IT UNIVERSITY OF COPENHAGEN

Master Thesis

Personalized Gamification in Exercise Applications: A Practical Implementation

Author: Sebastian Karstensen
Supervisor: Daniel Cermak-Sassenrath

IT UNIVERSITY OF COPENHAGEN
Copenhagen, Denmark

A thesis submitted in fulfillment of the requirements for the degree of Master of Science.

September 2018

Abstract

Gamification has been proven to have a positive impact towards motivating most people. Nevertheless, some researchers are doubting the efficiency of gamification when not adapted to the individual user. This has led to several studies on correlations between persuasive strategies and user types, but not many have included a practical implementation of a personalized system. This is partly due to the fact that no general approach to the problem has been established yet. Based on a framework presented by Tondello, Orji, and Nacke, a GPS-based running application has been developed together with a recommender system, which will allow for personalization of the gamified elements depending on the user's Hexad type. To examine the recommender system, a total sample of 11 user types were collected. Three of the data providers participated in a one-hour user test followed by a focus group to investigate the benefits of personalizing the application. The results showed that recommender systems are a possible way of implementing personalized gamification, although they may yield most benefits when incorporated as a feature for suggestions or creating planned programs for the user. No conclusions could be made in regards to the efficiency of personalizing the application due to the small test group, but indications show that it may promote further engagement.

Keywords: Gamification, Personalization, Recommender Systems, Health Applications, Player Models, User Type, GPS.

Contents

List of Figures

List of Tables

1
Introduction

This chapter briefly describes the background for the project and then presents the research questions, including the underlying motivation. The chapter concludes with an outline of how the rest of the report is structured.

1.1 Background

Gamification, the use of game elements in a non-game context, is a constantly expanding field. As new technologies are being developed, it has become much easier to track the performance of users. The phenomenon has been widely applied in exercise applications due to the increased popularity of wearables and GPS-enabled devices. Generally, most studies concerning the usage of gamification in active games has shown promise in motivating users towards further physical activity, but the research also reveals that the effort is rarely effective for all test subjects. Where some people find encouragement in social components of exergames, others are motivated by the fictional points and achievements of the applications. There is no 'one size fits all'. This has led people to question the efficiency of gamification, when it is not tailored to the user profile. By attempting to satisfies every motivational factor, is it not possible that some components, which actually have a positive impact on the individual user, might be drowned out by an overload of persuasive actions? Would it be more beneficial to personalize each user's experience based on their profile, their actions, and their environment?

1.2 Problem Statement

As researchers have started to question the efficiency of gamification when used as a one-size-fits-all approach, it has become increasingly popular to investigate the effects of personalizing the gamified application to the individual. The majority of the work completed in this area relates to determining the correlation between specific user types (found through models such as Big Five/Five Factor or BrainHex) and persuasive strategies or game mechanics, and indicates that personalizing the application based on player type is very likely to further engage users, also in health applications.

However, there are some limitations to the previous studies: first of all, when testing the persuasiveness of gamification in exergames, the user testing has mainly been conducted through use of storyboards or videos, and it is questionable whether the effect of the persuasive strategies will be similar when implemented in an actual, practical application. Secondly, the suggested personalization models are static and are not updated in real time based on the ongoing interaction between the user and the system. Tondello, Orji, and Nacke have suggested a general framework for using recommender systems in personalizing gamification, but we have yet to see an implementation of the recommendations for actual users of a gamified application. Some researchers argue that gamification systems only have a positive short-term effect, and that the motivational impression is only compelling when the user is new to the system. It is possible that implementing an "on-demand", personalized system could strengthen the lifetime of the persuasive strategies and help motivate users for a long-term perspective as well.

The goal of this thesis is therefore to answer the questions: **What are the benefits of personalized gamification when implemented in a practical application? How can a recommender system be implemented to enable on-demand personalization?**

To test this hypothesis, a prototype will be developed based on the framework presented by Tondello, Orji, and Nacke. The goal is to apply personalized gamification to exercise, and thus the intended system will be a geolocation-based, mobile application, in which users will have to complete running activities to progress in the game. Based on their user types, the recommender system will suggest activities that are most-likely to motivate the user, using different game mechanics and persuasive strategies. To evaluate the recommender system, a sample of user profiles will be collected, so the results of using real-world data in the system can be examined. Furthermore, the running application will be distributed to a group of users for a one-hour testing session. For comparison, the test will be split into three parts, each containing a different version of the application. The first and second version will include the game elements that are most and least likely to motivate them according to the recommender system, whereas the third does not implement the recommender system at all, and instead display all activities for the user. The test concludes with a focus group interview. The results will then be analyzed and evaluated.

1.3 Outline

The report is structured as follows:

Chapter 2 provides an overview of the most recent research done in the area of personalized gamification, as well as some general background information about gamification, player modelling, and recommender systems.

Chapter 3 describes the project execution, what activities were conducted during testing, and comments on the validity of the study.

Chapter 4 gives a detailed description of the designed recommender system and the developed application 'Runderful', which were both used for testing.

Chapter 5 shows the collected data and results of the user tests.

Chapter 6 presents findings based on the results and covers limitations of the project.

Chapter 7 concludes the report by summarizing the findings and relating them to the research questions.

2

State of the Art

This chapter provides the reader with the knowledge necessary for understanding the activities undertaken during this project, along with a summary of former research done in this area. First, the concept of gamification is defined, which is followed by its usage in health application. Afterwards, a section on player modelling describes three different concepts for modelling the player type and why it is relevant in gamification. Finally, an overview of previous attempts at personalizing gamification is presented, including suggested frameworks, which eventually ties in to a short recount of recommender systems and their potential usage.

2.1 Gamification

Deterding et. al defines gamification as "the use of game design elements in non-game contexts" [Deterding et al., 2011]. They argue that gamification, first and foremost, relates to 'games' and not 'play', meaning that the application of gamification has to take place in a context structured by rules, and in which players attempt to achieve a certain goal. Although the user of a gamified application may change their behaviour to be more 'playful' (defined by Caillois as 'paidia'), the usage of gamification does not fit well in the more improvisational and free-form setting of 'play'.

Furthermore, it is important to note that gamified applications "[...] merely incorporate elements of games". Defining what exactly classifies as game elements turns out to be difficult, and can lead to continuous discussions. For now, the definition suggested by Deterding et. al will be accepted; game elements used in gamification appear in the majority of games and "play a significant role in gameplay". [Deterding et al., 2011, p. 12]

In terms of 'non-game context', it is suggested that the only distinction to make is specifically the usage of game-elements in games, as this would simply be viewed as 'game design'. Although it is argued that implementations such as 'achievement systems' (aka meta-games) in games could be seen as gamification, it is specifically tied to non-game contexts for the sake of simplicity. As games and gamification are already closely related, the user's ability to distinct between the two leads to complications for the research (in terms of user testing).

Based on this, gamification can be thought of as followed: applying characteristic game elements and mechanics to an otherwise non-game related context, in an attempt to increase user engagement.

2.2 Usage in Health Applications

One popular application of gamification is in health-related settings, for instance to promote further physical activity of the user or encourage a healthier diet. In their 12-week study of 117 college students, Huang et. al. found that playing exergames can indeed have a positive impact on the participant's physical fitness [Huang et al., 2017, p. 313]. The study indicates that even in a gamified setting, the exercise activity is still beneficial for a person's health, thus making it relevant to increase user participation. This is further supported by Johnson et. al., who, through a literature review of papers investigating the effect of gamification on health and well-being, found that "gamification of health and wellbeing interventions can lead to positive impacts, particularly for behaviours, and is unlikely to produce negative impacts" [Johnson et al., 2016, p. 101]. Sardi et. al. performed a similar review of the research done on gamification in e-Health, and identified that "the major advantage of gamification in the health context is, perhaps, that of ensuring users' regular engagement and increasing their immersion into the e-Health solution" [Sardi et al., 2017, p. 41]. Therefore, even by simply being utilized on existing physical exercise, and not implemented as an exergame, gamification appears to motivate users towards additional activity.

However, similar to how people prefer playing different games, there is also a difference in what gamification elements motivates us the most. Although their study on using gamification in a step counting application showed an increase in participations, Nakashima et. al. noted that the application was "effective especially for people with high competitive spirit" [Nakashima et al., 2017, p. 2355]. Seeing as the main motivation technique was implementing a 'Leaderboard' (which is commonly associated with competition and comparing one's result to others'), it is not surprising how the mobile app proved significantly well among the competitive participants. On the other hand, when studying the relation between competition and competitive individuals, Song et. al. discovered a negative reaction to the competitive setting from participants who were less competitive [Song et al., 2013, p. 1706]. Similar responses could be expected from differences in other personal characteristics such as social longing, autonomy, etc.

Therefore, even though gamification has been proven to have a positive effect in most cases, the differences in people still creates challenges in making it useful for everyone. González and Adelantado brings up 'personalization' as an option for solving the problem [González and Adelantado, 2017, p. 4]. By applying the correct form of motivation towards the right person at the right time, it is possible to avoid alienating users, who are not driven by the gamification mechanic. One way to personalize gamification is through the use of player modelling, which has been an important input for creating adaptive video games.

2.3 Player Modelling

As the growth of the video games industry has raised player expectations to new heights, it has become increasingly difficult for game designers to make games that people enjoy. In order to ensure that games are encouraging for all types of players, some game developers have started customizing the gameplay to the individual's preferences, trying to make sure that everyone playing their game has an enjoyable experience [Busch et al., 2016b, p. 146]. Due to the evolution of computer technologies such as artificial intelligence and machine learning, the next step in player-centred design approach is becoming a reality, with the possibility of making on-demand personalization [Bakkes et al., 2012, p. 1]. Because personalization has grown in popularity, a wide range of models for determining the type of user has also developed. By knowing the user type, it is possible for the designer to choose specific game elements and mechanics, which are more likely to resonate with the user's motivational drivers. Some of the proposals for modelling player types are described in the following section.

2.3.1 Bartle: Achiever, Explorer, Socialiser, and Killer

One of the most well-known examples of player modelling is Bartle's taxonomy of player types [Bartle, 1996]. In order to describe the playing style of MUDs (Multi-User Dungeons), Bartle discovered four types of players: **Achievers** have the main goal of gathering points and rising in level. **Explorers** are driven by the unknown, these players try to discover the parts of the game that others may never see. **Socialisers** aren't necessarily interested in the gameplay, but mostly compelled by the interaction with other players in the network. **Killers** get joy from upsetting other players, usually by making it difficult or impossible for them to reach their goals. The player types can be seen in figure 2.1, where they are displayed in relation to the two-dimensional playing styles defined as part of the model.

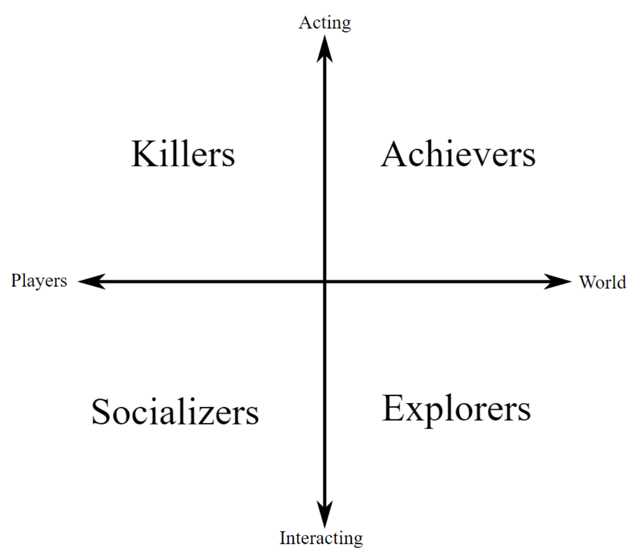


Figure 2.1: Bartle taxonomy of player types.

2.3.2 Bateman and Nacke: BrainHex

Unlike Bartle's player types, which are limited to MUDs and MMORPGs, the BrainHex Model can be used to clarify the given player experience in any game [Bateman et al., 2015]. Based on other game design models, players are divided into 7 different categories: Seeker, Survivor, Daredevil, Mastermind, Conqueror, Socialiser, and Achiever. Please note that the archetypes of Socialiser and Achiever are still present in this player model, even though this model is based on neurological findings, and the research method is much different from the one conducted by Bartle 14 years earlier.

2.3.3 Marczewski: User Types Hexad Scale

As an alternative for those using player types specifically in a gamified contexts, Marczewski introduced the Gamification User Types Hexad [Marczewski, 2015]. As seen on figure 2.2, the model consists of "six user types that differ in the degree to which they can be motivated by either intrinsic (e.g., self-realization) or extrinsic (e.g., rewards) motivational factors" [Busch et al., 2016a]. Inspired by the intrinsic motivators in self-determination theory (Mastery, Autonomy, Relatedness), the suggested user types are:

- **Philanthropist:** driven by purpose, and prefers actions that help others, such as sharing and giving.
- **Socialiser:** motivated by interactions with other users and expand their network.
- **Free Spirit:** wishes to explore and create, and is less interested in following the main paths suggested by the system.

- **Achiever:** competition is key, and challenges are a necessity. Skill improvement and progression are the main motivators.
- **Player:** fueled by extrinsic rewards, such as points and badges.
- **Disruptor:** does not feel constrained by the limitations of the game and looks for the ability to change and innovate (for instance, through developer tools).

To standardize the procedure of defining the player type based on the Hexad scale, Tondello et al. have developed a survey response scale consisting of 24 question, which can describe a user's preferences [Busch et al., 2016a]. Furthermore, during the same research, it was also confirmed that there was a positive correlation between the user type and what was predicted as as preferred design element. This indicates that the Hexad scale can be a useful tool for determining the main motivations of a user, which gameful applications can then be personalized around.

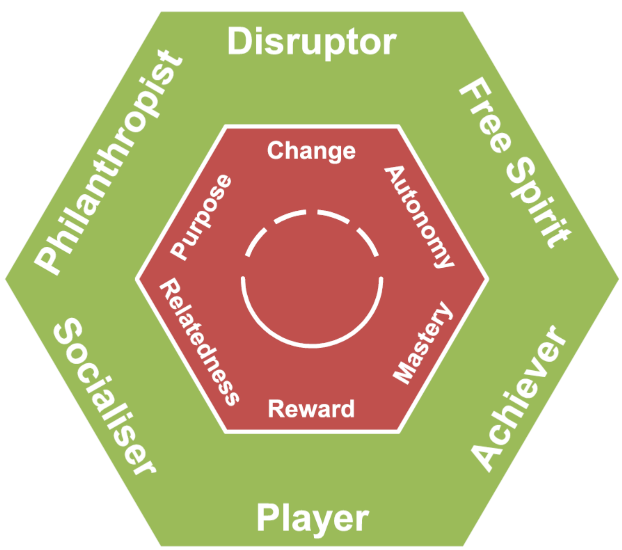


Figure 2.2: Gamification User Types Hexad.

2.4 Personalized Gamification

Research has shown that developers might be able to "increase efficiency and user adoption or better support business processes through a higher and long-term user engagement" [Böckle et al., 2017, p. 159] by adapting their gamification mechanics to the user profiles. Orji et al. found that "people's personality traits play significant roles in their responsiveness and preference for various persuasive strategies" [Orji et al., 2017, p. 1015]. This was further indicated in a following study on how persuasive strategies affect different gamification user types [Böckle et al., 2018, p. 10], specifically based on the types found in Marczewski's User Type Hexad (see section 2.3.3). Other studies, such as the one conducted by Halko and Kientz [Halko and Kientz, 2010], also supports the hypothesis of a close relation between specific personalities and persuasive strategies. This clearly indicates that there is a higher chance of motivating people through gamification by appealing to their user type.

However, there are still doubts about the payoff of personalization. First of all, it may not always be possible to customize your application to the specific user. Secondly, if it is possible, there is no guarantee that the extra work needed will be beneficial in terms of user engagement compared to time spent implementing the system. For instance, Orji et. al suggests two approaches to personalized gamification: either a one-size-fits-all approach, which will motivate as many people as possible, or tailoring the application to the individual type [Orji et al., 2013, p. 2473]. It is worth considering only targeting one segment, if the majority of people are motivated by the same game elements.

One major concern in the research is found in the way the testing has been conducted. In the previously mentioned studies, which all shown positive signs of personalization, test participants were only presented a storyboards or videos of persuasive examples and then surveyed on how they would react. Since gamification usually aims at motivating people towards doing stuff they are not already encouraged to do, it is possible that the results found through surveys may not correlate to applying the personalization in a practical setting, where users will actually complete the intended activities. Few experiments have been done in a real-world setting, but Monterrat et al. found that "members of the group with adapted features spent 39% more time on the learning environment than the members of the group with counter-adapted features" [Monterrat et al., 2015, p. 305], indicating that the results will be similar.

Part of the reason for the limited practical experiments in research is that no general process of personalizing a gamified application has been found yet. Knutas et al. managed to develop a ruleset and translate it to "machine-format rules that can be used as a plugin algorithm for computer-supported collaborative learning environments" [Knutas et al., 2017, p. 5], but the algorithm is only applicable in learning systems. A different approach was suggested by Tondello et al., who used recommender systems for ranking various activities and game elements based on user data and contexts [Tondello et al., 2017, p. 427]. The framework includes specific topics that should be considered when developing a personalized system, but the overall concept is general enough to be applied to almost any application, and will therefore be examined in this project. Since recommender systems play a key-role in this proposal, the main theory behind these systems will briefly be described in the following section.

2.4.1 Recommender Systems

According to Balabanovic and Shoham, recommendation services can be categorized into three different approaches [Balabanovic and Shoham, 1997]:

- **Content-based Recommendations**: Items in the system are analyzed and scored on a list of attributes, which can be compared to the user's profile. As the user rates more items, their profile will be updated, and new items will be recommended. Because the recommendations are based on the preferences of the individual user, the system is able to provide optimal suggestions without the data of thousand of similar users (an issue known as the "cold start problem"). On the other hand, since the system solely relies on the item attributes, it may be difficult to assign the correct features for the recommendations to be sufficient [Ricci et al., 2011, p. 78].

- **Collaborative Recommendations**: Instead of recommending items that are similar to what the user has liked in the past, the system suggests items based on users with the same previous ratings [Balabanovic and Shoham, 1997, p. 67]. Unlike content-based recommendations, there is no need to decide on a set of features for the items to be rated, as the only data used is other users' rankings. Even though the system becomes more precise with increased user participation, new items will always suffer from the cold start problem, and will therefore require a significant amount of user ratings before proper recommendations can be given.

- **Hybrid Recommendations**: A mix of the two previously mentioned approaches. A known example is to obtain the user profile through content-based recommendations, which are then compared to other users by applying collaborative-filtering. Hybrid recommendations usually solves the problems seen in pure content-based and collaborative systems [Balabanovic and Shoham, 1997, p. 68].

Known as one of the most popular recommendation techniques, the neighbourhood-based methods have been widely used due to their simplicity, efficiency, and

	Attribute A	Attribute B	Attribute C
Item 1	2	4	2
Item 2	1	5	3
Item 3	5	2	4

Table 2.1: Example: Items for nearest-neighbour algorithm.

stability [Ricci et al., 2011, p. 113]. In it's most elementary form, the nearest-neighbour algorithm uses the Pythagorean theorem to calculate the distance between two items. The closer the items are to each other, the more similar they are. The size of the vector used for calculations can be extended based on the complexity of the items to be recommended. The distance can be calculated as such:

$$distance_{item1,item2} = \sqrt{(a_{i1} - a_{i2})^2 + (b_{i1} - b_{i2})^2 + ... + (n_{i1} - n_{i2})^2}$$

As an example, take three different items, each with three attributes, shown in table 2.1. The distance between Item 1 and Item 2 would be:

$$distance_{i1,i2} = \sqrt{(2-1)^2 + (4-5)^2 + (2-3)^2} = \sqrt{3} = 1.73...$$

In the same way, the distance between Item 1 and Item 3 can be calculated:

$$distance_{i1,i3} = \sqrt{(2-5)^2 + (4-2)^2 + (2-4)^2} = \sqrt{17} = 4.12...$$

Since the distance between Item 1 and Item 2 is much shorter than the one between Item 1 and Item 3, it is concluded that the first pair of items is more similar. It is preferable to display the similarity in a scale of 0 to 1, where 0 means that there is no similarity and 1 means that the two items are identical. Therefore, the following formula is generally used to convert the distance to a “similarity score”, where d is equal to the distance between the two items:

$$similarity_{i1,i2} = \frac{1}{(1+d)}$$

From the previous example, Item 1 and Item 2 has a similarity score of 0.37, whereas Item 1 and Item 3 has a similarity score of 0.20.

3 Method

This chapter briefly covers the research methods conducted in the project in order to answer the research questions.

Since prior research has mostly investigated the effect of personalizing gamification through videos or storyboards, it was chosen to go through the actual process of producing a real-world, personalized application. An exercise application was developed with various game elements for motivating different user types. Instead of choosing other popular gamification areas such as education or work, exercise was selected because it requires intense physical effort from the user. Whereas testers previously only saw an image and said "Yes, that will motivate me to run more", having a practical application forces them to perform the actual exercise.

Meanwhile, a concept for how a recommender system could be used for the application was designed based on one of the frameworks presented in previous research. Although the system was never implemented in the application, it was made possible to choose exactly which activities and feedback mechanics each user was shown, thus emulating the functionality of the recommendations. The experiences gathered throughout the development process are documented in the section 6.1, and can be used as a guideline for future implementations.

In a network of university students, any people interested in running were asked to fill out the User Type Hexad survey to collect a sample of user profiles. A total of 11 responses were gathered. These were used to compare the participants' Hexad type with the recommender systems suggestions. Out of the people surveyed, 3 were interested in participating in an hour-long group test session. Based on their survey answer, they tested three versions of the application:

1. Users were only shown the activity and feedback that were **most** likely to motivate them according to the recommender system.
2. Users were only shown the activity and feedback that were **least** likely to motivate them according to the recommender system.
3. Users were shown all activities and feedback mechanics. Participants were asked to try the activity that they had yet to experience, and were then asked to use the application however whatever they wanted.

Before each version, which took about 20 minutes to test, users were presented with the user guide (see appendix A) for their given activity and feedback, and were able to ask questions about anything incomprehensible. After testing a version of the application, users were then asked to complete a short survey about the activity and feedback they tested, before moving on to the next version.

Due to an injury, one of three users were biking instead of running. This is not optimal for competitive reasons, but by asking the injured tester to bike at a regular pace they did not post unrealistic running times to the application, which other runners might find demotivating. A large amount of artificial user data was generated to make sure that the activities involving other users of the application had the motivational foundation needed to be effective (a leaderboard is not very compelling, if you are the only one on it).

To support the survey answers, a focus group session was conducted with the same three participants afterwards. This allowed users to share ideas and thoughts, and gave a clear understanding of how a user uses the application, including what parts motivates them. Due to the small size of the test group, the test results should more be seen as a proof of concept than any of real validity. Because the testing only took place for a limited amount of time, it is difficult to assume what test participants will think of the system after using it for a while. This requires a quantitative study with more testers and a longer testing period.

Comparing the challenges found during development with the test results can aid in the discussion of whether or not the effort of personalizing the system is worth the extra time. It should give a clearer view of which areas of gamification should be personalized and which should not.

4

Project Experiment

This chapter describes the process of developing a recommender system for a running application, and then shows how it can be implemented to personalize the application based on a user's profile. In the first section, fundamental choices on how to design the recommendation system are covered. This leads into the second section, which details the recommendation process implemented in the test application "Runderful", and also outlines the different game elements and why they have been implemented.

4.1 Designing the Recommender System

4.1.1 Items to be Rated

In their suggested framework, Tondello, Orji, and Nacke describes three types of inputs given to a recommender system: activities, game elements, and persuasive strategies [Tondello et al., 2017, p. 4]. Although exercise can take form of many different activities (walking, running, biking, etc.), it was decided to focus exclusively on running to simplify the application. Instead, different types of game elements were developed, as well as a range of feedback mechanics to display after a given activity (see section 4.2.2 and 4.2.3).

4.1.2 Input Data

In order to match the selected game elements to a specific situation of usage, Tondello et al. presents different types of data that can aid the recommender system in determining the input most likely to satisfy the user. First of all is the user profile, which might cover anything from the user's age, gender, and height, to more intricate measures such as domain specific preferences or intrinsic motivation. In the application implemented for the project, this will be the only source of data used in the recommender system. The other two suggestions, context types and transactions, will not be used as to limit the scope of the application. Although context could help in recommmending activities that are close to the user, it will mainly impact the usability of the system and not the motivation behind its usage. Transactions are used to improve the accuracy of the recommender system, but doesn't see much benefit until the application has been used by a large amount of people over a longer period of time, something that is not part of this project.

4.1.3 Recommendation Methods

As described in section 2.4.1, there are mainly three methods for doing the actual recommendation of the items: content-based, collaborative, or a mix of both. Choosing one or the other will depend on the specific case and what information is available to the system, as well as which context it will be implemented in. Given that this project is limited in size, the cold start problem will be too evident for new users to get any good recommendations. It was therefore chosen to recommend the game elements based on their specific attributes (described in detail in section 4.2.1.3). Tondello, Orji, and Nacke mentions two additional techniques: Context-aware recommendation and Machine-learning recommendation [Tondello et al., 2017, p. 428]. Both of these approaches could be used for enhancing the 'on-demand' personalization required for this project; a context-aware system will include the current time and the user's location in its calculations, and applying machine learning algorithms is likely to improve the recommender system over time. However, none of these techniques will be implemented in the suggested solution as it is out of the scope of this thesis, which aims to investigate the possible effectiveness of real-time personalization, but not the optimization of it.

4.2 'Runderful': Example of an Implementation

To test this approach to personalizing gamification, part of the project involved prototyping a practical application. This led to the development of 'Runderful', a GPS-based web-application aimed towards encouraging its users to run more. The application consists of three different types of running activities, which each incorporate separate game mechanics. After completing an activity, the user is shown one of three types of feedback. Because one of the activities involve team-based mechanics, users are randomly assigned to one of three teams when registered in the system.

4.2.1 Recommending Activities & Feedback

In the application, the system displays different activities and feedback mechanics based on the user's profile, which is found through the User Types Hexad framework. In order to translate the user type to a score comparable to features of the game elements, the correlation matrix found by Tondello et. al. [Busch et al., 2016a, p. 243] is used. The matrix displays how motivating each individual game element is seen by the different user types, measured by Kendall's τ.

This numerical value will be used to calculate the individual users association with each game element, and therefore provides a vector that can similarly be applied to game mechanics for comparison. Depending on complexity and scope of the application, one can choose either all or just some of the game mechanics for the vector - the more mechanics are included, the more effort is needed when scoring game elements (explained in section 4.2.1.3). For the system implemented in this project, a total of 12 game mechanics were selected based on their relevance for the game elements (the full list can be found in the appendix, table B.1).

In the example seen throughout this section, only three game elements will be used for the sake of simplicity: Challenges, Guild or Teams, and Nonlinear Gameplay.

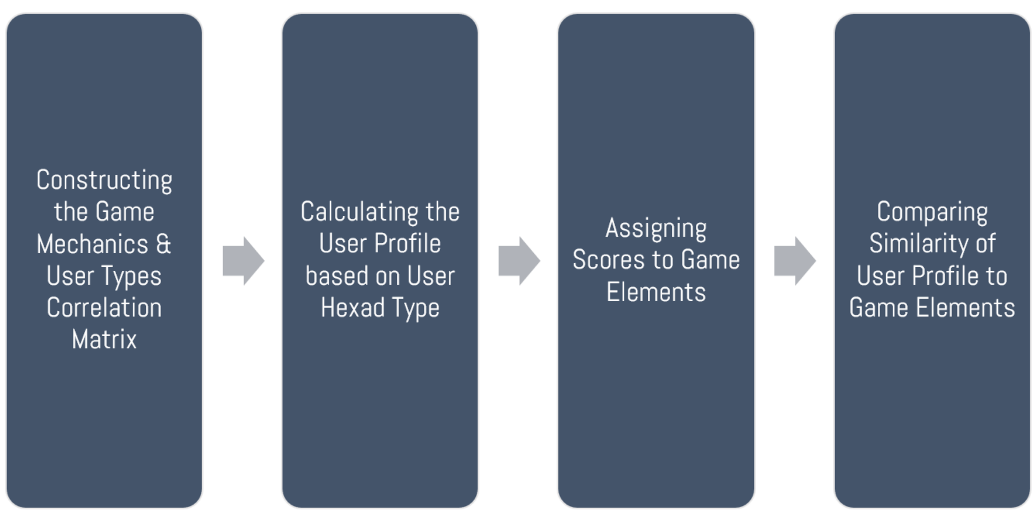


Figure 4.1: The Runderful recommendation process.

	Challenges	**Guild or Teams**	**Nonlinear Gameplay**
Socialiser	0.0	0.179	0
Free Spirit	0.412	0.0	0.221
Achiever	0.436	0.0	0.0
Disruptor	0.207	0.169	0.0
Player	0.317	0.192	0.0
Philanthropist	0.212	0.0	0.179

Table 4.1: Example: Correlations between game elements and Hexad user types.

4.2.1.1 Constructing the Game Mechanics & User Types Correlation Matrix

The first step is to simplify the correlation values. This is done to understand what range of scores will be used for the profile. In this project, it was chosen to map the correlation as follows:

- Insignificant correlations are mapped to the score of 0
- Correlations between 0 and 0.2 are mapped to the score of 1
- Correlations between 0.2 and 0.4 are mapped to the score of 2
- Correlations above 0.4 are mapped to the score of 3

Table 4.1 displays the correlation between Hexad type and game elements, and table 4.2 shows the result of simplifying the values using the beforementioned method.

	Challenges	**Guild or Teams**	**Nonlinear Gameplay**
Socialiser	0	1	0
Free Spirit	3	0	2
Achiever	3	0	0
Disruptor	2	1	0
Player	2	1	0
Philanthropist	2	0	1

Table 4.2: Simplified correlations of table 4.1.

4.2.1.2 Calculating User Profiles

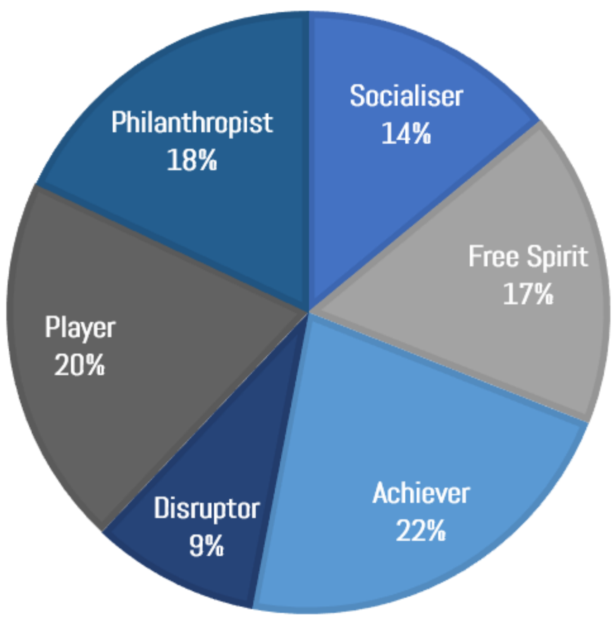


Figure 4.2: User Type Hexad for User A

The succeeding example calculates the user profile for User A, who's Hexad user type can be seen in figure 4.2. In this step, the weight of each user type will be multiplied by the simplified correlation values, and then summed to get a final score for each game mechanic. The sum gives a measurement of

CHALLENGES	**Correlation Score**	**Hexad Score**	**Result**
Socialiser	0	14 %	$0 * 0.14 = 0.00$
Free Spirit	3	17 %	$3 * 0.17 = 0.51$
Achiever	3	22 %	$3 * 0.22 = 0.66$
Disruptor	2	9 %	$2 * 0.09 = 0.18$
Player	2	20 %	$2 * 0.20 = 0.40$
Philanthropist	2	19 %	$2 * 0.19 = 0.38$
Total			**2.13**

Table 4.3: Calculations of user profile for game mechanic Challenges.

	Challenges	**Guild or Teams**	**Nonlinear Gameplay**
Socialiser	0	0.14	0.00
Free Spirit	0.51	0.00	0.34
Achiever	0.66	0.00	0.00
Disruptor	0.18	0.09	0.00
Player	0.40	0.20	0.00
Philanthropist	0.38	0.00	0.18
Total	**2.13 $\approx$ 1**	**0.43**	**0.52**

Table 4.4: Resulting user profile for User A.

how likely the user is to be motivated by the game element. Table 4.3 shows calculations of the score User A will get for the game mechanic Challenges, table 4.4 shows the resulting scores for all three mentioned game mechanics used in the example.

It was decided to set a score limit of 1 for each game mechanic, as it otherwise created issues when comparing the profile with the game elements, due to huge fluctuations in user profiles. By setting a range of 0 to 1, it is easy to see which game mechanics are motivating the user (those above 0.5), and which are not (those below 0.5).

4.2.1.3 Scoring Game Elements

Once the user profile has been found, it needs to be compared to a similar profile for each game element implemented in the system. Setting a score for these game elements is up to the developer, as it will be very individual from application to application. In an attempt to simplify this process, the project application only uses three types of scores:

- 0 = there is no correlation between the game mechanic and the activity/feedback.
- 0.5 = there is somewhat of a correlation between the game mechanic and the activity/feedback.
- 1 = there is a high correlation between the game mechanic and the activity/feedback.

In an attempt to balance the scores, the recommender system was tested with a user profile that was “dominant” in each individual Hexad type, and then compared to the suggested game elements. For instance, the recommender system was run on a Hexad profile of 25% Free Spirit and 15% for the rest of the user types. Seeing as the Exploration activity is aimed towards user with the type of Free Spirit, the resulting recommendations may have led to a possible tweaking of the scores to fit the appropriate activity. This was repeated for all Hexad types several times while adjusting the scores of the game elements, until the output was deemed appropriate.

The activities and feedback mechanics developed for this project is described more in detail in section 4.2.2 and 4.2.3, which also covers the decisions behind how each game element has been scored.

	UP	**GE**	**Result**
Challenges	1	0.5	$(1-0.5)^2 = 0.25$
Guild or Teams	0.43	0.0	$(0.43-0)^2 = 0.185$
Nonlinear Gameplay	0.52	1	$(0.52-1)^2 = 0.230$
Distance			$\sqrt{0.25+0.185+0.230} = 0.815$
Similarity			$\frac{1}{1+0.815} = 0.55$

Table 4.5: Distance and similarity for User A and the Exploration activity.

UP = User Profile, GE = Game Element

4.2.1.4 Comparing Similarity

In this example, we will use the Exploration activity. For the formerly used game mechanics, the activity has been scored as follows: Challenges = 0.5, Guild or Teams = 0, Nonlinear Gameplay = 1. By comparing this to the user profile found for User A, the distance and similarity score between the activity and the user profile can be calculated using the nearest neighbour algorithm. The results are shown in table 4.5.

By calculating the similarity score for multiple game elements, it is possible to determine which one the user is most likely to find motivating. To conclude the example, table 4.6 includes the scoring for each activity implemented in the project, and calculates the distance and similarity based on User A from the example. As seen from the results, User A is most likely to enjoy the activity Time Trials, as the similarity score is the highest. Please note that the example used in this section only included three game mechanics, whereas the application implemented in the project uses twelve game mechanics (and thus a 12-dimensional vector) for calculating similarity.

The next sections covers the different game elements implemented in the test application, which are rated using the recommendation process explained

	Time Trials	**Exploration**	**Crew Tagging**
Challenges	1	0.5	1
Guild or Teams	0	0	1
Nonlinear Gameplay	0	1	0
Distance	0.675	0.815	0.772
Similarity	$\frac{1}{1+0.675} = \mathbf{0.60}$	$\frac{1}{1+0.815} = 0.55$	$\frac{1}{1+0.772} = 0.56$

Table 4.6: Activity recommendation for User A.

above. Each activity and feedback mechanic will be described, including why they are relevant, how they were scored, and to whom they are aimed towards motivating.

4.2.2 Running Activities

The Runderful application allows users to participate in three different types of activities taking place at various locations scattered throughout the parks of Copenhagen. Each activity has been assigned a score (see table 4.7) for every game mechanic used in the recommender system, which allows it to be compared to the individual user's profile.

4.2.2.1 Time Trials

In the Time Trials activity (figure 4.3), users attempt to run a specified route as fast as possible. The route is selected by the user from a list of routes tied to each location, and the activity can be started once the user navigates to the "Start"-area marked on the map. Once begun, a timer will start, and the user has to run to the first checkpoint. Upon reaching the checkpoint, the icon on the map disappears, and the user continues towards the next. This will then be repeated until the "Finish"-area is reached, at which point the timer will stop and the user is shown one of the feedback screens.

This activity is mainly aimed towards competitive users that enjoy challenging themselves, since running against the clock encourages people to not only complete the route, but also complete it as fast as possible. Furthermore, the variety of routes gives an impression of completing different levels, and there is a sense of progression in trying to improve your own time. Therefore, the Time Trials activity has been scored with a 1 in Challenges, and a 1 in Levels/Progression.

The user type from the Hexad scale that is most likely to be motivated by this activity is the Achiever, whereas Free Spirit and Disruptor may not enjoy the constricted form of gameplay.

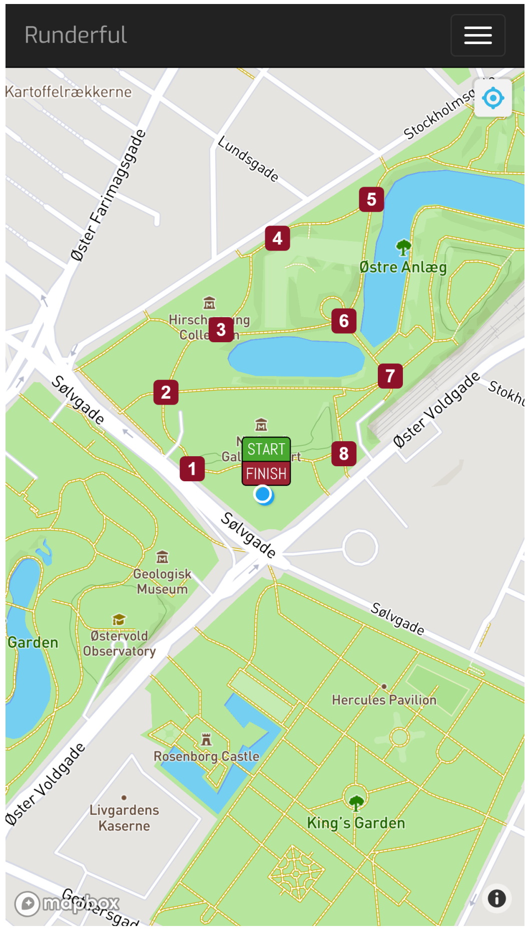


Figure 4.3: Screenshot of the Time Trials activity.

4.2.2.2 Exploration

In the Exploration activity (figure 4.4), users are tasked with finding a virtual "orb" placed on the map. The only information presented to the user is a small signal-icon, which increases as the user gets closer to the orb, as well as a label indicating the current distance to the orb. Using the tools provided, users then have to triangulate the position of the orb by moving around on the map, until they have positioned themselves close enough to the target, whereas a "Collect"-button will appear on screen. When pressed, the orb will be collected, and the user will be shown one of the feedback screens.

Because users can address the objective however they like, and doesn't necessarily have to complete the activity in the fastest time, there is less of a challenge to this activity than Time Trials, and only gets a score of 0.5 for Challenges. In contrast, the gameplay is much more liberal, which is why the activity has been scored with a 1 in both Exploratory Tasks and Nonlinear Gameplay. Given the collection-aspect of finding and retrieving orbs, it has also scored a 1 in Collection and Trading.

The user types from the Hexad scale that are most likely to be motivated by this activity are the Free Spirit and partly Disruptor, due to the non-linear approach to the task.

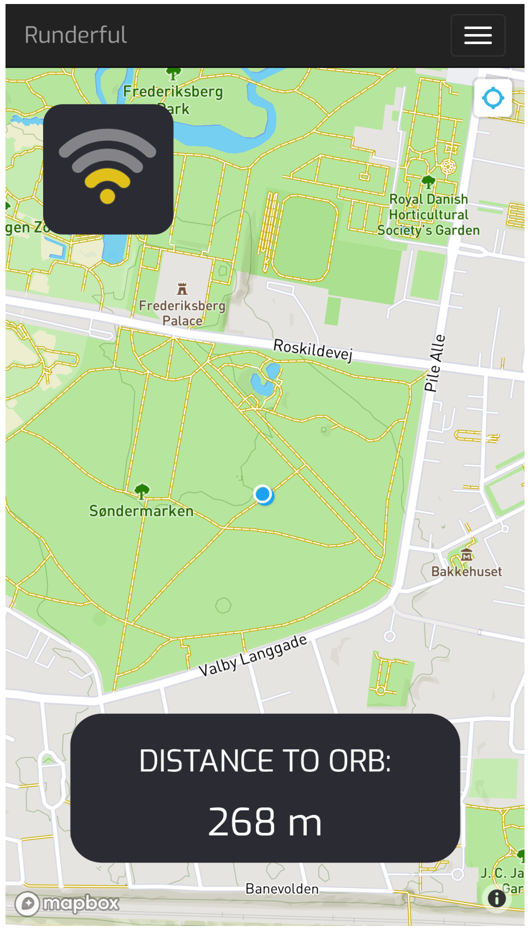


Figure 4.4: Screenshot of the Exploration activity.

4.2.2.3 Crew Tagging

In the final activity, Crew Tagging, users have to represent their team by "tagging" routes throughout the city (figure 4.5). Similarly to Time Trials, the user will select a route which has to be completed. However, no time is displayed for the activity. Instead, when a user completes the route, the route will be displayed as tagged by that user's team in the menu screen. For another team to tag this route, someone from their team would have to complete the route. It is always the team who most recently completed the route that has tagged it, no matter how quickly the run is completed. Upon completion of the route, the feedback screens are shown to the user.

Aimed at bringing social play into the mix, this activity scores a 1 on both Guild or Teams, Social Comparison, and Social Competition, as well as a 1 in Challenges. Focusing more on the team aspect and less on personal progression, there is still a challenge to completing the routes, but the motivation to do it comes from other users, not from beating your own time.

The user types from the Hexad scale that are most likely to be motivated by this activity are the Philanthropist and somewhat Socialiser, since the team challenges gives a sense of belonging and interacting with different users of the system.

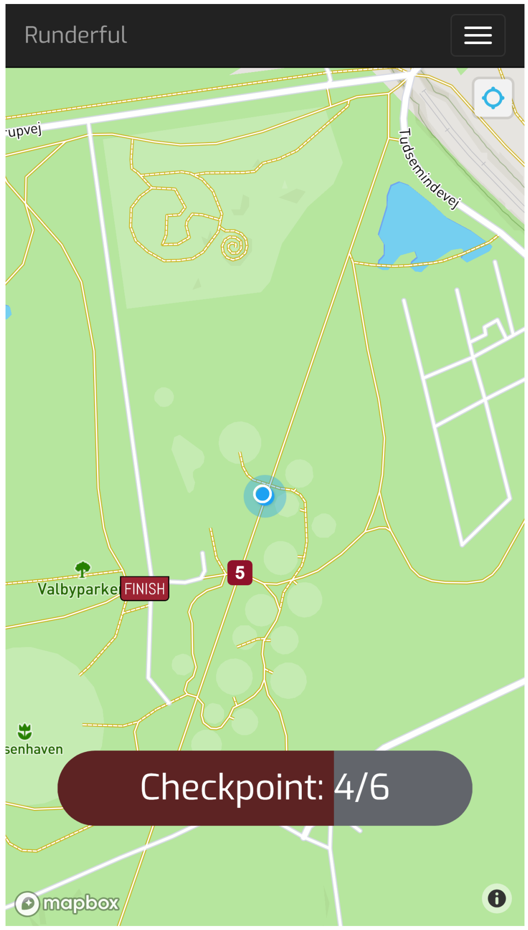


Figure 4.5: Screenshot of the Crew Tagging activity.

4.2.3 Feedback Mechanics

After completing an activity, users will be shown one of three types of feedback. Every feedback mechanic needs to be customized to the individual activity, so each has three different versions: one for Time Trials, another for Exploration, and a final one for Crew Tagging. Like the activities, the

Game Element	**Time Trials**	**Exploration**	**Crew Tagging**
Badges/Achievements	0	0	0
Challenges	1	0.5	1
Collection and Trading	0	1	0
Customization	0	0	0
Exploratory Tasks	0	1	0
Guild or Teams	0	0	1
Leaderboards	0	0	0
Levels/Progression	1	0	0
Nonlinear Gameplay	0	1	0
Points	0	0	0
Social Comparison	0	0	1
Social Competition	0	0	1

Table 4.7: Scoring matrix for the running activities.

feedback types have also been assigned a scoring vector to be used in the recommender system, although not specifically for each version of the feedback. The scoring vectors can be seen in table 4.8.

4.2.3.1 Leaderboard

The Leaderboard feedback (figure 4.6) shows how the user compares to other users of the system. In the Time Trial and Exploration activities, users are ranked based on their completion time of a route or the total amount of orbs they've collected, whereas the Crew Tagging activity displays how big a percentage of all routes the user's team has currently tagged. This type of feedback aims at allowing users to compete against each other, even though they may not be using the system simultaneously, and is highly encouraging for people of competitive nature such as Achievers. Seeing other players in the system can also be quite motivating for Socialisers. At the same time, there is always a goal to reach (in form of another person/team to beat), so the feedback also present the user with somewhat of a challenge. Therefore, this mechanic scores 1 in Leaderboards, Social Comparison, and Social Competition, and 0.5 in Challenges.

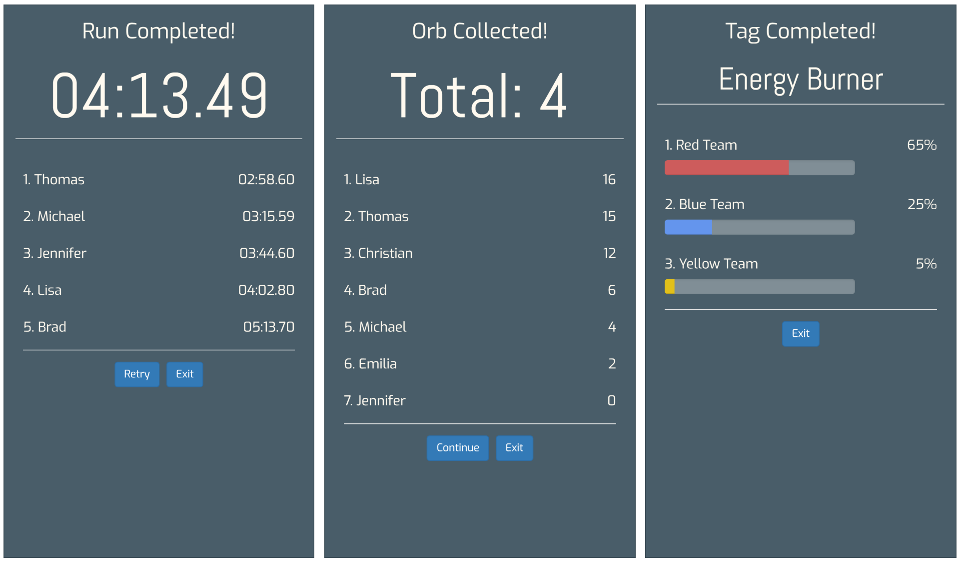


Figure 4.6: Screenshots of the Leaderboard feedback.

Left: Time Trials, middle: Exploration, right: Crew Tagging

4.2.3.2 Medals

To target users motivated by collecting, the application includes the Medals feedback (figure 4.7). Based on how successful the user has performed the activity, they will be shown a bronze, silver, or gold medal. For Time Trials, each medal has a specific time limit associated with it, which the user must beat in order to unlock the medal. Medals for Exploration and Crew Tagging are unlocked based on how many activities of that type have been completed in the current day (for instance, 3 collected orbs in one day gives you a silver medal).

Because of the collection component of this feedback mechanic, it has scored a 1 in Badges/Achievements and Collecting and Trading. Users of type Player are likely to find this feedback motivating, as it involves extrinsic rewards. It could also be a good motivator for Achievers, as they can see how they are progressing in the system (by improving their time on a route or completing more tags in one day).

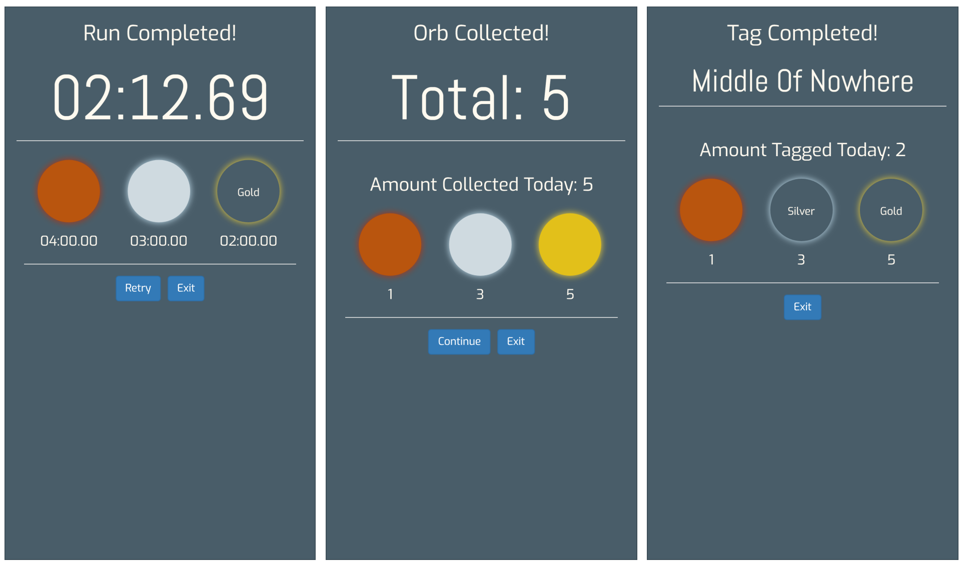


Figure 4.7: Screenshots of the Medals feedback.

Left: Time Trials, middle: Exploration, right: Crew Tagging

4.2.3.3 Runner Points

In the Runner Points feedback (figure 4.8), users earn virtual points for completing activities. Based on the amount of points they have accumulated in total, users will be assigned a current "runner level", and shown how many points are needed to reach the next level. To encourage continuous usage of the application, the first activity of the day earns users an additional 10 points, giving them an incentive to use the system more frequently.

Runner Points has been given a score of 1 for Challenges, Levels/Progression, and Points, and fits perfectly with users of the Player user type.

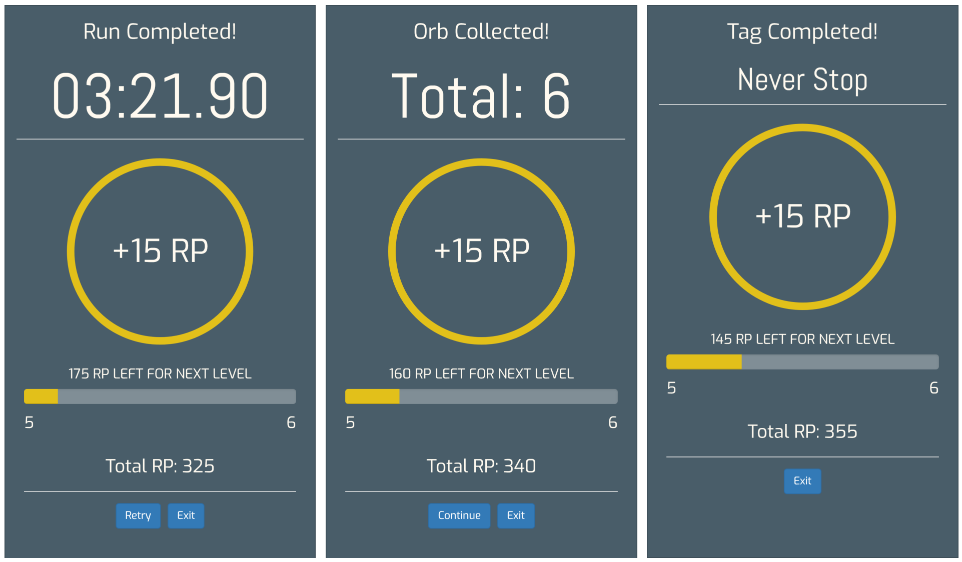


Figure 4.8: Screenshots of the Runner Points feedback.

Left: Time Trials, middle: Exploration, right: Crew Tagging

Game Element	**Leaderboard**	**Medals**	**Runner Points**
Badges/Achievements	0	1	0
Challenges	0.5	0	1
Collection and Trading	0	1	0
Customization	0	0	0
Exploratory Tasks	0	0	0
Guild or Teams	0	0	0
Leaderboards	1	0	0
Levels/Progression	0	0	1
Nonlinear Gameplay	0	0	0
Points	0	0	1
Social Comparison	1	0	0
Social Competition	1	0	0

Table 4.8: Scoring matrix for the feedback mechanics.

5
Results

This chapter is split into two parts: the first part presents the results of the project in the form of graphs and tables, whereas the second part recounts the key take-away points from the data.

5.1 Collected Data

5.1.1 Recommendation Results

The user type sample data was used to investigate the recommender system, and can be seen in its entirety in appendix C.1. Figure 5.1 shows the resulting recommendations for each activity, and figure 5.2 shows the same for the feedback mechanics.

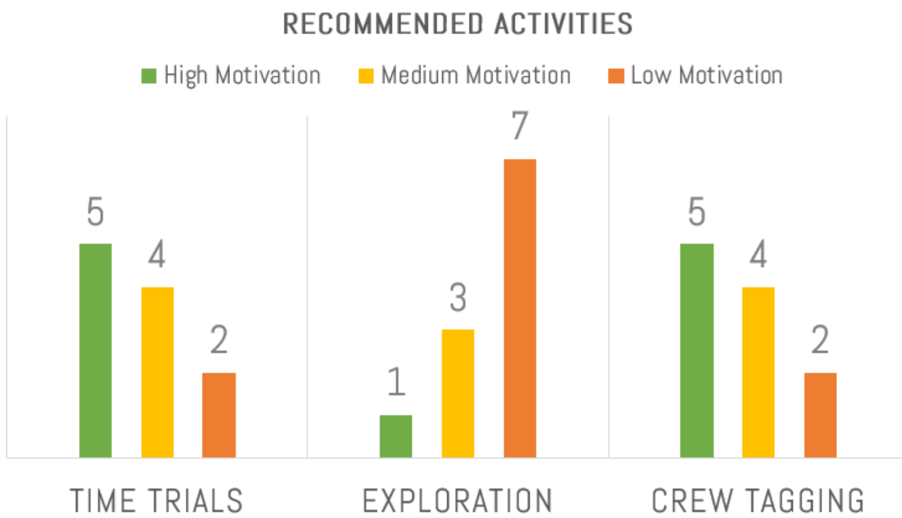


Figure 5.1: Distribution of recommendation of activities.

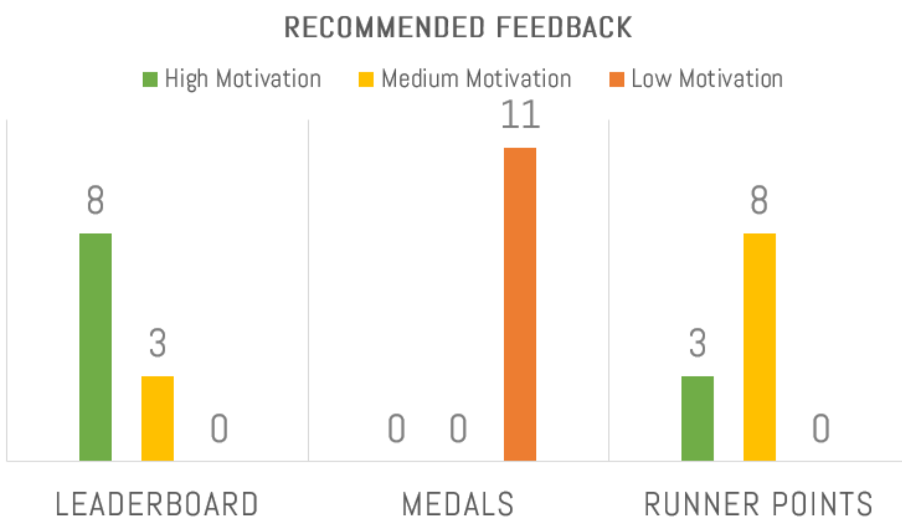


Figure 5.2: Distribution of recommendation of feedback.

Figure 5.3 shows the distribution of main Hexad types among the 11 participants. Please note that one participant scored equally among two types, therefore the total size of the sample is 12.

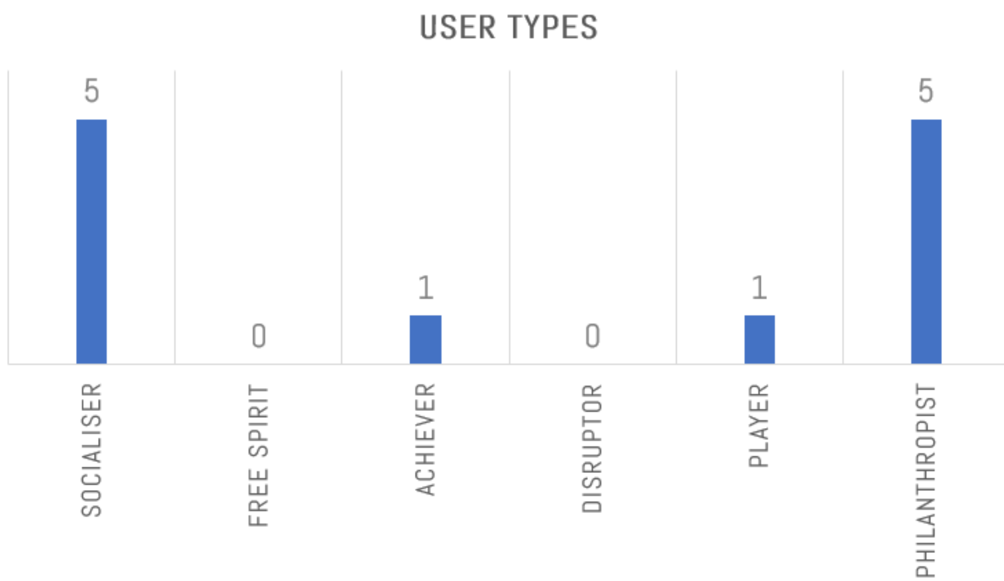


Figure 5.3: Distribution of user types from data sample.

To compare user types with recommendation, table 5.1 and 5.2 displays the activities and feedback mechanics that were **most** recommended in relation to the users' main Hexad type. Table 5.3 and 5.4 displays the activities and feedback mechanics that were **least** recommended.

	Time Trials	Exploration	Crew Tagging
Socialiser	2	0	3
Free Spirit	0	0	0
Achiever	1	0	0
Disruptor	0	0	0
Player	0	0	1
Philanthropist	2	1	2

Table 5.1: High motivation activities for main user types.

	Leaderboard	Medals	Runner Points
Socialiser	3	0	2
Free Spirit	0	0	0
Achiever	0	0	1
Disruptor	0	0	0
Player	1	0	0
Philanthropist	5	0	0

Table 5.2: High motivation feedback for main user types.

	Time Trials	Exploration	Crew Tagging
Socialiser	1	4	0
Free Spirit	0	0	0
Achiever	0	0	1
Disruptor	0	0	0
Player	1	0	0
Philanthropist	0	4	1

Table 5.3: Low motivation activities for main user types.

	Leaderboard	Medals	Runner Points
Socialiser	0	5	0
Free Spirit	0	0	0
Achiever	0	1	0
Disruptor	0	0	0
Player	0	1	0
Philanthropist	0	5	0

Table 5.4: Low motivation feedback for main user types.

5.1.2 User Testing

The survey answers collected during the test session with the three participants can be seen in table 5.5. The notes from the focus group can be found in appendix C.2.

	Tester A	Tester L	Tester J
User Type (Main)	Socialiser	Philanthropist	Achiever
Recommended Activity	Crew Tagging	Crew Tagging	Time Trials
Recommended Feedback	Leaderboard	Leaderboard	Runner Points
Discouraging Activity	Exploration	Exploration	Crew Tagging
Discouraging Feedback	Medals	Medals	Medals
How motivating did you find activity 1 (1-10)?	5	8	10
How motivating did you find feedback 1 (1-10)?	5	3	10
How motivating did you find activity 2 (1-10)?	10	7	6
How motivating did you find feedback 2 (1-10)?	9	9	8
Did you feel more motivated by being able to select the activity by yourself?	Yes	No	Yes
Did you feel more motivated by being shown all three types of feedback?	No	Yes	No
If you were to go for a run with the application now, which activity would you prefer?	Crew Tagging	All of them	Time Trials
If you were to go for a run with the application now, which feedback would you like to be shown?	Leaderboard	Medals, RP	Leaderboard, RP

Table 5.5: Survey answers from test session.

5.2 Analysis

5.2.1 Resulting Recommendations

Of the recommended activities, Time Trials and Crew Tagging ended up being recommended equally much, whereas Exploration was only recommended to one person. For 7 out of 11 people, the Exploration activity was ranked least motivating, whereas the rest of the users were evenly split between Time Trials and Crew Tagging.

The Medals feedback was not recommended to anyone, and ended up being the lowest ranking feedback in all 11-cases. The majority of users ended up with Leaderboard as their recommended feedback, with Runner Points only being suggested to 3 people.

Only taking the main type into account, the distribution of Hexad user types skewed heavily towards two groups: Socialiser and Philanthropist. Only one person was categorized as Achiever, and the same goes for Player. There were nobody of type Free Spirit or Disruptor.

Looking at the comparison of user type with the recommended activity, it is seen that the users of type Socialiser got recommended Crew Tagging in 3 out of 5 cases. This fits greatly with the intentions of the activity, which were aimed at delivering social play. Similarly, the one Achiever got recommended Time Trials, which delivers the competition and skill progression elements that are advised for this user type. The Philanthropists were almost evenly split between the three different activities. Because no users got Free Spirit as their main type, it is understandable that the Exploration activity was only recommended once.

In terms of feedback, the Leaderboard was recommended the most. All Philanthropist users got this recommendation, as well as 3 out of 5 Socialisers (even the one Player type, which otherwise appear to be more motivated by the two other feedback mechanics, was recommended this mechanic). Surprisingly, the Achiever, who were the one most-likely to be motivated by leaderboards, was not recommended this feedback but instead Runner Points.

On the other hand of the scale, by looking at the least recommended game elements, 80% of Philanthropists and Socialisers got the Exploration activity.

Seeing as Philanthropist is the only other user type than Free Spirit to have a correlation with Exploratory Tasks and Nonlinear Gameplay, this result is somewhat unexpected. The rest of the low-ranking activities somewhat correlated to the Hexad type. Medals was calculated to be the least motivating feedback mechanic for every user.

5.2.2 User Motivation

The survey results from the testing session showed different results. Two users found the recommended activity more motivating than the activity ranked least-motivating by the system, while the feedback worked opposite, with two users preferring the low-scoring feedback mechanic. Two out of three felt more motivated by being able to select the activity yourself, while being shown all types of feedback yielded the opposite results. When asked which feedback the user would liked to be shown, nobody chose to display all three.

Based on the information collected from the focus group, the general opinion was that the application motivated them towards running more. When discussing the ability to choose the activities yourself, it was a question of what was most motivating: the activity you liked the most, or having the system tell you exactly what to do. It appeared that all activities worked as intended, although the feeling of belonging to a team was somewhat vague when doing Crew Tagging. The feedback mechanics had different effects on each tester, but all agreed that the Leaderboard feedback could be improved by only showing your friends or closest competitors. Similarly, all three participants could see potential in a more story-driven activity.

Although no technical issues occurred during testing, users reported some usability problems. Most significant was the fact that you have to run and look at a screen, which can be quite difficult for people not used to this. It was suggested to implement the activities only using sounds, since the checkpoints of the routes could be learned and you only need indications of when you have triggered the next one.

6
Discussion

This chapter interprets the results of the project in an attempt to answer the research questions. It reflects on the efficiency of personalizing gamification, and covers both the benefits and challenges of using recommender systems in the process. Finally, a list of limitations for the thesis is presented.

6.1 Findings

6.1.1 Recommender System Design

The main challenge in designing the recommender system was balancing the score of each game element, so the appropriate recommendations were made for the different users. As can be seen from the test results, this was almost successful, except for the fact that the Medals feedback was never suggested to anyone. This is not a requirement for the system, but seeing as the sample data contained many different user types, it is likely that nobody will ever have this feedback recommended, which indicates that the scores have not been balanced properly. As covered in section 2.4.1, the process of assigning features can be quite difficult and is a known issue with content-based recommendations that could be solved by using collaborative filtering.

Apart from the balancing, the general personalization approach of using a recommender system functioned quite well. For simplicity's sake, it is worth considering cutting out the correlation matrix between user type and game mechanic. By using the Hexad user types instead of game mechanics for the recommendation vector, each game element simply needs to be scored compare to the different user types. The research of matching game mechanics to Hexad types could still be used, but instead of using the measured correlations directly, one could simply apply it to assign scores to game elements based on the user type ("From 1 - 5, how likely is an Achiever to enjoy this activity?").

In other applications, recommender system are mainly used on an extensive selection of item. Having a limited amount of activities, in this case only running, makes it challenging to truly personalize the recommendations, because the options are quite limited. However, it does allow for easier implementation of different game mechanics, as it may not be possible to combine every activity (example: jumping jacks) with every mechanic (example: exploratory tasks). The system could have been expanded further by including the user's fitness level or preferences for running (long runs or short intervals), but this would require each individual route to be scored on these two variables as well.

It's also important to ask the question: what is the goal of the recommendations? Recommender systems are great for exactly that, "recommending", and perhaps works best when used as a guiding hand rather than with the goal of customizing an entire application. By adding the personalization as a feature, either as "Recommended Activities" or in conjunction with activities that are procedurally generated, the user can themselves choose if they wish to utilize the capability of the system, instead of letting the developer make that choice.

6.1.2 Motivational Effects of Personalization

6.1.2.1 User Type Correlations

Considering the difficulties with balancing the recommender system, it could be interesting to remove the recommendations from the test and simply look at the user type: did the Hexad types correlate to which activity and feedback the testers enjoyed?

User "L" preferred Medals and Exploration, but scored only 18% in Player and 14% in Free Spirit. This does not correlate with previous research, which shows that the non-linear gameplay presented by the Exploration activity should mainly motivate users of the Free Spirit type. Similarly, the user was surveyed to be significantly of the Socialiser type (20%), but did not enjoy the Leaderboard feedback much, which otherwise offers both social comparison and competition. The Hexad type and the gamification preferences do not seem to correspond for this user.

User "A" preferred Exploration when running alone, but claimed that Crew Tagging would motivate him much more if other users of the system consisted of his friends (the low score for this activity in the survey was mainly due to lack of real life competitors - this user quickly realized that the data in the system was generated). The Leaderboard was ranked high in motivation for this user, and is usually aimed towards competitive people. Given that "A" scored 21% in Socialiser, both activity and feedback seem to match the Hexad type.

User "J" preferred Time Trials and was only somewhat encouraged by Crew Tagging. The medals did not motivate much either, but Runner Points greatly inspired to further exercise. Having a score of 22% in Achiever, this closely correlates to the competition and improvement elements of Time

Trials. Even though Medals and Runner Points have many similarities, it can be considered that the Medals were only motivating until the gold had been reached, whereas Runner Points would keep the user motivated, as there is no level cap. This would agree with the Achiever mentality of self-improvement, since the Medals feedback is missing a “carrot” once the user has completed enough activities to get the gold medal.

Overall, it appears that there can be a positive correlation between the Hexad user types and their expected motivational game elements. When making this comparison, it is suggested to not only look at the main user type, but to go deeper into the Hexad, as the tool is not meant to define a specific type, but instead to specify how much the individual identify with each type.

6.1.2.2 Negative Impacts

One of the key arguments for personalizing gamification is that users are being exposed to elements which have a negative impact on on their motivation. When choosing in which areas to apply the personalization, it is therefore important to contemplate: do personalizing of this game element have an effect on the user?

Being able to select between running activities seemed to be preferable over not being shown all of them, which is not that surprising since knowing about other activities doesn’t really appear to be demotivating. On the other hand, it would make sense to personalize the activities, if the persuasive strategy behind the application was to motivate users by specifically telling them what to do. In this case, it could be beneficial to select activities that are more likely to motivate them based on their profile. Therefore, it is not suggested that the user is blocked from participating in certain gameplay mechanics, unless they have been implemented as part of a personalized program.

The feedback seemed to have the biggest impact on the users' motivation. The individual preferences differed between the participants, but it was clear that demotivating feedback mechanics became an annoyance to the user, as they would much rather only see the elements they enjoyed. Furthermore, the test showed that even the individual mechanics can be personalized for better results. For instance, by only showing the closest competitors on a leaderboard, the user is constantly shown a goal that is just in range. Alternatively, players of different skill can be compared to each other by letting them compete on different parameters, but still be scored equally. If one was to implement an XP-system, where all users are ranked in the same way, yet give the same amount of points for different activities, the levelling mechanic can still motivate you, even if you are not the fastest or most active.

To summarize, personalization of gamification can be hard to decide on. There is no doubt that some game mechanics are much less efficient at motivating specific user types than others, but if given the option to select for themselves, most users will likely just avoid the elements that they don't find enjoyable. Instead, if you have to force your user through a series of tasks, applying your knowledge about that user to personalize the tasks can be a great way to ensure motivation.

6.2 Limitations

The implemented activities and feedback were not confirmed to be associated with the planned game mechanics. For instance, even though the Exploration activity was build with the intention to offer non-linear gameplay, this may only have been accomplished from the perspective of the developer. By consulting a group of gamification experts, it may turn out that the game elements did not achieve the expected experience of the user, which creates significant errors in the test results.

The size of the test group was quite small, and all results will have to be validated with more users utilizing the system simultaneously. With a bigger group of testers, it could be possible to implement the recommender system using collaborative filtering, and thus avoid the balancing issues presented earlier. Since each version of the application only was available for a very short amount of time, it would be preferable to conduct tests over longer periods (weeks or months). This would not only allow users to get more familiar with the different game elements, but could also provide interesting result on the efficiency of personalized gamification as a solution to provide long-term motivation.

During the testing session, it was observed that users appeared to have a tough time differentiating between activity and feedback. This could lead to scores in the survey being a mix of both activity and feedback, even though the two should be separated. In relation to this, the questions in the survey were not verified in any way, and may have been unclear to some test participants.

Finally, it is also worth noting that absolutely no attempts were made at adapting the user interface. A general layout was chosen for the application, but distinct user types may react differently to various appearances of the app.

7

Conclusion

This project set out to investigate the efficiency of personalizing gamification, specifically by implementing a recommender system in a practical application. Although the narrow testing sample size limits the validity of the research, there are still some key points to takeaway from this proof of concept.

Implementing a content-based recommender system requires a lot of effort to balance the system properly. The approach taken in this project was based on specific correlations found in previous research, which may be a step towards unnecessary complexity. To avoid the "personal" impact of having the developer of the application assign attributes to the game elements, collaborative filtering can be used alternatively, if the user base is large enough. However, one should always consider in which area the recommendations should apply, as suggestions or planned programs for the user appear to be most beneficial from employing recommender systems. No definitive claims can be made in regards to the efficiency of personalized gamifications in a practical application, but testing indicates that there are advantages of adapting gamification elements towards the specific user, maybe even with successful results of keeping the user motivated on a long-term basis.

Bibliography

[Bakkes et al., 2012] Bakkes, S., Tan, C. T., and Pisan, Y. (2012). Personalised gaming: A motivation and overview of literature. *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System*, pages 4:1–4:10.

[Balabanovic and Shoham, 1997] Balabanovic, M. and Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM, vol. 40, no. 3*, pages 66–72.

[Bartle, 1996] Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs.

[Bateman et al., 2015] Bateman, C., Nacke, L. E., and Mandryk, R. L. (2015). Brainhex: Preliminary results from a neurobiological gamer typology survey. *Health & Place, vol. 46, July 2017*, pages 328–336.

[Busch et al., 2016a] Busch, M., Diamond, L., Marczewski, A., Nacke, L. E., Tondello, G. F., and Wehbe, R. R. (2016a). The gamification user types hexad scale. *CHI PLAY '16 Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pages 229–243.

[Busch et al., 2016b] Busch, M., Mattheiss, E., Hochleitner, W., Hochleitner, C., Lankes, M., Fröhlich, P., Orji, R., and Tscheligi, M. (2016b). Using player type models for personalized game design – an empirical investigation. *Interaction Design and Architecture(s)*, pages 145–163.

[Böckle et al., 2017] Böckle, M., Novak, J., and Bick, M. (2017). Towards adaptive gamification: A synthesis of current developments. *Proceedings of the 25th European Conference on Information Systems (ECIS)*, pages 158–174.

[Böckle et al., 2018] Böckle, M., Novak, J., and Bick, M. (2018). Personalizing persuasive strategies in gameful systems to gamification user types. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 435:1–435:14.

[Deterding et al., 2011] Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". *MindTrek '11 Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 9–15.

[González and Adelantado, 2017] González, C. S. G. and Adelantado, V. N. (2017). Gamification and active games for physical exercise - a review of literature. *1st Workshop on Gamification and Games for Learning (GamiLearn'17).*

[Halko and Kientz, 2010] Halko, S. and Kientz, J. A. (2010). Personality and persuasive technology: An exploratory study on health-promoting mobile applications. *Persuasive Technology*, pages 150–161.

[Huang et al., 2017] Huang, H.-C., Wong, M.-K., Lu, J., Wei-Fan, H., and Teng, C.-I. (2017). Can using exergames improve physical fitness - a 12-week randomized controlled trial. *Computers in Human Behavior, vol. 70, May 2017*, pages 310–316.

[Johnson et al., 2016] Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., and Hides, L. (2016). Gamification for health and wellbeing - a systematic review of the literature. *Internet Interventions, vol. 6, November 2016*, pages 89–105.

[Knutas et al., 2017] Knutas, A., van Roy, R., Hynninen, T., Granato, M., Kasurinen, J., and Ikonen, J. (2017). Profile-based algorithm for personalized gamification in computer-supported collaborative learning environments. *Proceedings of the 1st Workshop on Games-Human Interaction (GHITALY 2017).*

[Marczewski, 2015] Marczewski, A. (2015). User types. *Even Ninja Monkeys Like to Play: Gamification, Game Thinking & Motivational Design*, pages 69–84.

[Monterrat et al., 2015] Monterrat, B., Desmarais, M., Lavoué, E., and George, S. (2015). A player model for adaptive gamification in learning environments. *Artificial Intelligence in Education*, pages 297–306.

[Nakashima et al., 2017] Nakashima, R., Sato, T., and Maruyama, T. (2017). Gamification approach to smartphone-app-based mobility management. *Transportation Research Procedia, vol. 25, 2017*, pages 2344–2355.

[Orji et al., 2013] Orji, R., Mandryk, R. L., Vassileva, J., and Gerling, K. M. (2013). Tailoring persuasive health games to gamer type. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2467–2476.

[Orji et al., 2017] Orji, R., Nacke, L. E., and Di Marco, C. (2017). Towards personality-driven persuasive health games and gamified systems. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1015–1027.

[Ricci et al., 2011] Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B. (2011). *Recommender Systems Handbook*. Springer.

[Sardi et al., 2017] Sardi, L., Idri, A., and Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-health. *Journal of Biomedical Informatics, vol. 71, July 2017*, pages 31–48.

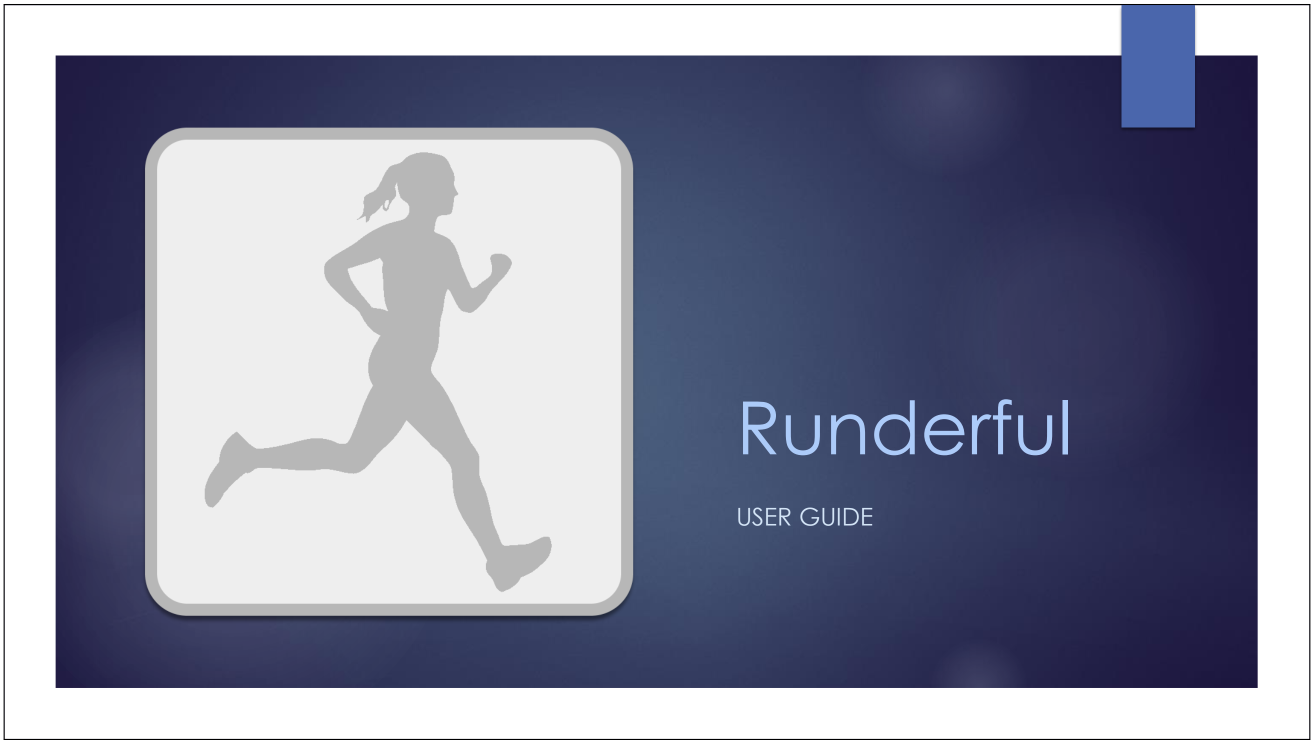
[Song et al., 2013] Song, H., Jihyun, K., Tenzek, K. E., and Lee, K. M. (2013). The effects of competition and competitiveness upon intrinsic motivation in exergames. *Computers in Human Behavior*, pages 1702–1708.

[Tondello et al., 2017] Tondello, G., Orji, R., and Nacke, L. E. (2017). Recommender systems for personalized gamification. *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 425–430.

Appendices

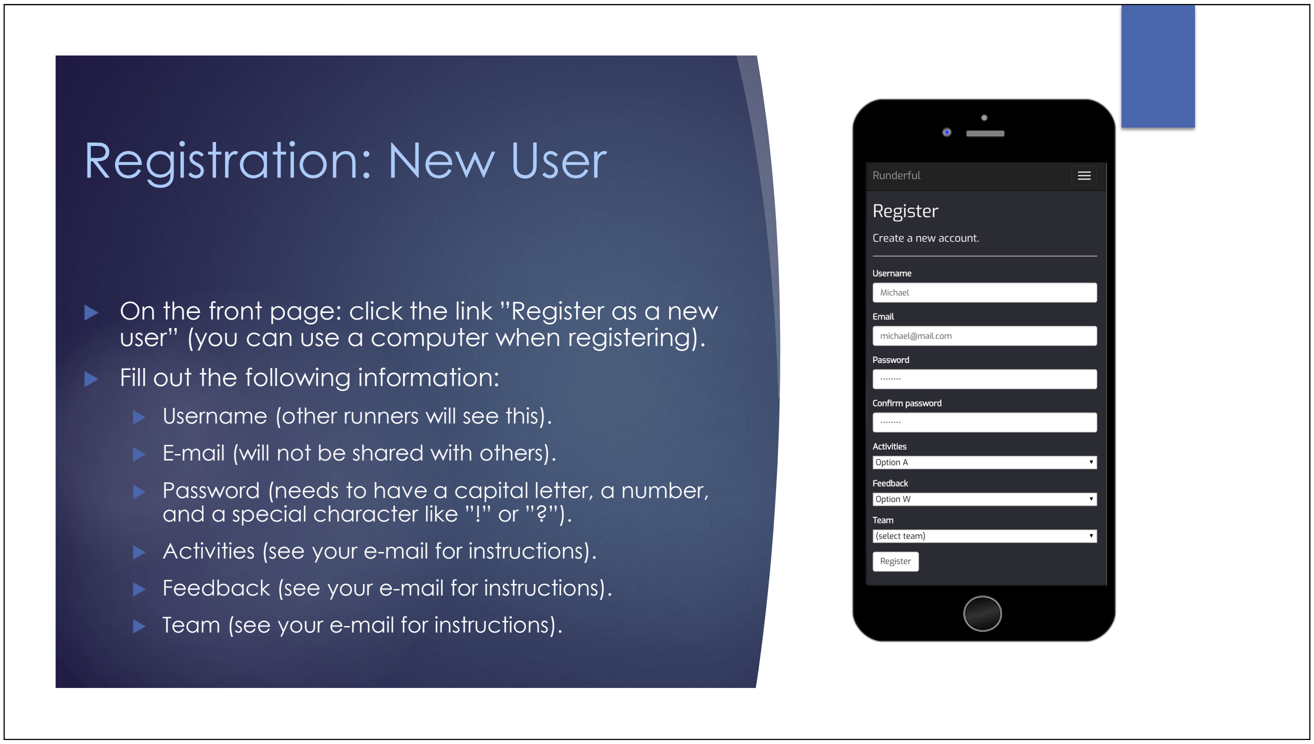
A

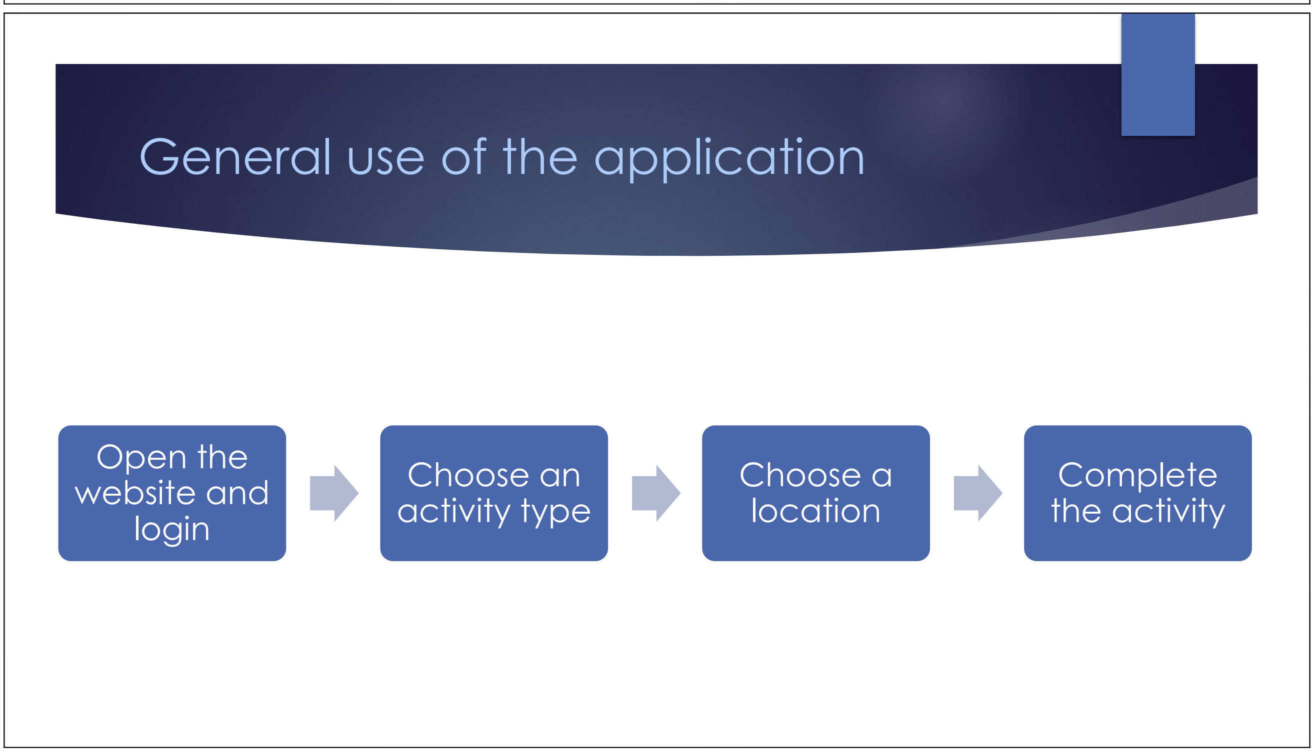
User Guide to Fitness App



Introduction

- "Runderful" – a GPS-based running app.
- Consists of various activities taking place in Copenhagen.
- To access the application, open a browser on your mobile phone.
- For the GPS to track your location, the web browser needs to be open druing use. If your phone locks, the position will be lost until you return to the application.
- Link: https://runderful.azurewebsites.net/.
- Using GPS requires a significant amout of energy, so make sure your phone battery is charged before starting a run.
- Internet access is needed for loading the activities, but does not require a lot of data (approx. 10 Mb for 1 hour of usage).


Registration: New User
On the front page: click the link "Register as a new user" (you can use a computer when registering).
Fill out the following information:
Username (other runners will see this).
E-mail (will not be shared with others).
Password (needs to have a capital letter, a number, and a special character like "!" or "?").
Activities (see your e-mail for instructions).
Feedback (see your e-mail for instructions).
Team (see your e-mail for instructions).
Runderful
Register
Create a new account.
Username
Michael
Email
michael@mail.com
Password
Confirm password
Activities
Option A
Feedback
Option W
Team
(select team)
Register


General use of the application
Open the website and login
Choose an activity type
Choose a location
Complete the activity

General use of the application

Open the website and login

Choose an activity type

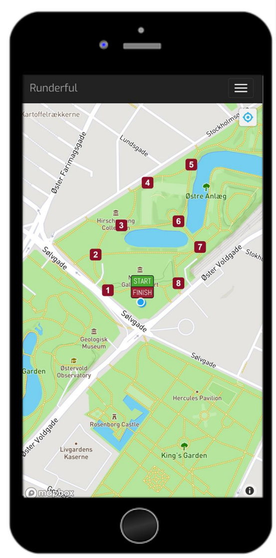
Choose a location

Choose a route

Complete the activity

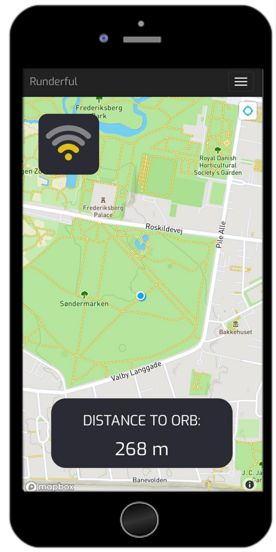
Activity: Time Trials

- Goal: Complete the route as fast as possible.
- Rules:
  - Click the "GPS"-icon in the top right corner to start tracking your location.
  - Go to the "Start"-icon on the map.
  - Once your location is in the start area, a "Start"-button will appear on the screen. Press it to start the race.
  - Run from start to finish by following the checkpoints on your map.
  - Complete the route as fast as you can, but be careful of your surroundings.
  - *Note: It may be a good idea to learn the route (either by running it at a slower pace, or looking at the map) before making an attempt at the time trial.*



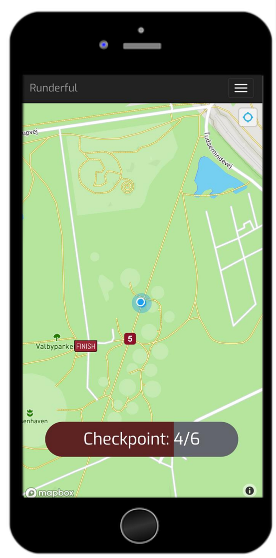
Activity: Exploration

- Goal: Find the hidden orbs by using the radar.
- Rules:
  - Click the "GPS"-icon in the top right corner to start tracking your location.
  - The radar shows your current distance to the orb.
  - Use the information to determine the location of the orb and move yourself to its position.
  - Once close enough (< 50 m), the "Collect"-button will appear on screen.
  - Press it to collect the orb and score a point.
  - *Note: since GPS is not always reliable, use the map to determine if your location is correct. If not, press the "GPS"-button to disable and then reenable location tracking.*

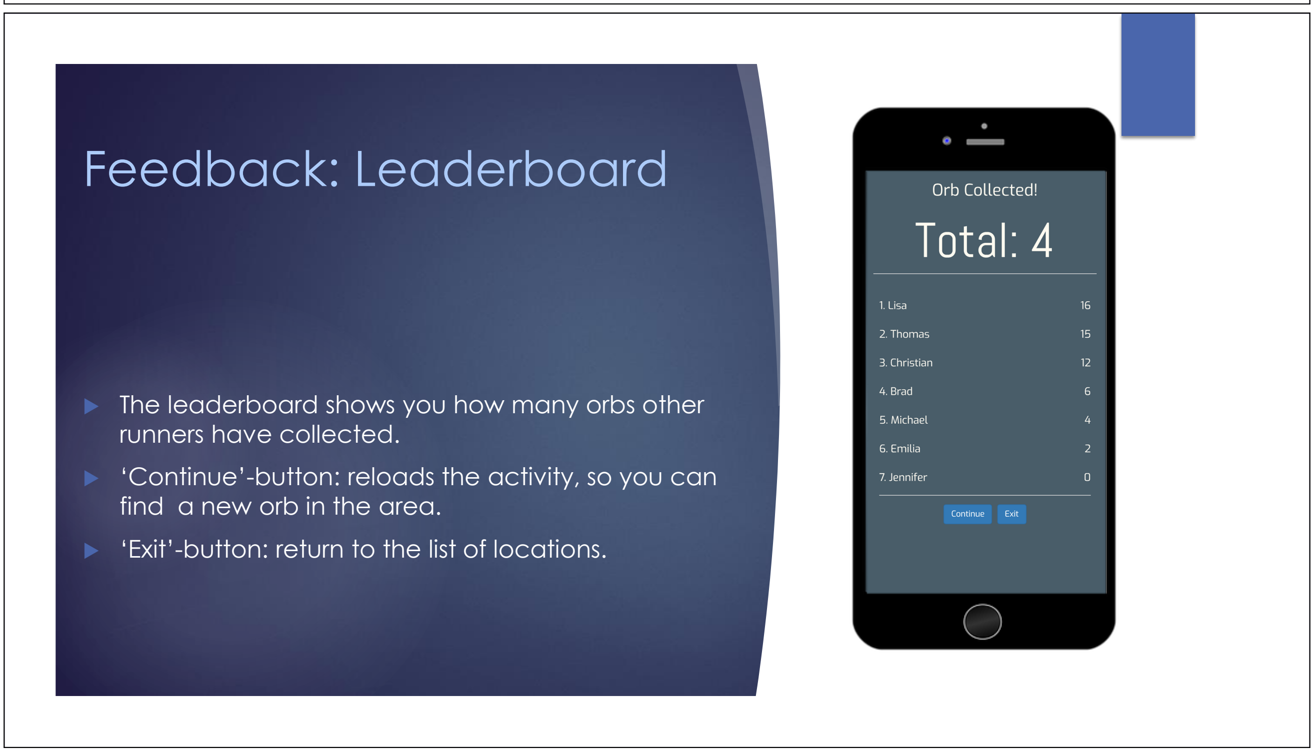


Activity: Crew Tagging

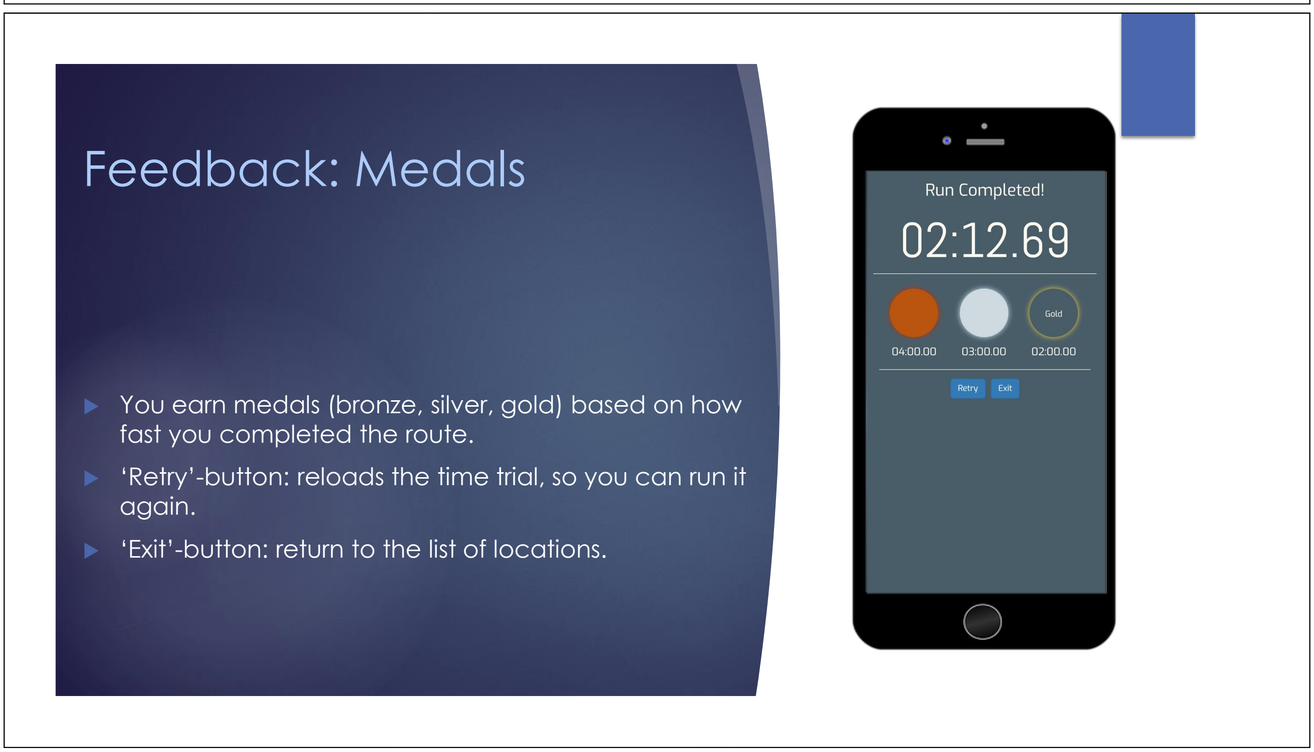
- Goal: Represent your team by "tagging" routes of the city.
- Rules:
  - Click the "GPS"-icon in the top right corner to start tracking your location.
  - Go to the "Start"-icon on the map.
  - Once your location is in the start area, a "Start"-button will appear on the screen. Press it to start the race.
  - Run from start to finish by following the checkpoints on your map. Time of completion does NOT have an influence.
  - Complete a route to tag it with your team's color.
  - *Note: Tag routes that are either untagged or tagged by another team to aid your team in taking over the city.*




Feedback: Leaderboard
The leaderboard shows you the completion times of other runners.
'Retry'-button: reloads the time trial, so you can run it again.
'Exit'-button: return to the list of locations.
Run Completed!
04:13.49
1. Thomas 02:58.60
2. Michael 03:15.59
3. Jennifer 03:44.60
4. Lisa 04:02.80
5. Brad 05:13.70
Retry
Exit

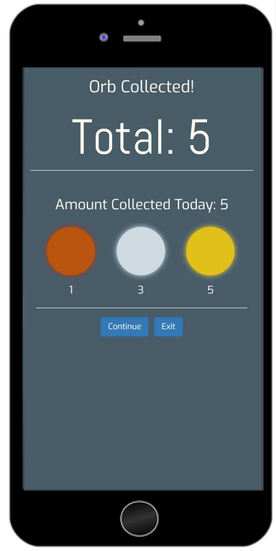

Feedback: Leaderboard
The leaderboard shows you how many orbs other runners have collected.
'Continue'-button: reloads the activity, so you can find a new orb in the area.
'Exit'-button: return to the list of locations.
Orb Collected!
Total: 4
1. Lisa 16
2. Thomas 15
3. Christian 12
4. Brad 6
5. Michael 4
6. Emilia 2
7. Jennifer 0
Continue
Exit


Feedback: Leaderboard
The leaderboard shows you how many of the routes have been tagged by each team (measured in a percentage of total routes).
'Exit'-button: return to the list of locations.
Tag Completed!
Energy Burner
1. Red Team
65%
2. Blue Team
25%
3. Yellow Team
5%
Exit


Feedback: Medals
You earn medals (bronze, silver, gold) based on how fast you completed the route.
'Retry'-button: reloads the time trial, so you can run it again.
'Exit'-button: return to the list of locations.
Run Completed!
02:12.69
Gold
04:00.00
03:00.00
02:00.00
Retry
Exit

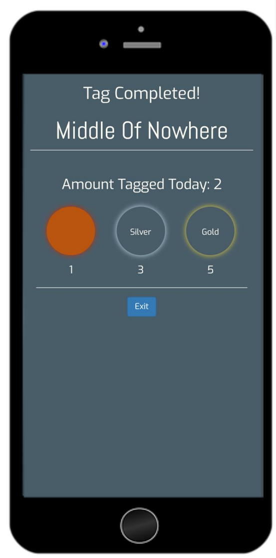
Feedback: Medals

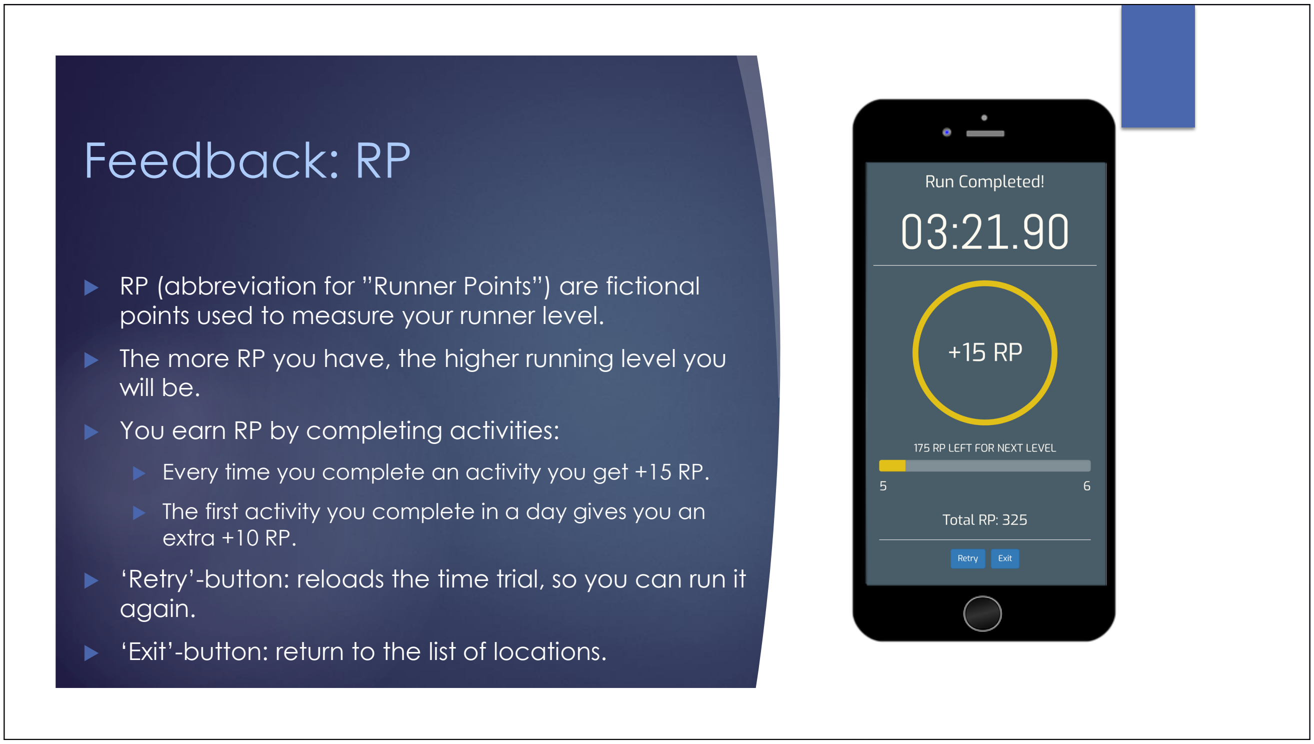
- You earn medals (bronze, silver, gold) based on how many orbs you have collected in a day.
- 'Continue'-button: reloads the activity, so you can find a new orb in the area.
- 'Exit'-button: return to the list of locations.

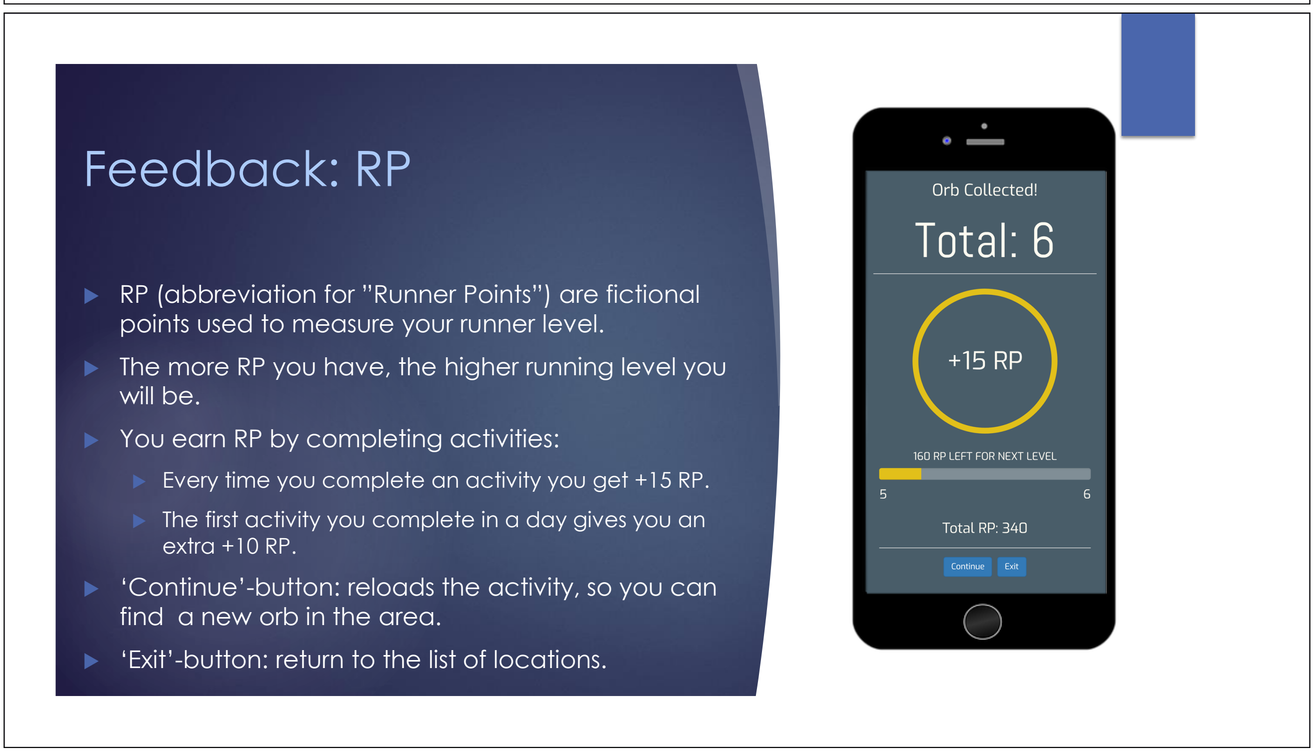


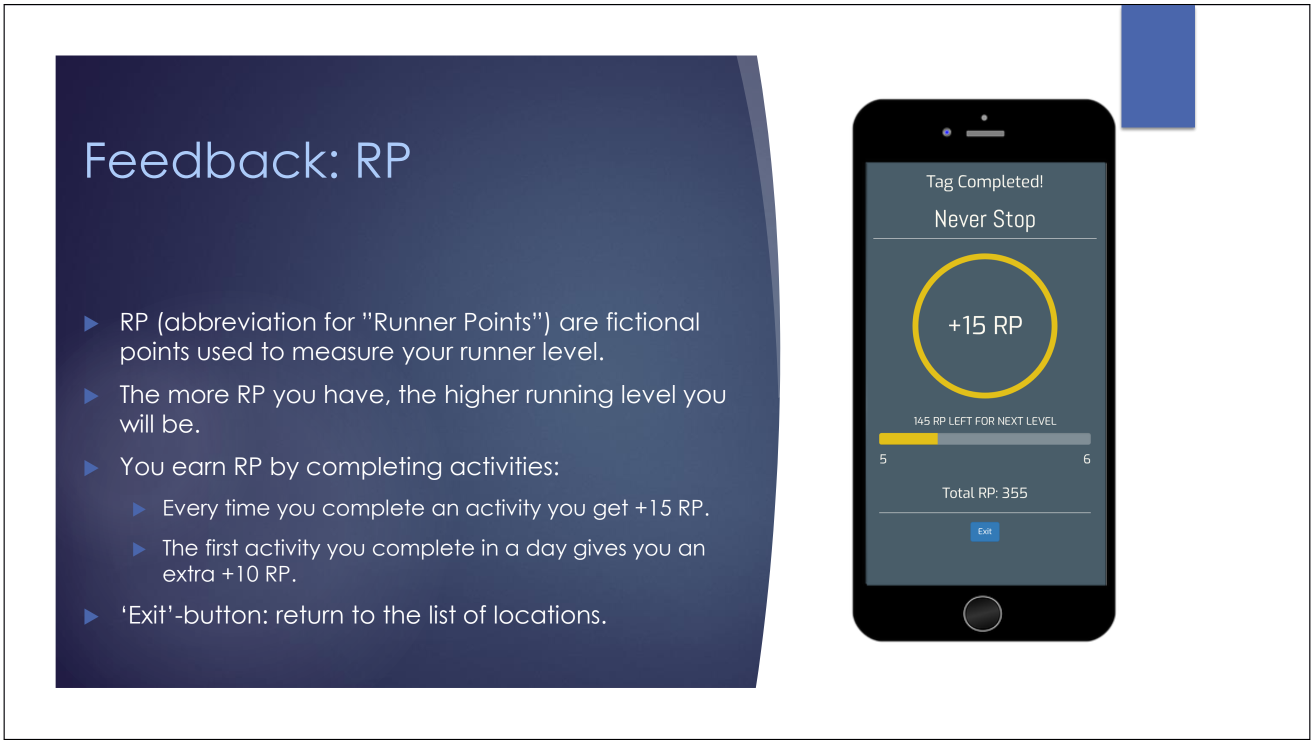
Feedback: Medals

- You earn medals (bronze, silver, gold) based on how many tages you have completed in a day.
- 'Exit'-button: return to the list of locations.




Feedback: RP
RP (abbreviation for "Runner Points") are fictional points used to measure your runner level.
The more RP you have, the higher running level you will be.
You earn RP by completing activities:
Every time you complete an activity you get +15 RP.
The first activity you complete in a day gives you an extra +10 RP.
'Retry'-button: reloads the time trial, so you can run it again.
'Exit'-button: return to the list of locations.
Run Completed!
03:21.90
+15 RP
175 RP LEFT FOR NEXT LEVEL
5
6
Total RP: 325
Retry
Exit


Feedback: RP
RP (abbreviation for "Runner Points") are fictional points used to measure your runner level.
The more RP you have, the higher running level you will be.
You earn RP by completing activities:
Every time you complete an activity you get +15 RP.
The first activity you complete in a day gives you an extra +10 RP.
'Continue'-button: reloads the activity, so you can find a new orb in the area.
'Exit'-button: return to the list of locations.
Orb Collected!
Total: 6
+15 RP
160 RP LEFT FOR NEXT LEVEL
5
6
Total RP: 340
Continue
Exit


Feedback: RP
RP (abbreviation for "Runner Points") are fictional points used to measure your runner level.
The more RP you have, the higher running level you will be.
You earn RP by completing activities:
Every time you complete an activity you get +15 RP.
The first activity you complete in a day gives you an extra +10 RP.
'Exit'-button: return to the list of locations.
Tag Completed!
Never Stop
+15 RP
145 RP LEFT FOR NEXT LEVEL
5
6
Total RP: 355
Exit

B

Recommender System

B.1 Selected Game Mechanics

	Socialiser	Free Spirit	Achiever	Disruptor	Player	Philanthropist
Badges/Achievements	1	0	2	0	2	0
Challenges	0	3	3	2	2	2
Collection and Trading	1	1	1	0	2	0
Customization	0	1	0	1	1	0
Exploratory Tasks	0	2	0	0	1	1
Guild or Teams	1	0	0	1	1	0
Leaderboards	1	0	0	1	2	0
Levels/Progression	1	2	2	0	2	0
Nonlinear Gameplay	0	2	0	0	0	1
Points	1	2	1	0	2	0
Social Comparison	1	0	0	0	2	0
Social Competition	2	2	1	2	2	0

Table B.1: Game mechanics and their simplified correlations used in the recommender system.

C

Test Data

C.1 Recommendations and Hexad User Types

	Activity			Feedback			Hexad
	High	Medium	Low	High	Medium	Low	Main User Type
User 1	TT	CT	E	RP	L	M	Soc. (19%)
User 2	CT	TT	E	L	RP	M	Soc. (20%)
User 3	CT	E	TT	L	RP	M	Pla. (21%)
User 4	E	TT	CT	L	RP	M	Phi. (22%)
User 5	TT	CT	E	RP	L	M	Soc. (19%)
User 6	TT	CT	E	L	RP	M	Phi. (19%)
User 7	CT	E	TT	L	RP	M	Soc. (19%)
User 8	CT	TT	E	L	RP	M	Phi. (22%)
User 9	CT	TT	E	L	RP	M	Soc./Phi. (21%)
User 10	TT	E	CT	RP	L	M	Ach. (22%)
User 11	TT	CT	E	L	RP	M	Phi. (21%)

Table C.1: Recommendations and main Hexad type of user sample.

TT = Time Trials, E = Exploration, CT = Crew Tagging

L = Leaderboard, M = Medals, RP = Runner Points

C.2 Focus Group Notes

C.2.1 General thoughts about the application

- Fun when running becomes play!
- Short routes are better ("I've completed 3 routes today").
- The application tells me what to do, so I don't have to consider where to run myself.
- Competing against each other is motivating.
- Have more activities to choose between is nice.

C.2.2 Tester J

- Time Trials was really great – it was nice to try and improve your own time.
- In Exploration, it was fun to actually explore the area.
- Crew Tagging did not work well when you were just a few people or alone. Maybe you could make team competitions lasting 1 week/-month?
- If the distance between me and my competitors is too big, my motivation will decrease. The same happens, when other people are able to see how little I run (if you are not an experienced runner, it's not nearly as fun).
- The Medals were least motivating for me ("I got a gold medal, what's next?"), whereas the Runner Points were the most fun ("I gotta get to the next level").

C.2.3 Tester L

- Exploration was a bit weird to begin with but worked much better once I completely understood it. Since you don't know where the next orb will be, there is always a moment of surprise.
- Crew Tagging was difficult to understand. The Leaderboard was a bit confusing, and I was more motivated by the progress bar, which was shown during the activity. Checkpoints gives a good overview of the progress throughout the route.
- "The routes make me run places I wouldn't originally run".
- If you were a sports team, you could use the application to make running practice more of a game.
- Going from Crew Tagging to Exploration to Time Trials was like a training session (the intensity is rising for each activity).
- "I didn't care about how fast SpartaMichael was, I would rather watch my own progress with medals".
- Regarding being shown all three types of feedback: it was nice to know that you weren't the slowest, but otherwise, I just look at the feedback I prefer.
- Runner Points could be motivating over longer periods of time.
- Some sort of story mode could be interesting. "Every park could have some fairy tale routes". "Collect these three things and save the princess". Whenever you reach a checkpoint or complete a route you would get a reward (a video, an animation, a smiley).
- Possible as both a running and walking application.

C.2.4 Tester A

- You completely forget about "the hard parts" of running.
- Crew Tagging was similar to shooting games. If you were to meet a couple of people and separate into teams, it would be easy to run against each other. Alternatively, instead of marking complete routes, you could just mark checkpoints/coordinates.
- Exploration worked great when you were running alone.
- Should you see the entire leaderboard, or just the users closest to you (or the daily/weekly best)? If it was just your friends, it would be very motivating (maybe with the possibilities of making groups).
- It could be an idea to compete on different parameters (running the fastest/longest/most).
- The Runner Points system could be expanded with stories and characters.

C.2.5 Did you have any issues using the application (technical, usability, etc.)?

- Too much battery drain.
- In Time Trials it is possible to improve your own score and continuously run the same routes, but Crew Tagging would need more routes in the same area for the routes to not be tagged too quickly.
- Watching a screen while running was quite challenging. Good when it is shorter distances, but long routes would be annoying.

C.2.6 Any improvements/comments/questions?

- Exploration could work using only sounds, where a sound would either get louder as you get closer to the orb, or a voice tells you how far away it is.

- It is possible to learn the routes, which means you only have to press 'Start' and 'Stop' (great if you are running with the phone in an armband).

- Being able to make your own routes, so you could compete against your friends could be cool.

- The app could suggest the next route or show the distance to the closest routes.

- Add a weekly/monthly running program to the application, which pushes the user to perform activities ("You have to do X, Y, and Z before going home").