

Investigating the Relationship Between User Ratings and Gamification – A Review of mHealth Apps in the Apple App Store and Google Play Store

Manuel Schmidt-Kraepelin
Karlsruhe Institute of Technology
manuel.schmidt-kraepelin@kit.edu

Scott Thiebes
Karlsruhe Institute of Technology
scott.thiebes@kit.edu

Ali Sunyaev
Karlsruhe Institute of Technology
sunyaev@kit.edu

Abstract

Gamification of mHealth apps is regarded as a promising approach to counteract decreasing long-term motivation of mHealth app users. Although gamification has received tremendous attention from researchers interested in mHealth apps, little is known about the extent to which gamification is used in real world mHealth apps today and whether the implementation of gamification actually pays off for app developers by, for example, positively influencing user ratings. Within this research, we investigate the implementation of game mechanics for 1,000 apps from the Apple App Store and Google Play Store as well as the potential relationship between the degree of gamification of mHealth apps and their user ratings. While our results highlight a high degree of adoption of gamification for both app stores, they also indicate a positive relationship between the degree of gamification of an mHealth app and user ratings for the Apple App Store only.

1. Introduction

Since the opening of the Apple App Store in 2008 the number of apps available on the App Store has reached a total of 2,200,000 apps in 2017 [36]. Even more so, according to Apple, app developers earned \$20 billion on the App Store in 2016 alone [4]. Although apps in the categories *Health & Fitness* and *Medical* account for a comparatively low combined share of only 4.87% of all apps on the App Store [35], almost 60% of US smartphone users have installed at least one health-related app (henceforth mHealth app) on their smartphones [22]. Overall, the value of the mobile health industry amounted to \$6.7 billion in 2012 and is expected to grow to \$58.8 billion by 2020 [37].

Considering the surge of apps available to users in both, the Apple App Store and Google Play Store [40], it is becoming increasingly difficult for mHealth app

developers to attract potential users' attention, getting them to download their app, and subsequently motivating them to use their apps over a sustained period of time. Once downloaded, almost 70% of app users use their apps infrequently [6], with 26% of all downloaded apps being used only once and 74% being discontinued altogether by the tenth time [34].

In recent years, gamification has become one of the most popular approaches to address users' infrequent and decreasing long-term use of apps [22] and has garnered much attention, especially from researchers interested in mHealth apps [17]. By applying game design elements to non-game contexts [11], gamification aims to foster users' intrinsic motivation [16], making the usage of apps or other systems more engaging and fun [14]. Researchers have investigated the use of gamification in diverse health-related contexts [31] such as medication adherence, or disease self-management [2]. However, despite gamification's popularity within academia, past studies focused on outlining and applying different game mechanics to mHealth apps [27], as well as resulting psychological or behavioral effects [17]. Little is known about how extensively gamification is actually implemented outside research studies, in real-world mHealth apps [24], and whether it is beneficial for developers of mHealth apps to incorporate gamification into their apps or only unnecessary overhead. To this end, user ratings along with the name of the app and an app's ranking in the respective app store play an important role in the decision process of choosing an app to download, since users primarily rely on information provided by the app selection screen [12]. Apps with higher user ratings have higher ranks in top lists of app stores, which in turn leads to an increase in visibility and download numbers [29]. Within this research, we therefore aim approach the aforementioned research gap by answering the following research questions:

RQ1: How extensively is gamification used in real-world mHealth apps?

RQ2: What is the relationship between the use of gamification and mHealth apps' user ratings?

In the past, extensive efforts have been put into research on online recommender systems as well as the effects of user ratings [e.g., 1, 3]. More recently and with the emergence of mobile app stores such as the Apple App Store and Google Play Store, researchers have increasingly focused on the analysis of user ratings and reviews provided on these app stores, including what features users value most and what users complain about most [15], the relationship between comments and app ratings [29], identifying spam reviews [8], and the effects of app bugs and errors on user ratings [7]. Moreover, some researchers have investigated sampling issues related to research studies on samples of apps [25], as well as the development of quality measures for apps, based on reviews and ratings [39]. Although mHealth apps are a frequent context for such studies [e.g., 26], to the best of our knowledge, we are among the first to explicitly investigate the relationship between mHealth apps' user ratings and their use of gamification. To answer our research questions, we manually downloaded 1,000 mHealth apps from the Apple App Store and Google Play Store and coded the use of game elements by those apps. While our results highlight a statistically significant difference in the use of game elements between good and poor-rated apps for the Apple App Store, results for the Google Play Store are mixed. Our results also show that the use of game elements is positively correlated with the number of user ratings in both mobile app stores. With our research, we contribute to the scientific knowledge base on gamified mHealth apps by providing insights into the actual use of gamification in a large number of real-world mHealth apps in the Apple App Store and Google Play Store. We are among the first to take a more developer-centric perspective in our analysis of gamified mHealth apps by investigating the relationship between the use of gamification and user ratings.

This paper proceeds as follows. The next section provides an outline of gamification and related game elements as well as an overview of research on gamified health apps. Section three describes our research approach, whereas section four presents our results. We discuss our results in section five and conclude our paper in section six.

2. Background on gamified mHealth apps

2.1. Game elements in gamification

Literature provides two prevailing definitions for gamification. First, Huotari and Hamari [19] define gamification as a process of enhancing services with motivational affordances for gameful experiences.

Hamari et al. [17] advanced this conceptualization by introducing the three main parts of gamification (i.e., implemented motivational affordances, resulting psychological outcomes, and further behavioral outcomes) and outlining their relationships. Second, Deterding et al. [11] define gamification as “the use of game design elements in nongame contexts”. Kari et al. [21] propose that these definitions arise from two different understandings of gamification (i.e., process view and experience view). In addition to that, past research around gamification differentiates diverse types of game elements, the most prominent classification being the Mechanics, Dynamics, Aesthetics (MDA) Framework [43]. Mechanics are functional components of gamified applications that provide various actions and control mechanisms to enable user interaction [18]. Dynamics determine the runtime behavior of mechanics concerning players' inputs and outputs over time and aesthetics refer to the “desirable emotional responses evoked in users when they interact with the gamified system” [5]. As this work aims to analyze the actual use of game mechanics in mHealth apps, we take a process view of gamification [21] and thus align our understanding with the definition by Deterding et al. [11].

Within this work, we base our analysis on the gamification elements identified by Thiebes et al. [41] because, to the best of our knowledge it is the most comprehensive collection of gamification elements in literature as it has been developed by systematically reviewing game elements proposed in gamification literature. In their study, Thiebes et al. [41] classify the proposed gamification elements in five clusters: System Design, Challenges, Rewards, Social Influences, and User Specifics. However, they do not clearly specify which of the gamification elements are being categorized as mechanics, dynamics, or aesthetics respectively. Thus, we analyzed the gamification elements proposed by Thiebes et al. [41] and decided for each gamification element whether it was classified as game mechanic, game dynamic or game aesthetic. In order to do so, every game element's description was reviewed and checked against the MDA framework by three researchers. Table 1 shows the results of this classification. In our research, only game mechanics are relevant for further analysis because, per definition, game mechanics are the functional components that gamify an app and thus they reveal whether and to what extent an app is gamified from a process point of view. Game dynamics and game aesthetics, on the other hand, are different for every user and thus relate more to the experience view of gamification. One and the same game mechanic can lead to a gameful and fun experience for one user and to the opposite for another [19].

Table 1: Gamification elements proposed by Thiebes et al. (2014)

Cluster	Type ¹	Gamification elements
System design	Mechanics	Feedback, Audible feedback, Reminder, Meaning, Interaction concepts, Visually resembling existing games, Fantasy
	Dynamics & Aesthetics	-
Challenges	Mechanics	Goals, Time pressure, Progressive disclosure
	Dynamics & Aesthetics	-
Rewards	Mechanics	Achievement, Point system, Badges, Bonus, Loss aversion
	Dynamics & Aesthetics	Ownership
Social influences	Mechanics	Shadowing, Leaderboards, Virtual goods
	Dynamics & Aesthetics	Status, Collaboration, Reputation, Competition, Envy, Social facilitation, Conforming behavior, Altruism
User specifics	Mechanics	User levels, Ideological incentives, Virtual character
	Dynamics & Aesthetics	Self-expression

¹Assignment of Game Element Type was part of this research

2.2. Gamification in mHealth apps

For mHealth apps, gamification has become an important topic. Especially when it comes to bringing behavioral interventions into real life contexts, gamified mHealth apps are a promising approach to overcome barriers to health behavior change [10]. In particular, mHealth apps have the potential to, for example, facilitate the management and prevention of (chronic) diseases or supporting healthier lifestyles by promoting physical activity or a healthy diet [32]. In mHealth apps, gamification is primarily applied for motivating individuals to continue using an mHealth app more regularly or promoting the completion of activities or tasks that are associated with positive health outcomes [38]. Accordingly, several studies exist that outline and apply different game mechanics to mHealth apps [27] and most studies investigate the psychological or behavioral effects that occur when introducing specific game mechanics to a certain mHealth app [17]. In addition, several studies propose frameworks that aim to guide researchers and practitioners in their journey of gamifying systems such as mHealth apps [e.g., 2]. However, it remains unclear to which extent game mechanics have been adapted in real world apps and whether they can have an impact on user ratings or popularity measures and thus app success. Lister et al. [24] tried to address this gap by coding 132 apps regarding the ten effective game elements proposed by Reeves and Read [30] and six core components of gamification for health, which were derived by reviewing existing literature. However, their analysis is limited to the Apple App Store and thus disregards the Google Play Store which also accounts for a great amount of app downloads. In addition, their analysis is not based on an established classification of game mechanics and does not specifically analyze the effect of their implementation on user ratings or other popularity metrics.

3. Methods

3.1. App selection

In order to draw a representative sample of mHealth apps we used the repository by Xu and Liu [42], which is the most comprehensive repository of mHealth applications for the two prevailing mobile app stores. By the time of data collection in August 2017 it consisted of 41,298 apps in the category “Health & Fitness” and 27,509 apps in the category “Medical” from the US Apple App Store as well as 25,035 apps in the category “Health & Fitness” and 11,195 apps in the category “Medical” from the US Google Play Store. We decided to use the US version of the app stores as they contain over 98% of available apps on the market and thus have the highest ratio of available apps of all app stores [42].

To ensure feasibility of the study and a high level of data quality, we excluded apps that met at least one of the following exclude criteria. (1) Apps that were not free for download in order to avoid bias arising from users’ higher expectations towards apps they have to pay for (Apple App Store: N=4,253; Google Play Store: N=1,680). (2) Apps with less than ten user ratings to ensure that single user ratings did not have too much influence on the overall analysis [9] (Apple App Store: N=59,931; Google Play Store: N=23,961). This led to a reduced data set of 4,623 mHealth apps from the Apple App Store and 10,589 mHealth apps from the Google Play Store. Based on the reduced set of mHealth apps we drew random samples of N = 250 for successful and N = 250 for unsuccessful apps for both app stores respectively. In line with extant research [15], we used three stars as the threshold value for the classification of successful and unsuccessful apps. This provided us with a total sample of 1,000 mHealth apps in four different groups. Figure 1 shows the overall process of sample selection for both app stores. A full list of selected apps is available from the authors on request.

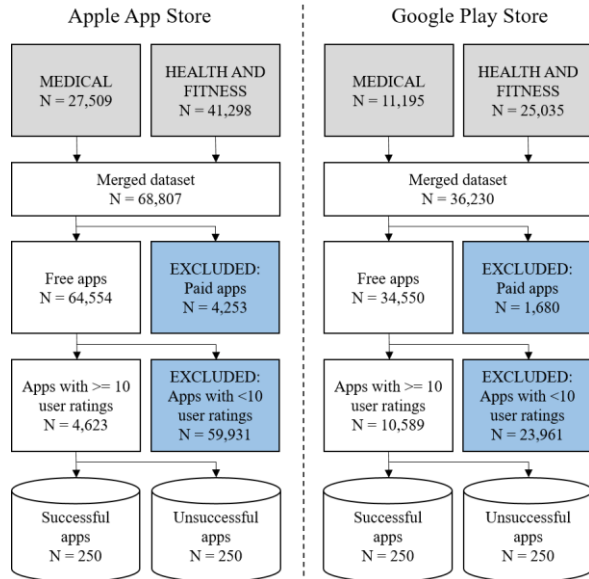


Figure 1: App selection process

3.2. App analysis

Each app of the sample was coded regarding the implementation of the game mechanics proposed by Thiebes et al. [41] (see Table 1). In order to reduce coding subjectivity, each app was downloaded, tested, used for a sufficient period of time to experience all features and subsequently coded by two research assistants independently. In total, four research assistants were recruited from undergraduate and graduate students at a German university. Prior to data analysis, the research assistants were trained in the understanding and interpretation of the game mechanics. Research assistants only coded whether a specific game mechanic was implemented or not implemented but not to what extent or in which quality a game mechanic was implemented. In case the coding differed, an app was discussed until consensus was found. In order to ensure a high level of inter-coder reliability, research assistants compared and discussed their results after analyzing and coding an initial set of 50 mHealth apps. Cohen's Kappa was used to measure inter-coder reliability. Overall, a Cohen's Kappa coefficient of .63 was measured. Thus, according to established guidelines, a substantial level of agreement between coders was achieved [23].

During app analysis, additional exclude criteria were applied. Apps that met at least one of the following additional exclude criteria were also excluded from analysis and replaced with an app from the similar group to ensure a sample size of 250 for every group. (3) Apps that were not available for download at the time of analysis. (4) Apps that

required additional devices such as wearables in order to experience all app features. (5) Apps that did not work properly or crashed during analysis. (6) Apps that required a professional account, such as apps that were only available to registered physicians.

4. Results

Of the 1,000 apps that were downloaded, tested, and coded in total, 722 apps contained at least one game mechanic (Apple App Store: 286/500; Google Play Store 436/500). The overall mean value of implemented game mechanics was 2.34 with a standard deviation of 2.52 (Apple App Store: $M = 1.71$, $SD = 2.32$; Google Play Store: $M = 2.98$, $SD = 2.55$). A Mann-Whitney U Test indicated that apps from the Google Play Store implemented significantly more game mechanics ($Mean\ rank = 591.11$) than apps from the Apple App Store ($Mean\ rank = 409.89$), $U=79685.5$, $p<.0001$.

The most common game mechanics implemented were (1) *Interaction concepts* (538/1000), (2) *Feedback* (357/1000), (3) *Reminder* (320/1000), (4) *Meaning* (235/1000), and (5) *Goals* (209/1000). Table 2 shows the five most common game mechanics for both app stores and both groups of apps, respectively. The overall least common game mechanics were (1) *Bonus* (6/1000), (2) *Loss aversion* (10/1000), (3) *Badges* (11/1000), (4) *Virtual goods* (12/1000), and (5) *Visually resembling existing games* (12/1000). Appendix A shows an overview of all game mechanics and their number of codings for both app stores.

Table 2: Most common game mechanics

Game mechanic	Apple App Store				Google Play Store			
	Rank	Successful apps	Unsuccessful apps	Sum	Rank	Successful apps	Unsuccessful apps	Sum
Interaction concepts	1	128	55	183	1	178	177	355
Feedback	2	80	41	121	2	108	128	236
Reminder	3	66	25	91	3	101	128	229
Meaning	5	57	21	78	4	65	92	157
Goals	4	53	33	86	5	58	65	123

For both app stores combined, the successful apps collectively implemented 1,306 game mechanics. The mean value of implemented game mechanics was 2.61 with a standard deviation of 2.69. The unsuccessful apps, collectively implemented 1,038

game mechanics with a mean value of 2.08 and a standard deviation of 2.30. In the Apple app store, the successful apps implemented 592 game mechanics in total ($M = 2.37$; $SD = 2.64$) and the apps categorized as unsuccessful overall used 263 game mechanics ($M = 1.05$; $SD = 1.73$). In the Google Play Store sample, the successful apps collectively implemented 714 game mechanics ($M = 2.86$; $SD = 2.73$) and the unsuccessful apps used 775 game mechanics in total ($M = 3.10$; $SD = 2.35$). Figure 2 gives an overview of how many game mechanics were implemented in both app stores and for both groups of apps.

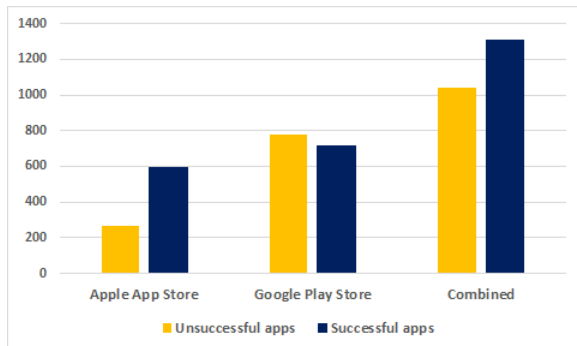


Figure 2: Amount of coded game mechanics

In order to analyze whether specific clusters of game mechanics are more extensively used in mHealth apps than others, we performed additional analyses based on the clusters proposed by Thiebes et al. [41]. Therefore, we first measured for every cluster the total number of times an element within the cluster was coded. Based on this value we calculated a normalized cluster index (C_i) by dividing the total number of codings by the amount of mechanics a cluster contains and the underlying sample size. The resulting value C_i states to which extent a certain cluster was implemented on average per app for a specific group of apps. Table 3 gives an overview on both measures for successful, unsuccessful and all apps based on the combined sample of Apple App Store and Google Play Store. The results show that the game mechanics within the cluster *System Design* ($C_i=.2336$) are the most common game mechanics in mHealth apps, followed by game mechanics in the cluster *Challenges* ($C_i=.121$). The clusters *Social Influences* ($C_i=.0343$), *User Specifics* ($C_i=.034$), and *Rewards* ($C_i=.0282$) all showed substantially lower implementation values. In addition, we also calculated both measures for the Apple App Store and Google Play Store individually. However, results, particularly the ranking of clusters, did not differ substantially from the results of the overall sample.

Table 3: Most common clusters

Cluster	Successful apps		Unsuccessful apps		All apps	
	Total # coded	C_i	Total # coded	C_i	Total # coded	C_i
System design	890	.2543	745	.2129	1635	.2336
Challenges	203	.1353	160	.1067	363	.121
Social influences	50	.0333	53	.0353	103	.0343
User specifics	66	.044	36	.024	102	.034
Rewards	97	.0388	44	.0176	141	.0282

To further investigate the relationship between the use of game mechanics and user ratings, we calculated Mann-Whitney U Tests comparing the two groups of successful and unsuccessful apps concerning the amount of mechanics they implemented. We chose a Mann-Whitney U test over a t-test due to non-normal distribution of the data. The Mann-Whitney U tests indicate that in the Apple App Store successful apps implemented significantly more game mechanics than unsuccessful apps ($U=20,312.5$; $p<.0001$). In the Google Play Store successful apps implemented slightly less game mechanics than unsuccessful apps ($U=27,823.5$; $p=.032$). We also performed Spearman correlations to investigate the relationship between the amount of implemented game mechanics and available measures for app success (i.e., average user rating and number of user ratings). We decided to analyze number of user ratings as a measure for popularity as download or install numbers were not available for all apps. For the Apple App Store, we found a positive significant correlation between the amount of implemented game mechanics and average user rating ($r=.364$; $p<.0001$) and a positive significant correlation between amount of game elements and number of user ratings ($r=.216$; $p<.0001$). For the Google Play Store, we did not find a significant correlation between the amount of implemented game mechanics and average user rating ($r=-.065$; $p=.144$), but a positive significant correlation between amount of game elements and number of user ratings ($r=.226$; $p<.0001$).

5. Discussion

5.1. Implications

Analyzing the implementation of game mechanics for 1,000 mHealth apps from the Apple App Store and the Google Play Store revealed some interesting insights into the current dissemination of gamification in mHealth apps. Out of 1,000 investigated mHealth apps, 722 (72.2%) contained at least one game mechanic. Compared to the study performed by Lister et al. [24] from 2014 (they stated

that 52.5% of the analyzed apps in their sample contained at least one element of gamification) this is an substantial increase. This finding can be attributed to the fact that the concept of gamification is still rather new but has become increasingly popular, especially for mHealth apps, during the last years [20]. However, it might also be influenced by the underlying coding schemes as the collection of game elements proposed by Thiebes et al. [41] interprets gamification rather broadly compared to other collections of gamification elements such as the elements of gamification used by Lister et al. [24]. Analysis of the normalized cluster indices for the five clusters of game elements revealed that game mechanics in the clusters *System design* and *Challenges* are substantially more common in mHealth apps than game mechanics in other clusters. In particular, the game mechanics *Interaction concepts*, *Feedback*, *Reminder*, *Meaning*, and *Goals* were most common. When comparing these results to extant research on user preferences for gamification [e.g., 28, 33], it becomes clear that there still seems to be a gap between user preferences and the actual implementation of gamification. In particular, some game mechanics that are highly preferred by users such as *User levels* and *Point Systems* do not get the attention by app developers that users would hope for. However, this does not mean that implementing simple point systems will eventually lead to app success as research in other fields often enough showed that user preference statements and actual user behavior can substantially differ. In fact, despite frequent criticism that gamification is simply "pointification" [13], in real-world mHealth apps we see that point systems are actually not used that much.

The results of our study suggest heterogeneous landscapes for mHealth apps in the Apple App Store and the Google Play Store. More precisely, our results show that mHealth apps in the Google Play Store implement game mechanics more extensively than mHealth apps in the Apple App Store. A potential reason for this is that app developers aim to satisfy heterogeneous needs of different target groups and to adapt their apps accordingly. Another reason might be that app design guidelines and review processes for the Apple App Store are much stricter and that Apple is known for rejecting apps if they differ too much from their design guidelines (e.g., cluttered with gamification elements). For app developers, this means that they have to be aware of the app landscape they want to contribute to and the corresponding user needs before they decide to implement gamification or not. In addition, the results of our study suggest that successful and

popular mHealth apps in the Apple App Store make significantly more extensive use of game mechanics than unsuccessful apps. From our point of view, this observation can be ascribed to three potential underlying reasons. (1) Gamification really "works" [17] and makes mHealth apps more engaging and fun as suggested by extant research. This higher level of fun and engagement then positively influences the ratings that users of the Apple App Store assign gamified mHealth apps. (2) Developers of good and successful mHealth apps have recognized the industry trend for more gamification in mHealth apps in an early stage and implemented game mechanics accordingly. (3) The proposed game mechanics by Thiebes et al. [41] are not limited to such game mechanics that are traditionally associated with gamification (e.g., *Badges*, *Point systems*, *Leaderboards*), but also contain game mechanics that are traditionally more associated with good application design per se (e.g., *Interaction concepts*). Contrary to these results, analyzing the results of the Google Play Store sample revealed a mixed picture of the relationship between the use of game mechanics and user ratings in mHealth apps. According to the Mann-Whitney U Test, successful mHealth apps in the Google Play Store implemented even slightly less game mechanics than unsuccessful apps. However, Spearman correlations slightly amend these results as we did not find a significant relationship between amount of implemented game mechanics and average user ratings but a positive relationship between amount of game mechanics and amount of user ratings. Although, our study results can not completely explain these findings, they might be attributed to the fact that the Google Play Store in general has a higher level of implemented game mechanics. As a result, the often cited novelty effect of gamification [17] might have already reduced the positive effects of gamification for users of the Google Play store. Contradicting to these results, the positive significant relationship between amount of game mechanics and number of user ratings as a measure for app popularity indicates a potentially positive relation between gamification and app success within the Google Play Store. An additional explanation for these contradicting results might be the presence of other highly influential factors that drive users' ratings for mHealth apps in the Google Play Store and that potentially overshadow our analysis concerning the relationship between average user ratings and gamification.

5.2. Limitations & Future Research

The findings of this study should be interpreted in consideration of some key limitations. First, we limited our analysis to free apps which may have excluded relevant available paid mHealth apps. However, since there is a recent industry trend for free mobile apps [24], we are confident that this sample sufficiently represents the majority of apps available in the Apple App Store and Google Play Store. Second, different teams of research assistants were recruited to code mHealth apps from the Apple App Store and the Google Play Store and thus the coders' perceptions of game mechanics might have influenced the different results for both app stores. In addition, some game mechanics are more prone to subjective coding than others. For example, coding whether a *Point system* exists in an mHealth app is easier than the assessment of *Interaction concepts*. We aimed to counteract these problems by intensively coaching coders in their understanding of game mechanics and the overall coding process. In addition, Cohen's Kappa indicates a substantial level of agreement between coders. Nevertheless, we cannot completely rule out the possibility that specific interpretations of the game mechanics evolved within coding teams during data analysis as coders from one team did not interact with coders from other teams. Third, due to a different level of available information, the repository did not contain a similar level of data quality for both app stores. For example, average user ratings in the Apple App Store were rounded to the typical star ratings (1, 1.5, 2, [...] 5) while average user ratings in the Google Play Store were reported accurately. Fourth, our analysis is limited to the implementation of single game mechanics as well as the amount of implemented game mechanics and their relationship to user ratings. By doing so, we do not consider effects that arise from the perception of the design as a whole (e.g., contradicting game mechanics) although this might have a substantial effect on user ratings. Unfortunately, within this research we were not able to ask users for their perception of gamification concepts implemented in the respective mHealth apps. Finally, in order to ensure feasibility of the study, we only downloaded and analyzed a subset of available mHealth apps for both app stores. According to Martin et al. [25], analyzing subsets of mined app store information can lead to a sampling bias and thus threaten validity especially with regard to inferential statistics. We aimed to account for this problem by purposefully sampling different groups of applications. In addition, we cannot rule out the possibility that we reviewed and coded some apps

that were mistakenly assigned to one of the categories although they had no health related functionality and thus potentially skewed the results to some extent.

Our study creates various starting points for future research. First, we only considered quantitative measures of user ratings to investigate the relationship between user ratings and the implementation of game mechanics in mHealth apps. Future studies could delve deeper and additionally analyze the actual content of user app reviews (i.e., textual feedback provided by users on app stores [15]) in order to better understand which aspects of gamification are associated with positive or negative effects on user ratings. Second, our study is limited to the analysis of mobile applications in the categories "Medical" and "Health and Fitness". In addition, future research could take a more holistic approach and analyze users' perceptions of gamification concepts as a whole and its influence on user ratings. Future research could additionally investigate whether our findings hold for different categories or whether the application context significantly influences the relationship between gamification and user ratings. With regard to the operationalization of app success, we suggest future research also takes a look at other measures (e.g., total number of downloads, app store rankings) that are of similar importance for app developers. Concerning the relationship between the implementation of gamification and app success, it might also be interesting to study the evolution of specific apps and the impact of gamification on these apps over time. In order to do so, future studies could, for instance, analyze the effects that occur to user ratings after a new app version containing one or multiple (new) game mechanics is released. Finally, our study showed mixed results concerning the relationship between implementation of gamification and user ratings. Particularly notable is the difference between apps from the Apple App Store and the Google Play Store concerning the use of gamification and its effect on user ratings. Since we were not able to fully explain this observation in this study (our results do not tell us anything about causality), future studies should delve deeper into this topic and find profound explanations (e.g., different target groups) for the difference in the app landscape between both app stores. Due to the novelty of the research field and the exploratory nature of our research, future studies could make use of qualitative research methods (e.g., interviews with users of apps from both app stores) in order to gain a better understanding concerning the characteristics of heterogeneous user groups and factors that drive user ratings in mobile app stores.

6. Conclusion

Gamification of mHealth apps is a promising approach to counteract decreasing long-term motivation of mHealth app users and thus a popular measure to promote health behavior change. However, it is unclear how extensively gamification is used in real world mHealth apps today and whether the implementation of gamification also pays off for app developers by, for example, positively influencing user ratings. In this study, we contribute to the scientific knowledge base on gamification in mHealth apps by analyzing the use of gamification in 1,000 mHealth apps from the Apple App Store and the Google Play Store. The results of our study support the assumption that gamification has reached a certain level of popularity in mHealth apps over the last years. However, there still exists a gap between the use of game mechanics in real-world mHealth apps and users' gamification preferences. It is on future developers of mHealth apps to diminish this gap by carefully assessing gamification preferences of potential users and designing their gamification concepts accordingly. With regard to the relationship between the use of gamification in mHealth apps and user ratings, the results of our study are mixed. While our analysis shows a significant difference in the use of game mechanics between successful and unsuccessful apps for the Apple App Store, we did not find any such difference for the Google Play Store. In addition, our study suggests that the use of gamification in the Google Play Store is substantially more common than in the Apple App Store. Although this finding might be attributed to the strict design guidelines and review process by Apple, it is on future research to delve deeper into this research area and identify more profound explanations for this finding.

7. References

- [1] Adomavicius, G. and Tuzhilin, A., "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", *IEEE transactions on knowledge and data engineering*, 17(6), 2005, pp. 734-749.
- [2] Almarshedi, A., Wills, G.B., and Ranchhod, A., "The Wheel of Sukr: A Framework for Gamifying Diabetes Self-Management in Saudi Arabia", *Procedia Computer Science*, 63, pp. 475-480.
- [3] Amatriain, X., Pujol, J.M., and Oliver, N., "I Like It... I Like It Not: Evaluating User Ratings Noise in Recommender Systems", *International Conference on User Modeling, Adaptation, and Personalization*, Trento, Italy, Jun 22-26, 2009, pp. 247-258.
- [4] Apple, "App Store Shatters Records on New Year's Day", Retrieved from <http://www.apple.com/newsroom/2017/01/app-store-shatters-records-on-new-years-day.html>, accessed Jun 15, 2018.
- [5] Basten, D., "Gamification", *IEEE Software*, 34(5), 2017, pp. 76-81.
- [6] Bauer, A.M., Rue, T., Keppel, G.A., Cole, A.M., Baldwin, L.-M., and Katon, W., "Use of Mobile Health (Mhealth) Tools by Primary Care Patients in the Wwami Region Practice and Research Network (Wprn)", *The Journal of the American Board of Family Medicine*, 27(6), 2014, pp. 780-788.
- [7] Bavota, G., Linares-Vasquez, M., Bernal-Cardenas, C.E., Di Penta, M., Oliveto, R., and Poshyvanyk, D., "The Impact of Api Change-and Fault-Proneness on the User Ratings of Android Apps", *IEEE Transactions on Software Engineering*, 41(4), 2015, pp. 384-407.
- [8] Chandy, R. and Gu, H., "Identifying Spam in the Ios App Store", *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality*, Lyon, France, Apr 16, 2012, pp. 56-59.
- [9] Dehling, T., Gao, F., Schneider, S., and Sunyaev, A., "Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on Ios and Android", *JMIR mHealth and uHealth*, 3(1), 2015, pp. e8.
- [10] Dennison, L., Morrison, L., Conway, G., and Yardley, L., "Opportunities and Challenges for Smartphone Applications in Supporting Health Behavior Change: Qualitative Study", *Journal of medical Internet research*, 15(4), 2013, pp. e86.
- [11] Deterding, S., Dixon, D., Khaled, R., and Nacke, L., "From Game Design Elements to Gamefulness: Defining Gamification", *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, Tampere, Finland, Nov 4-6, 2011, pp. 9-15.
- [12] Dogruel, L., Joeckel, S., and Bowman, N.D., "Choosing the Right App: An Exploratory Perspective on Heuristic Decision Processes for Smartphone App Selection", *Mobile Media & Communication*, 3(1), 2015, pp. 125-144.
- [13] Esteves, J.M., "The Perils of Gamification Trivialization: How and Why Gamification Is Failing to Deliver Loyalty", *Academy of Management Proceedings*, 2017(1), 2017, pp. 13560.
- [14] Flatla, D.R., Gutwin, C., Nacke, L.E., Bateman, S., and Mandryk, R.L., "Calibration Games: Making Calibration Tasks Enjoyable by Adding Motivating Game Elements", *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, Santa Barbara, California, USA, October 16 - 19, 2011, pp. 403-412.
- [15] Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., and Sadeh, N., "Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store", *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, Illinois, USA, August 11 - 14, 2013, pp. 1276-1284.
- [16] Hamari, J. and Koivisto, J., "Working out for Likes: An Empirical Study on Social Influence in Exercise

- Gamification", *Computers in Human Behavior*, 50, pp. 333-347.
- [17] Hamari, J., Koivisto, J., and Sarsa, H., "Does Gamification Work?--a Literature Review of Empirical Studies on Gamification", 47th Hawaii International Conference on System Sciences (HICSS), 2014 Waikoloa, HI, Jan 6-9 2014, pp. 3025-3034.
- [18] Hunnicke, R., Leblanc, M., and Zubek, R., "Mda: A Formal Approach to Game Design and Game Research", Proceedings of the Nineteenth National Conference of Artificial Intelligence Challenges in Games AI Workshop, San José, California, Jul 25-26, 2004.
- [19] Huotari, K. and Hamari, J., "Defining Gamification: A Service Marketing Perspective", Proceedings of the 16th International Academic MindTrek Conference, Tampere, Finland, Oct 3-5, 2012, pp. 17-22.
- [20] Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., and Hides, L., "Gamification for Health and Wellbeing: A Systematic Review of the Literature", *Internet Interventions*, 6, pp. 89-106.
- [21] Kari, T., Piippo, J., Frank, L., Makkonen, M., and Moilanen, P., "To Gamify or Not to Gamify?: Gamification in Exercise Applications and Its Role in Impacting Exercise Motivation", BLED 2016: Proceedings of the 29th Bled eConference.
- [22] Krebs, P. and Duncan, D.T., "Health App Use among Us Mobile Phone Owners: A National Survey", *JMIR mHealth and uHealth*, 3(4), 2015, pp. e101.
- [23] Landis, J.R. and Koch, G.G., "The Measurement of Observer Agreement for Categorical Data", *biometrics*, 33(1), 1977, pp. 159-174.
- [24] Lister, C., West, J.H., Cannon, B., Sax, T., and Brodegard, D., "Just a Fad? Gamification in Health and Fitness Apps", *JMIR serious games*, 2(2), 2014, pp. e9.
- [25] Martin, W., Harman, M., Jia, Y., Sarro, F., and Zhang, Y., "The App Sampling Problem for App Store Mining", *IEEE/ACM 12th Working Conference on Mining Software Repositories*, Florence, Italy, 16-17 May, 2015, pp. 123-133.
- [26] Mendiola, M.F., Kalnicki, M., and Lindenauer, S., "Valuable Features in Mobile Health Apps for Patients and Consumers: Content Analysis of Apps and User Ratings", *JMIR mHealth and uHealth*, 3(2), 2015.
- [27] Miller, A.S., Cafazzo, J.A., and Seto, E., "A Game Plan: Gamification Design Principles in Mhealth Applications for Chronic Disease Management", *Health informatics journal*, 22(2), 2016, pp. 184-193.
- [28] Nour, M.M., Rouf, A.S., and Allman-Farinelli, M., "Exploring Young Adult Perspectives on the Use of Gamification and Social Media in a Smartphone Platform for Improving Vegetable Intake", *Appetite*, 120, pp. 547-556.
- [29] Pagano, D. and Maalej, W., "User Feedback in the Appstore: An Empirical Study", 21st IEEE International Requirements Engineering Conference, Rio de Janeiro, Brazil, Jul 15-19, 2013, pp. 125-134.
- [30] Reeves, B. and Read, J.L., *Total Engagement: How Games and Virtual Worlds Are Changing the Way People Work and Businesses Compete*, Harvard Business Press, 2009.
- [31] Schmidt-Kraepelin, M., Thiebes, S., Baumsteiger, D., and Sunyaev, A., "State of Play: A Citation Network Analysis of Healthcare Gamification Studies", *European Conference of Information Systems*, Portsmouth, UK, Jun 23 - 28, 2018.
- [32] Schmidt-Kraepelin, M., Thiebes, S., Tran, M.C., and Sunyaev, A., "What's in the Game? Developing a Taxonomy of Gamification Concepts for Health Apps", *Proceedings of the 51st Hawaii International Conference on System Sciences*, Waikoloa, Hawaii, USA, Jan 3-6, 2018, pp. 1217-1226.
- [33] Schöbel, S., Söllner, M., and Leimeister, J.M., "The Agony of Choice--Analyzing User Preferences Regarding Gamification Elements in Learning Management Systems", *International Conference on Information Systems*, Dublin, Ireland, pp. 1-21.
- [34] Smith, D., "Motivating Patients to Use Smartphone Health Apps", Retrieved from <http://www.prweb.com/releases/2011/04/prweb5268884.htm>, accessed Jun 15, 2018.
- [35] Statista, "Most Popular Apple App Store Categories in December 2016, by Share of Available Apps", Retrieved from <https://www.statista.com/statistics/270291/popular-categories-in-the-app-store/>, accessed Jun 15, 2018.
- [36] Statista, "Number of Available Apps in the Apple App Store from July 2008 to January 2017", Retrieved from <https://www.statista.com/statistics/263795/number-of-available-apps-in-the-apple-app-store/>, accessed Jun 15, 2018.
- [37] Statista, "Value of the Global Mhealth (Mobile Health) Market 2012-2020", Retrieved from <https://www.statista.com/statistics/295771/mhealth-global-market-size/>, accessed Jun 15, 2018.
- [38] Stepanovic, S. and Mettler, T., "Gamification Applied for Health Promotion: Does It Really Foster Long-Term Engagement? A Scoping Review", *European Conference on Information Systems*, Portsmouth, UK, Jun 23 - 28, 2018.
- [39] Stoyanov, S.R., Hides, L., Kavanagh, D.J., Zelenko, O., Tjondronegoro, D., and Mani, M., "Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps", *JMIR mHealth and uHealth*, 3(1), 2015.
- [40] Sunyaev, A., Dehling, T., Taylor, P.L., and Mandl, K.D., "Availability and Quality of Mobile Health App Privacy Policies", *Journal of the American Medical Informatics Association*, 22(e1), 2014, pp. e28-e33.
- [41] Thiebes, S., Lins, S., and Basten, D., "Gamifying Information Systems-a Synthesis of Gamification Mechanics and Dynamics", *European Conference on Information Systems*, Tel Aviv, Israel, Jun 9-11, 2014, pp. 1-17.
- [42] Xu, W. and Liu, Y., "Mhealthapps: A Repository and Database of Mobile Health Apps", *JMIR mHealth and uHealth*, 3(1), 2015, pp. e28.
- [43] Zichermann, G. and Cunningham, C., *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*, O'Reilly Media, Inc., Gravenstein Highway North, Sebastopol, CA, 2011.

Appendix A

Table A-1: Game mechanics, descriptions, and total number of codings

Game mechanics	Cluster	Description	Apple App Store			Google Play Store			Combined			
			Rank	Successful apps	Unsuccessful apps	Rank	Successful apps	Unsuccessful apps	Rank	Successful apps	Unsuccessful apps	Sum
Interaction concepts	System design	An attractive user interface with stimulating visuals and exciting interaction concepts, as well as a high degree of usability.	1	128	55	1	178	177	1	306	232	538
Feedback	System design	Immediate feedback that is used to keep the player aware of progress or failures in real time.	2	80	41	2	108	128	2	188	169	357
Reminder	System design	Reminder of past behavior of the user.	3	66	25	3	101	128	3	167	153	320
Meaning	System design	Taking into consideration the background that the user brings to the activity and the context into which the specific activity is placed. Game elements emerge from aspects of the underlying activity that are meaningful to the user.	5	57	21	4	65	92	4	122	113	235
Goals	Challenges	Goals of the underlying activity that can be adapted as challenges for the user.	4	53	33	5	58	65	5	111	98	209
Audible feedback	System design	Implementing sound effects and / or background music.	6	51	25	7	34	35	6	85	60	145
Time pressure	Challenges	Creating time pressure on activities, (e.g., through counters or hourglasses).	7	29	9	6	29	46	7	58	55	113
Achievement	Rewards	A reward for completing a clear and desirable goal.	9	22	5	8	31	13	8	53	18	71
Ideological Incentives	User specifics	The notion of influencing user behavior through influencing their attitudes and values and thus educating the user on a deeper level. Ideological incentives make it possible to motivate users by themselves.	8	16	17	10	22	8	9	38	25	63
Shadowing	Social influences	Describes a method where users attempt to improve their previous records.	10	15	8	9	8	25	10	23	33	56
Point system	Rewards	Reward users for completing actions, whereby a numeric value is added to their overall total score.	11	15	5	12	14	9	11	29	14	43
Progressive disclosure	Challenges	Continuously increasing users' skills by progressive disclosure of both knowledge and challenge. This helps to ensure that the challenges in the app match the player's skill level.	14	12	0	11	22	7	12	34	7	41
Leaderboards	Social influences	Are used to track and display desired actions, using competition with peers to drive valuable behavior.	13	12	5	13	6	12	13	18	17	35
Fantasy	System design	Evokes images of objects or situations that are not actually present. This can make the experience more emotionally appealing to users.	12	11	7	15	5	5	14	16	12	28
User levels	User specifics	Indicate the proficiency of the player in the overall gaming experience over time.	15	8	1	14	12	3	15	20	4	24
Virtual character	User specifics	Visually represents the user within the app. Users might design their Avatar regarding their own preferences.	16	3	3	16	5	4	16	8	7	15
Virtual goods	Social influences	Non-physical, intangible objects that can be purchased or traded.	18	4	0	18	5	3	17	9	3	12
Resembling existing games	System design	Creating a visual design, which is very similar to existing games. For example, designing the system similar to the well-known Tetris game.	18	3	1	17	3	5	18	6	6	12
Badges	Rewards	Optional rewards and goals whose fulfilment is outside the scope of the core activities of the app.	17	4	1	20	4	2	19	8	3	11
Loss aversion	Rewards	Influences user behavior not by a reward, but by punishment in case the targeted goal is not achieved.	20	1	1	19	2	6	20	3	7	10
Bonus	Rewards	Rewards for having completed a series of challenges or core activities.	21	2	0	21	2	2	21	4	2	6