone* [iOS] Blizzard Entertainment.

Brockmann T., Stieglitz S. & Cvetkovic A. (2015) "Prevalent business models for the Apple App Store," in Proceedings of the Wirtschaftsinformatik 2015, Osnabrück, Germany. Paper 81.

Björk, S. & Holopainen, J. (2005) Patterns in Game Design. Charles River Media, Hingham, MA.

Chen, Y., & Xie, J. (2005) "Third-party product review and firm marketing strategy," in Marketing Science, 24(2), pp. 218-240.

Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007) "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," in Journal of Interactive marketing, 21(4), pp. 23-45.

Eliashberg, J. & Shugan, S.M. (1997) "Film critics: Influencers or predictors?" in Journal of Marketing, 61(2), pp. 68-78.

EEDAR (2010) The Influence of Professional Critic Reviews. EEDAR/SMU Behavioral Study.

EEDAR (2014) Deconstructing Mobile & Tablet Gaming. EEDAR's Annual Report.

Electronic Arts (2012) *The Simpsons: Tapped Out* [iOS] Electronic Arts.

Fields, T., & Cotton, B. (2012) Social Game Design. Monetization Methods and Mechanics. Morgan Kaufmann, Waltham, MA.

Fishlabs (2010) *Galaxy On Fire 2* [iOS] Fishlabs.

Fullerton, T., Swain, C. & Hoffman, S. (2004) Game Design Workshop: Designing, Prototyping, and Playtesting Games. CMP Books, San Francisco, CA.

Greenwood-Ericksen, A., Poorman, S.R. & Eludamos, R.P. (2013) "On the validity of Metacritic in assessing game value," Journal for Computer Game Culture. 7 (1), pp. 101-127.

Hamari, J. (2015) "Why do people buy virtual goods? Attitude towards virtual good purchases versus game enjoyment," in International Journal of Information Management, 35(3), pp. 299-308.

Hamari, J. & Keronen, L. (2016) "Why do people buy virtual goods? A literature review," in Proceedings of the 49th Annual Hawaii International Conference on System Sciences (HICSS), Hawaii, USA, January 5-8, 2016. DOI: 10.1109/HICSS.2016.171

Hamari, J. & Lehdonvirta, V. (2010) "Game design as marketing: How game mechanics create demand for virtual goods," in International Journal of Business Science & Applied Management, 5(1), pp. 14-29.

Hamari, J. & Järvinen, A. (2011) "Building customer relationship through game mechanics in social games," in M. Cruz-Cunha, V. Carvalho, P. Tavares (Eds.), Business, Technological and Social Dimensions of Computer Games: Multidisciplinary Developments, IGI Global, Hershey, PA.

Jellyoasis (2012) *Elf Defense Eng* [iOS] Jellyoasis.

Karppi, T. & Sotamaa, O. (2012) "Rethinking playing research: DJ HERO and methodological observations in the mix," in Simulation & Gaming, 43(3), pp. 413-429.

King (2011) *Candy Crush Saga* [iOS] King.

Lankoski, P. & Björk, S. (2015) "Formal analysis of gameplay," in Lankoski, P. & Björk, S. (Eds.) Game Research Methods: An Overview. ETC Press, pp. 23-35.

Lehdonvirta, V. (2009) "Virtual item sales as a revenue model: identifying attributes that drive purchase decisions," in Electronic Commerce Research, 9(1-2), pp. 97-113.
Lin, H. & Sun, C.-T. (2011) "Cash trade in free-to-play online games," in Games and Culture, 6 (3), pp. 270-287.
Luton, W. (2013) Free-to-Play: Making Money from Games You Give Away. New Riders.
Machine Zone (2013) *Game of War - Fire Age* [iOS] Machine Zone.
Madgarden (2012) *Punch Quest* [iOS] Rocketcat Games.
Martin, M. (2015) "PewDiePie passes 10 billion views on YouTube," in VG247. Available at https://www.vg247.com/2015/09/07/pewdiepie-1-billion-views-youtube/ (accessed May 2016)
Matulef, J. (2012) "Irrational Games job ad lists 85+ Metacritic score as a requirement," in Eurogamer. Available at http://www.eurogamer.net/articles/2012-07-26-irrational-games-job-ad-lists-metacritic-score-as-a-requirement (accessed May 2016)
Myers, D. (2010) Play Redux: The Form of Computer Games. Ann Arbor: The University of Michigan Press.
Mäyrä, F. (2008) An Introduction to Game Studies. SAGE, London, UK.
Nieborg, D.B. & Sihvonen, T. (2009) "The new gatekeepers: The occupational ideology of game journalism," in Breaking New Ground: Innovation in Games, Play, Practice and Theory. Proceedings of DiGRA 2009.
Orland, K. (2014) "Steam Gauge: Do strong reviews lead to stronger sales on Steam?" in Ars Technica. Available at http://arstechnica.com/gaming/2014/04/steam-gauge-do-strong-reviews-lead-to-stronger-sales-on-steam/ (accessed May 2016)
Paavilainen, J., Hamari, J., Stenros, J. & Kinnunen, J. (2013) "Social network games: Players' perspectives," in Simulation & Gaming, 44 (6), pp. 794-820.
Pinchefsky, C. (2013) "Social network games shed 10 million players...in one month," in Forbes. Available at http://www.forbes.com/sites/carolpinchefsky/2013/04/17/social-network-games-shed-10-million-players-in-one-month/ (accessed May 2016)
Riot Games (2009) *League of Legends* [PC Computer, Online Game] Riot Games.
Rose, M. (2012) "Report: Obsidian missed Fallout: New Vegas bonus by one Metacritic point," in Gamasutra. Available at http://www.gamasutra.com/view/news/165900/Report_Obsidian_missed_Fallout_New_Vegas_bonus_by_one_Metacritic_point.php (accessed May 2016)
Rovio Entertainment (2011) *Angry Birds Rio HD* [iOS] Rovio Entertainment.
Schell, J. (2008) The Art of Game Design: A Book of Lenses. Morgan Kaufmann, Burlington, MA.
Siwek, S.E. (2014) Videogames in the 21st Century - The 2014 Report. Entertainment Software Association. 2014.
Supercell (2012) *Clash of Clans* [iOS] Supercell.
Supercell (2012) *Hay Day* [iOS] Supercell.
Time (2015) "The 30 most influential people on the Internet," in Time. Available at http://time.com/3732203/the-30-most-influential-people-on-the-internet/ (accessed May 2016)
Totilo, S. (2008) "Low Metacritic scores cause game publishers to withhold developer royalties," in MTV Multiplayer. Available at http://multiplayerblog.mtv.com/2008/05/29/low-metacritic-costs-developers/ (accessed May 2016)
Wargaming (2010) *World of Tanks* [PC Computer, Online Game] Wargaming.