

---

# IPP report

## Gamifying Spaced Repetition Software

---

s1700260

### Abstract

Spaced repetition software (SRS) aims help students learn new concepts by using flashcards. Current solutions lack schemes to motivate and encourage users to continue the learning process. Gamification is a technique that adds game principles and elements into existing software to increase user engagement. This report presents a scheme to gamify an existing SRS by integrating it with a current game. The ultimate objective is to improve user engagement and enhance the effectiveness of spaced repetition by providing a more appealing version of the SRS. The effects of the gamification will be evaluated using a methodology designed to test the application with groups of participants, collect usage data, and analyse the results to find characteristic behaviours of every group. A detailed set of tasks, execution plan and potential risks are provided.

### 1. Introduction

Learning requires a considerable amount of mental effort, which greatly depends on the perception of learners about the source of knowledge and the context of learning (Salomon, 1983). It follows that under conditions where people perceive the source of knowledge in a positive manner, the learning process requires less mental effort, therefore, it is easier, more pleasant, and provides better results (Salomon, 1984).

Traditional learning schemes have tried to include elements from new technologies with the objective of facilitating the learning process. One of such approaches is known as serious games (Michael & Chen, 2005). This alternative aims to take advantage of the pedagogical value of fun, competition and leisure to accomplish that goal. Serious games have demonstrated to provide benefits in the acquisition of new knowledge and skills (Graafland et al., 2012).

Serious games are narrowed to specific fields and contexts. Moreover, the considerable number of resources necessitated to develop such type of solutions makes it impractical to implement them in broader contexts. Under such circumstances, other alternatives have emerged to leverage the benefits of leisure and entertainment, but including characteristics of flexibility and adaptation. One of such approaches is known as gamification.

The main objective of gamification is to increase the user engagement of existing products, services or processes by including game elements and principles to make them more appealing. The range of scenarios of usage of this technique is wider and includes organizational productivity (Zichermann & Cunningham, 2011), physical exercise (Hamari & Koivisto, 2015) and learning (Hamari et al., 2016).

In the learning context, gamification has showed to be effective for different target audiences including primary school students (Boticki et al., 2015), undergraduate students (Slish et al., 2015), and the general public (DiSalvo, 2014). The types of learning products, platforms, and services that have implemented gamification are wide including mobile applications (Su & Cheng, 2015), websites (Gené et al., 2014) and desktop software (Cheong et al., 2013). These conditions ratify the flexible nature of gamification and its ability to be adapted to different environments.

Before the rise of technological techniques designed to leverage the benefits of good perceptions of learning contexts and sources of knowledge, there were other alternatives aimed to facilitate the acquisition of new knowledge. One of these techniques was oriented to exploit the spacing effect phenomenon (Hintzman, 1974) to increase the capacity of retention of learners about specific contents. This technique is known as spaced repetition.

One of the foundations of spaced repetition is the acquisition of new content by recurrent revisions in a series of short learning sessions scattered over fixed or variable intervals of time. These conditions increase the capacity of retention compared to the acquirement of knowledge in a single massive learning session. Thus, the difficulty of remembering new facts, concepts or definitions is diminished and mental processes like recognition and recall are improved.

Originally, spaced repetition is implemented with flashcards containing one or multiple related concepts. The flashcards are grouped based on how well the learner remembers their content. Then, the group of flashcards with more challenging content is presented to the learner more frequently. Flashcards move from one group to another as the learner progresses in each session.

Spaced repetition has also taken advantage of new technologies. Different pieces of software have been developed to make it more accessible through different platforms and interfaces. Each of these implementations has its own char-

acteristics depending on the target audience. Some solutions are focused on topics like learning new languages ([von Ahn Luis, 2018](#)), whereas others have broader and more general scopes ([Marcus, 2018](#)) and ([Damien, 2018](#)).

The current document presents a proposal to gamify an existing SRS. The ultimate objective is to create a more appealing version to increase the user engagement aimed to improve the effectiveness of spaced repetition. The proposed gamification scheme requires integrating the SRS and a game into a single piece of software. The integration can be achieved in two ways, which main difference is the context between the SRS and the game.

The integration of both components (SRS and game) will deliver two versions of the final application to be evaluated. The evaluation will require testing the application by groups of participants. The data collected will serve to make analysis, compare the effectiveness of every version, and discuss the relevance of the proposed approach. The following sections describe the design of the solution, the methodology for evaluation, and the tasks, risks, and execution plan.

## 2. Design

The usual way to implement gamification into existing solutions is to append game principles and elements throughout the piece of software. An example of this approach is the addition of a leaderboard of users based on points earned while using the application. Such leaderboard adds another motivation to use the application in order to reach the top of the classification. Overall, this method to implement gamification means that the entire user experience can be converted into a game context.

For the current proposal, the way to implement gamification is different to the traditional one. Rather than appending game elements into the SRS, the project aims to combine it with an existing game. Thus, two software components will be integrated as a single application. A brief description of those components and the details of the strategy to integrate them are given in the following subsections.

### 2.1. The spaced repetition component: AnkiDroid

Among all the available spaced repetition software only a few of them are open source. The most relevant option is Anki ([Damien, 2018](#)) which is distributed under the GNU Affero General Public License (AGPLv3) for most of its platform versions except iOS. Therefore, it is possible to modify and adapt the existing source code to new requirements.

Anki makes spaced repetition more accessible to a wider audience by providing web, desktop and mobile interfaces. Anki also provides a group of characteristics including different types of content in the flashcards (text, images, sounds), a multitude of decks compiled for several languages and various topics, and statistics of usage. It also provides features to customize the topics to learn.

The Anki version for Android devices is known as AnkiDroid. It provides an intuitive user interface that follows the best practices for mobile development. However, the lack of elements to motivate its usage might make it boring for some users, which makes it a good fit for the current project.



*Figure 1.* Example of AnkiDroid user interface. Image taken from the Google Play page of the application.

Anki implements spaced repetition based on the Super-Memo 2 (SM-2) algorithm ([Wozniak, 1990](#)). The original SM-2 algorithm defines the date for the next review of an item based on two elements: easiness and consecutive correct answers. Easiness refers to how easy to remember an item is; it is a numeric parameter that varies between 1.3 and 2.5, values that correspond to very difficult and very easy, respectively. The easiness of an item is not fixed, it is updated according to expression (1).

$$E' = E - (0.8 - 0.28q - 0.02q^2) \quad (1)$$

where:

- $E'$  = New easiness factor
- $E$  = Previous easiness factor
- $q$  = Quality of the response

The quality of the response is an integer value between 5 and 0. This parameter is assessed by the learner and measures how good was his response, being 5 a correct perfect response and 0 the worst possible incorrect response. The easiness factor is then combined with the number of consecutive correct answers ( $c$ ) to define the next date to display an item, as shown in expression (2). When an item is correctly answered, the base number of days for the next repetition is 6.

$$\text{Next date} = \begin{cases} \text{now} + 6E^{c-1} \text{ days} & \text{if correct} \\ \text{now} + 1 \text{ day} & \text{if incorrect} \end{cases} \quad (2)$$

The main difference in the SM-2 algorithm version implemented by Anki is the flexibility of the initial learning interval. As seen in expression (2), a new item can be displayed every day until the learner eventually remembers it. Anki understands that new items have to be displayed a number of times before the learner is able to remember them, but it does not punish him by showing failed items several times over a short period of days due to initial failures. Moreover, the learner can set the base number of days for correctly answered items; it does not have to be 6 for every learner.

In every learning session, Anki presents a set of cards to the user. Each card has two elements containing related information: the front and the back. For instance, cards used to learn a new language can have the words in the native language of the learner in the front, and the translations to the new language in the back. For every card, the application presents the front, the user analyses its content, and then the back is shown; later, the following card is displayed. The session ends when no more cards are left.

## 2.2. The game: 2048

Among the huge number of games available in Android, just a few of them are open source. The selected game for the current project is the popular 2048 ([Uberspot, 2017](#)). It is distributed under the MIT License, which is compatible with other licenses including AGPLv3. Therefore, it is possible to modify and adapt the existing source code to new requirements.

2048 is a sliding puzzle which ultimate objective is to merge numbered blocks until creating a one with the number 2048. The blocks are distributed over a grid of cells of variable size, but the default number of cells is 4x4. At the start of the game, two randomly positioned blocks holding the number 2 are displayed in the grid as seen in Figure 2.

A player has to slide the blocks horizontally or vertically. The blocks are moved in the chosen direction and they are stopped by the edges or when colliding with other blocks. If two blocks holding the same numbers collide, they are merged into a single block holding the sum of the values of the previous blocks. In every turn, a new block holding a power of two number is randomly positioned in the grid, being the power an integer between 1 and 10 as seen in expression (3).

$$\text{Number in block} = 2^x, x \in \{1, 2, \dots, 10\} \quad (3)$$

The game ends if one of the following states is reached:

1. A block holding the number 2048 is created.

2. The grid is full of blocks, none of them holding the number 2048.

The first state implies that the user won. Nonetheless, the user can continue playing to create blocks with numbers higher than 2048. The second state indicates that the user lost. Therefore, a new session has to be initiated to play again.

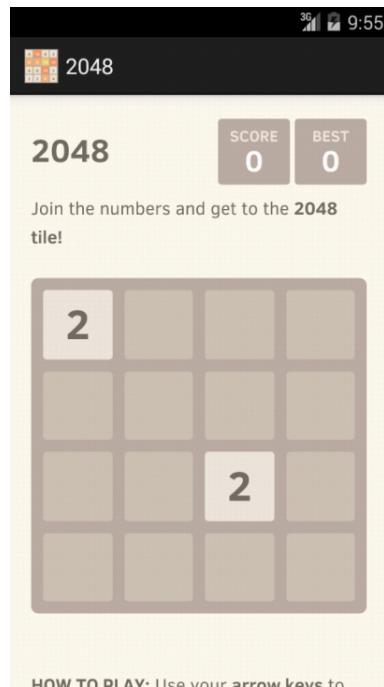


Figure 2. Example of the game 2048 user interface. Image taken from the Google Play page of the application.

## 2.3. Integration of Anki and 2048

Integrating two components into a single element poses a complexity in setting the balance between them. There are two strategies to cope with such situation. The first one simply combines the game and Anki in independent contexts. The second alternative uses a connection to create a dependency between the components.

### 2.3.1. INTEGRATION WITH INDEPENDENT CONTEXTS

In this option, using the spaced repetition component and playing the game will occur in interleaved stages, as seen in Figure 3. The implementation of this option would be straightforward, however, the lack of a real connection between both components might be a potential problem. First, it might be difficult for the user to understand which the main context of the application is. Moreover, switching between Anki and the game might be interpreted as an interruption of one over the other.

The main design decision for in option is to define the moment to switch the components. From the game, that moment will be reaching one of the ending conditions. In the other hand, switching from Anki will occur when the



Figure 3. Flow of spaced repetition and game components with no real connection between them.

learning session ends. This flow is fixed and it might make feel the users forced to do one of the activities to continue to the other. Alternatively, the users will be able to switch between components at any moment they want. Moreover, the users will have the option to skip the switching and continue using the current component.

### 2.3.2. INTEGRATION WITH CONNECTION

This alternative poses a clear definition of the main context of the application. In this case, one of the components will play the main role, and the other will be an addition. The integration will be done by creating a true connection between both components. The level of complexity of the implementation might be higher compared to the first option, but the potential benefits might be higher as well.

As described in the introduction section (1), when the source of knowledge is perceived in a positive way, the learning process requires less mental effort, and it is more pleasant. Based on this idea, it is safe to assume that playing a game is perceived more positively than executing spaced repetition. For this reason, the main context of the final application will be the game, and Anki will be the addition.

One of the objectives of the connection is to avoid the users interpreting the components as interruptions of each other. However, the main goal is to establish a relationship such the output of using Anki can be seen as a valuable resource to be used in the game, as seen in Figure 4.

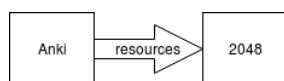


Figure 4. Flow of spaced repetition and game components with connection via resources provided from one component to the other.

Since the main context of the final application will be the game, Anki needs to be connected in a way such the users find worth using it. One alternative is to reward the users after using Anki. Then, the reward can be used during the game. Such rewards can be implemented as points to reach higher positions in a leaderboard, or as resources that can be used to ease the game.

A similar scheme was proposed in previous work (Johnson et al., 2012). There, the flow of the final application was split into two phases: points earning and game playing. During the first phase, the participants obtained points in two stages. In the first stage, the participants reviewed content using spaced repetition. The second stage consisted of a quiz where points were given based on the number of

correct answers. Then, those points could be used during the game.

There are other mechanisms that can be used to reward users and motivate them to use the spaced repetition component. However, for simplicity and ease of implementation, only the scheme of points earning will be taken into consideration. Therefore, the usage of the application will be split into two stages: points earning and game playing. The main difference with the scheme in (Johnson et al., 2012) is the flexibility of executing the points earning stage.

In the original scheme, the flow of the stages was fixed and the participants were able to earn points only at the beginning of each game session. For the current design, that flow is still valid and the points earning stage will be mandatory. However, the participants will also have another extra opportunity to earn points. They will be able to execute the spaced repetition component at any time while playing the game. The application might suggest that option to the participants, especially during challenging levels.

The number of points earned in the corresponding stage will be based on the easiness of each answered card; the more difficult the card, the higher the number of points. Then, while playing the game, its complexity can be reduced by exchanging the points with resources like the following:

- Remove a number of random blocks.
- Avoid the addition of new blocks for a number of turns.
- Allow new blocks with values higher than 2.
- Remove specific blocks.

Each of the available resources will be acquired by a number of points, which will depend on how much they ease the game; the higher the benefit of the resource, the higher its cost. Additionally, due to the simplicity of the original game, new features will be added. Those features will be initially locked; then, they will be available as the user progresses in Anki.

## 3. Evaluation

The evaluation process requires measuring the user engagement both versions of the final application. At first, user engagement can be seen as a subjective concept, thus difficult to measure. One first approach might be to perform a qualitative analysis of the application. To do so, the type of required information would be related to the perception of the user about the application. Such information might have the form of opinions, comments, and suggestions.

Collecting qualitative data can be expensive since it might require interviewing the participants after using the application. Alternatively, surveys or questionnaires could be utilised to collect that type of information. However, the

biggest problem is that this kind of information is difficult to analyse. Moreover, the information might be biased by the mood of the participants and other uncontrolled factors when collecting it. Under such conditions, it might be highly difficult to replicate the evaluation.

Some methods have been developed to cope with the difficulty of measuring user engagement. One of them is the development of a framework to measure it (O'Brien & Toms, 2010) by assessing parameters like perceived usability, aesthetics and felt involvement. Such assessment is performed through a series of surveys. Even though this option tries to minimize the problems related to collecting and analysing qualitative data, its main issue is that it is a general purpose framework that needs a considerable amount of information. Thus, it might be difficult to adapt it when the number of participants is small.

Due to the problems related to qualitative data, the evaluation of the current project will be done in terms of quantitative data. This will ease the collection of data and their analysis. Moreover, it will permit to replicate the evaluation or perform a new one with modified conditions to make comparisons. The details of the evaluation method will be described in the following subsections.

### 3.1. Participants

There will be two groups of participants. The first one will be the control group, the participants in this group will use the version of the application with independent contexts (2.3.1). The data obtained from this group will be used as a benchmark to measure the variation in user engagement. The second group will be the experimental group, the participants in this group will use the version of the application with a connection between the game and Anki (2.3.2).

The results obtained from the experimental group will be compared against the benchmark from the control group. The expectation is that there will be differences between both results. The analysis of those differences will allow drawing conclusions about the effectiveness of gamification for spaced repetition software.

The number of total participants will depend on the availability of subjects to participate in the evaluation of the final application. The expectation is to have between seven and ten people in every group. In any case, the number of members in both groups will be the same so the amount of collected data will be comparable.

AnkiDroid is a general-purpose implementation of spaced repetition. Hence, it can be used by any person that has access to an Android device. In fact, the published application in Google Play falls in the content rating *Everyone*, therefore, there is no restriction in the age of its users. Such situation means that the constraints in the selection of the participants are minimum.

Even though the minimum constraints in the selection of participants, it is important to maintain a level of homo-

geneity among them. Thus, the complexity of the analysis of the results from the control and experimental will be further reduced. A potential issue with this condition is the level of generalizability of the results. It might be difficult to extrapolate the results to other groups of the general population.

### 3.2. Quantitative data

Quantitative data are easier to collect and analyse. Collecting them will be done remotely by implementing features in the application such it can deliver statistics about its usage to a server. The type of data that will be gathered is wide and will come from Anki and the game. Examples of kinds of data include time using the game, time using Anki, number of correctly answer cards, number of earned points, and number of times the user skipped the switching. All this information can be analysed to find common characteristics of usage in every group of participants.

Eventually, there will be two parameters to study. The first one is *user engagement*, which requires to collect and analyse usage data of the application. The second parameter is *effectiveness of spaced repetition* when applied with gamification; this parameter requires to measure how much knowledge has been retained by the participants after using the application. The following definitions serve as a reference for the types of data that will be collected to measure both parameters.

- *User engagement*

1. **Time in session (TIS):** It is the time the users spend every occasion they use the application. It is measured in minutes per session.
2. **Frequency of sessions (FOS):** It is the time between sessions. It is measured in hours between two consecutive sessions.
3. **Frequency of spaced repetition features (FOSR):** It is the number of times spaced repetition features are used in every session. It is measured in number of times per session.

- *Spaced repetition effectiveness*

1. **Initial score (IS):** The score obtained in the initial test. It is measured as the number of correctly answered questions divided by the total number of questions.
2. **Score in session (SIS):** The score obtained in the quiz taken during a session of usage. It is measured as the number of correctly answered questions divided by the total number of questions.
3. **Final score (FS):** The score obtained in the final test. It is measured as the number of correctly answered questions divided by the total number of questions.

### 3.3. Hypotheses

The main objectives of the project are to increase the user engagement of AnkiDroid, and measure the effectiveness of spaced repetition when implemented with gamification. Therefore, the following hypotheses need to be confirmed or denied:

**H1:** The participants in the experimental group will spend more time using the application than the participants in the control group.

**H2:** The participants in the experimental group will obtain higher scores than the participants in the control group.

### 3.4. Analysis

The analysis of the collected information will be aimed to confirm or deny the hypotheses to some level of confidence. It will require making proper statistical analysis of the data from both groups. Initially, the analysis will be made using student's t-test and ANOVA, although other methods might be required.

## 4. Tasks, risks and schedule

The execution of the project is defined as a series of tasks or stages with specific goals. Each task has a duration and outcomes that might be required in following stages, as seen in Figure 5 and Table 1. Additionally, potential risks might affect the flow of the project in different levels as seen in Table 2. The next subsections describe the details of each task. Actual dates for the execution of each task can be found in Figure A.1 in the Appendix section.

### 4.1. Design

An overview of the general aspects of the design has been described previously (2). However, specific details about use case, application features, user interface components, and user interactions are still required. The objective of such details is to clearly establish the scope and context of the final application. Those details will be described in a specifications document.

The specifications document will serve as a guide to create two prototypes of the user interface on paper. The purpose of multiple prototypes is to reduce the effects of potentially missing important aspects that might affect the interaction between the users and the application. The prototypes will be presented to a few people to identify the elements in the design that make sense and those which do not. The results will be analysed and will serve to decide which characteristics of the prototypes are the best to combine them, make modifications and adjustments, and create the final prototype.

Additionally, a quick and rough evaluation of the existing interfaces of the game and Anki will be done. The purpose is to have a list of remarkable features and potential issues.

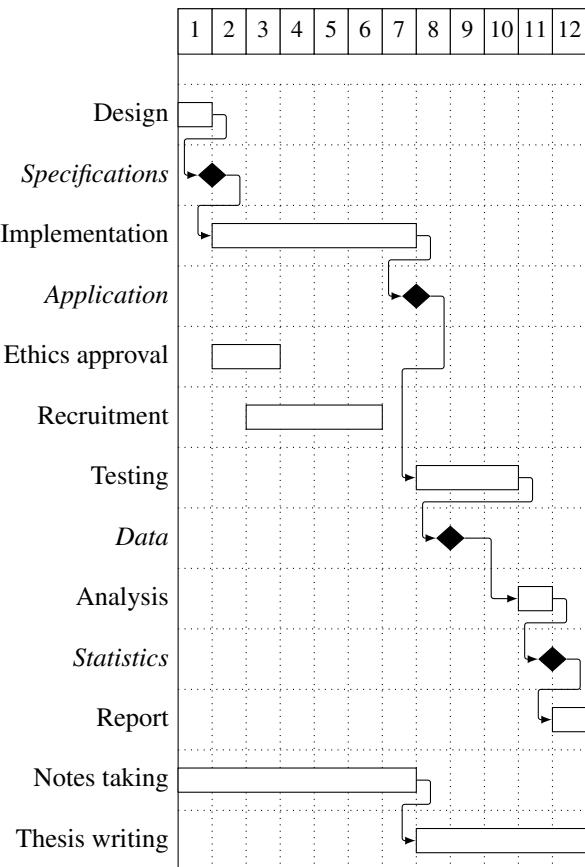


Figure 5. Gantt chart of execution of project tasks over a period of 12 weeks. Some tasks deliver milestones needed to start the following task.

Identifying those elements will serve to define which current elements will be useful in the integration and which ones might cause problems.

The execution of this task is critical since its outcome will guide the implementation and analysis stages. The specifications document will contain the details of the application from the perspective of the user, the types of information to be collected, and how they will be collected and stored. The outcome of this stage will be a final specifications document with detailed information about the features of the application along with sketches of the user interface. The duration of this task will be one week.

### 4.2. Implementation

This stage corresponds to implementing the application based on the specifications document. It is divided into iterative phases which will deliver updated versions of the application. Each new version will contain additional features and fixed bugs from the previous ones. The total number of phases will be five, and each one is expected to last one week. An additional iteration lasting one week will be needed to perform beta tests. The task will end once the beta tests have been executed and passed, and the final version of the application has been released.

Stage	Duration (weeks)	Outcomes
Design	1	Specifications document
Implementation	6	Final version of the application
Ethics approval	2	Approval from The School of Informatics
Participants recruitment	4	Signed consents
Testing	2-3	Data
Analysis	1	Statistics
Report	1	Report
Notes taking	7	Reference notes document
Thesis writing	5	Formal thesis

Table 1. Duration and outcomes of every stage of the project.

There are two potential risks: many features, and new bugs. The first one refers to the number of features and the time to implement them. Ideally, the features will be categorized by their importance and relevance to the application. Then, the most important ones will have to be implemented during the first iterations. If at the end of the fifth iteration some secondary features are not implemented yet, they will have to be discarded.

The second risk is the creation of new bugs when fixing previous ones or implementing new features. Features causing bugs might be discarded as long as their relevance is small. The effects of this risk can be minimized by the creation of test cases that will run every time a new feature is added or a previous bug is fixed. Alternatively, a test-driven development (TDD) strategy can be adopted. The creation of test cases can be time-consuming, therefore, it is necessary to keep a balance between them and the overall implementation process.

### 4.3. Ethics approval

Since testing the application will require the participation of humans, it will be necessary to obtain the corresponding approval from The School of Informatics. The objective of this stage is to make sure that the activities during the testing will follow the guidelines from the School Ethics Code and ethics regulations at the College and University, therefore, no harm will be caused to the participants. This stage will require to fill and submit a Level 1 Ethics Procedure form.

Additionally, the details of the project and the consent form to be signed by the participants will be provided. Both elements will be created out of the specifications document from the design stage. The purpose of the consent form is to make sure the participants understand the implications of the study. This task will need to be done early during the development of the project, ideally, after the specifications document is ready. The expectation is that the study will be approved within two weeks. If the study is not approved in the first submission, it is expected that feedback and recommendations will be provided to improve the next submission.

### 4.4. Recruitment

Ideally, potential participants will be contacted during the first steps of the implementation stage when a basic functional version of the application is already released. The purpose is to be able to provide screen-shots of the interface of the application. Thus, the potential candidates will have a rough idea of the look and feel of the application. The expectation is that this strategy will encourage potential candidates to participate in the study.

The actual participants will have to be contacted before the release of the final version of the application. It follows that the participants will be fully aware of their role, the objective of the testing, the duration of the testing period, and the information that will be collected. To do so, they will have to sign the consent form previously approved by The School of Informatics. This process is expected to last four weeks. The outcome will be the consent forms signed by the participants.

The potential risk in this stage is a small number of participants. This situation might pose a problem to collect sufficient data for the analysis stage. The effects of this risk are minimized by contacting the potential participants early in the development of the project. Moreover, the main way to recruit people will be in online communities like LinkedIn groups and Slack channels.

### 4.5. Testing

As mentioned, there will be two groups of participants to test the application: control and experimental. This stage will be conducted as a blinded experiment, thus, the participants will not be informed to which group they belong to. The purpose is to avoid potential biases due to the users knowing what the expected outcomes are. After signing the consent form, they will be given the resources to install the application.

The testing stage will last between two and three weeks. The actual duration will depend on the amount of data collected after the second week of study. During this time, the participants are expected to use the application on a daily basis to generate enough data. Daily notifications will be sent remembering to use the application. The data from the application will be collected remotely as long

Stage	Risks	Severity	Mitigation
Design	Miss important aspects	High	Multiple prototypes
Implementation	Too many features Bugs	Medium Medium	Categorization and prioritization Test cases & TDD
Ethics approval	Not approved	Low	Guidelines & feedback
Participants recruitment	A few participants	High	Early contact
Testing	Low usage Abandonments	High Medium	Notifications Extra participants
Analysis	Incorrect analysis	Low	Advice from experts
Report	Unproper conclusions	Low	Advice from experts
Notes taking	Delay in other stages	Low	Daily updates & weekly revision
Thesis writing	Unclear writing	High	Advice from experts

Table 2. Summary of potential risks in the project.

as the participants are connected to a free wifi service to avoid charges in their mobile service due to the use of the application.

The testing stage will require the participants to take an initial and a final quiz. Such quizzes will be taken within the application. The participants will be notified remotely about the quizzes. After completing the final quiz, the participants will be informed that testing period has ended and they will be free to uninstall the application. The outcome of this stage is a set of data collected while the participants used the application.

Participants might potentially abandon the study. Evidently, they will not be forced to continue in the testing. To reduce the affectation of this risk, an extra number of participants will be needed. However, the abandonment of participants will also be analysed to determine its causes. The application will have a feature to survey the users when they decide to leave the study. Participants will be encouraged to complete the survey before leaving the study.

Finally, the participants will use the application at any time during their daily activities. It follows that the application will be used in an uncontrolled environment. There are several potential issues for such condition. However, the most relevant one is that the user might forget to use the application, therefore not enough data will be generated. To diminish the effects of such situation, the application will send daily notifications to remember the users to use it. Evidently, the frequency of such notifications has to be moderated to avoid the users perceive them as intrusive.

#### 4.6. Analysis

The data collected during the testing stage will be used to perform analysis related to user engagement and spaced repetition effectiveness. Since the data will be gathered remotely, it will be possible to start the analysis before the end of the testing stage. The expectation is that in the middle of the testing stage there will be enough data that can be used to find trends or other characterizations in the usage of the application.

The analysis process can speed up by performing automated

routines to clean and pre-process the data on a daily basis. Once all the data have been collected, a simple processing will be required. The outcome of this stage will be a series of statistical parameters that characterize the usage of the application in the control and experimental groups. This information will be used to confirm or deny the hypotheses. This stage is expected to last one week.

The risk here is that the proposed analysis might not provide the results to confirm or deny the hypotheses. One alternative to reduce the effects of improper analysis will be looking for advice from experienced people in the field. Moreover, the vast amount of resources for data analysis will help to have a better understanding of the collected information.

#### 4.7. Report

The report will be aimed to interpret the results from the analysis data and identify possible causes for such results. Moreover, information about the limitations of the project will be provided. Either the hypotheses have been confirmed or denied, the report will provide the guidelines for further work in the same field or related ones. This stage is expected to last one week. A potential risk might be drawing improper conclusions. This issue can be minimized by getting advice from experts in the fields related to the project.

#### 4.8. Notes taking

This task will run since the beginning of the project until the start of the testing stage. It will provide a reference for the Thesis writing stage. The objective is to take notes about the execution of every task to understand how the project was developed including changes, problems, and solutions that might have found.

Notes taking could be time-consuming, therefore, it might affect the execution of the other tasks. The strategy to minimize those affectations is to add notes on a daily basis when possible. At the end of every week, a revision will be done. If is not possible to add all the details related to the activities done during the week, then the task will be skipped until the next week.

#### 4.9. Thesis writing

This task will provide a formal document with detailed information of the development and results of the project. Since the testing stage will depend mostly on the activities of the users, it is a good opportunity to start the writing of the thesis and avoid interruptions with the previous tasks. It will run until the end of the twelfth week. Once, the document is ready, it will be sent for revision of the supervisors. The feedback will be used to enhance the document and present its final version. An unclear writing might hinder the understanding of the document; advice from experts will be required to minimize such problem.

#### References

- Boticki, Ivica, Baksa, Jelena, Seow, Peter, and Looi, Chee-Kit. Usage of a mobile social learning platform with virtual badges in a primary school. *Computers & Education*, 86:120–136, 2015.
- Cheong, Christopher, Cheong, France, and Filippou, Justin. Quick quiz: A gamified approach for enhancing learning. In *PACIS*, pp. 206, 2013.
- Damien, Elmes. Anki, 2018. URL <http://apps.ankiweb.net/>.
- DiSalvo, Briana B Morrison Betsy. Khan academy gamifies computer science, 2014.
- Gené, Oriol Borrás, Núñez, Margarita Martínez, and Blanco, Ángel Fidalgo. Gamification in mooc: challenges, opportunities and proposals for advancing mooc model. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, pp. 215–220. ACM, 2014.
- Graafland, Maurits, Schraagen, Jan M, and Schijven, Marlies P. Systematic review of serious games for medical education and surgical skills training. *British journal of surgery*, 99(10):1322–1330, 2012.
- Hamari, Juho and Koivisto, Jonna. âIworking out for likesâ: An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, 50: 333–347, 2015.
- Hamari, Juho, Shernoff, David J, Rowe, Elizabeth, Coller, Brianno, Asbell-Clarke, Jodi, and Edwards, Teon. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54:170–179, 2016.
- Hintzman, Douglas L. Theoretical implications of the spacing effect. 1974.
- Johnson, Matthew W, Okimoto, Tomoko, and Barnes, Tiffany. Leveraging game design to promote effective user behavior of intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*, pp. 597–599. Springer, 2012.
- Marcus, Burnette. Flipquiz, 2018. URL <https://flipquiz.me>.
- Michael, David R and Chen, Sandra L. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.
- O'Brien, Heather L and Toms, Elaine G. The development and evaluation of a survey to measure user engagement. *Journal of the Association for Information Science and Technology*, 61(1):50–69, 2010.
- Salomon, Gavriel. The differential investment of mental effort in learning from different sources. *Educational Psychologist*, 18(1):42–50, 1983.
- Salomon, Gavriel. Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of educational psychology*, 76(4):647, 1984.
- Slish, Donald, Nash, Micael, and Premo, Joshua. Gamification increases scores of underperforming students in cell biology. In *EdMedia: World Conference on Educational Media and Technology*, pp. 870–876. Association for the Advancement of Computing in Education (AACE), 2015.
- Su, C-H and Cheng, C-H. A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, 31(3):268–286, 2015.
- Uberspot. 2048, 2017. URL <https://play.google.com/store/apps/details?id=com.uberspot.a2048>.
- von Ahn Luis, Hacker Severin. Duolingo, 2018. URL <https://www.duolingo.com/>.
- Wozniak, Piotr. Application of a computer to improve the results obtained in working with the supermemo method. *Text taken from PA Wozniak, Optimization of learning, Master's Thesis, University of Technology in Poznan, 1990.*
- Zichermann, Gabe and Cunningham, Christopher. *Gamification by design: Implementing game mechanics in web and mobile apps*. "O'Reilly Media, Inc.", 2011.

## A. Appendix

	2018													
	15-apr.	22-apr.	29-apr.	6-may.	13-may.	20-may.	27-may.	3-jun.	10-jun.	17-jun.	24-jun.	1-jul.	8-jul.	15-jul.
Design														
Implementation v1														
Implementation v2														
Implementation v3														
Implementation v4														
Implementation v5														
Beta testing														
Ethics approval														
Recruitment														
Testing														
Analysis														
Report														
Notes taking														
Thesis writing														

Figure A.1. Project plan with specific dates. Every column along the 2018 row represents a week. The header for every column is the start date of the week