FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Impact of Gamification on Subject Engagement and Data Quality in Health-care Surveys on Mobile Applications

Raul Viana

Mestrado em Engenharia Informática e Computação

Supervisor: João Correia Lopes

Second Supervisor: Gonçalo Gonçalves

October 12, 2022

Impact of Gamification on Subject Engagement and Data Quality in Health-care Surveys on Mobile Applications

Raul Viana

Mestrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

President: Prof. Rui Camacho
Referee: Prof. Benedita Malheiro
Referee: Prof. João Correia Lopes

October 12, 2022

Abstract

Gamification can be described as using video-game mechanics and elements in non-game contexts to improve user engagement and enhance playfulness, fun, competitiveness or other aspects. Its application is not limited to any particular discipline; it is used in fields related to education, workplace, business, sport, software and health.

Health studies may rely on surveys to collect data from subjects. These have the disadvantages of being very time and staff-consuming, along with the subsequent digitalisation burden. With the increasingly easier and cheaper access to mobile computing and wireless communications, new approaches arise. The surveys can be delivered inside mobile apps to be filled out by the subjects more conveniently.

Applying gamification features to health surveys can increase subject engagement. This would help improve the data quality and, therefore, help establish this mobile-based survey design more confidently. Also, it can probably speed up the studies' preparation phase and mitigate the data handling burden, increasing the number of published studies and knowledge produced.

The "Instituto de Saúde Pública da Universidade do Porto" (ISPUP), is developing a health study about chronic pain in adolescents. The data is being collected in an infrastructure explicitly developed for this kind of study, containing a mobile application through which the surveys are distributed. The population sample for this study comprises adolescents belonging to the "Geração XXI" cohort, randomized into two groups: gamified and non-gamified.

The data collected between 07/06/2022 and 24/08/2022 revealed the following results: a statistically significant higher *number of logins* in the gamified group and no statistically significant differences, in terms of subject engagement, between the two groups among the four *survey survival rates* and also on the several analysed data quality measures: *Likert variance*, *LongString*, *Non-Response* and *Open Answer Length*. Finally, the exploration of the gamification elements by the subjects fell behind expectations, with few purchased items on average and a high in-app coins balance after the completion of the fourth survey.

Within the limitations of this study, it is safe to conclude that the gamification elements could not induce subject engagement to enhance survey data quality. Therefore, the survey gamification implementation burden seems to overcome the potential benefits when applied in health studies. Even so, the implemented mobile application has been used to collect data for the ISPUP study through mobile phones, with all its inherent advantages.

Keywords: health pain survey, mobile app, gamification, engagement, data quality

Resumo

A gamificação pode ser descrita como a utilização de mecânicas e elementos de videojogos em contextos diferentes destes, de forma a aumentar o envolvimento dos utilizadores, bem como potenciar o divertimento, a diversão e o lado lúdico das atividades nestes diferentes contextos. A sua aplicação não está limitada a uma única disciplina; a gamificação é usada em diferentes áreas, relacionadas com a educação, ambiente de trabalho, empresarial, desporto, *software* e saúde.

A aplicação de elementos de gamificação a inquéritos de saúde pode potenciar o envolvimentos dos sujeitos. Isto poderá levar a um aumento da qualidade dos dados e, assim, ajudar a estabelecer este desenho de estudo como o *standard*. Ainda, poderá levar a uma diminuição do tempo da fase de preparação deste tipo de estudos e a uma diminuição dos gastos na logística para lidar com os dados, permitindo um maior número de trabalhos publicados.

O Instituto de Saúde Pública da Universidade do Porto (ISPUP), está a desenvolver um estudo de saúde acerca da dor crónica nos adolescentes. Os dados estão a ser recolhidos através de uma infra-estrutura especificamente desenvolvida para o efeito, contendo uma aplicação móvel na qual são distribuídos os inquéritos. A amostra deste estudo é constituída por adoslescentes pertencentes ao *cohort* Geração XXI, e foi randomizada em dois grupos: gamificado e não-gamificado.

Os dados recolhidos entre 07/06/2022 e 28/08/2022 revelaram os seguintes resultados: uma diferença estatísticamente significativa no número de logins entre os dois grupos, sendo que o grupo gamificado revelou um número de logins superior ao grupo não-gamificado. Não foram encontradas diferenças estatísticamente significativas entre entre os dois grupos nos *rácios de sobrevivência* dos quatro inquéritos nem nas diferentes medidas de qualidade dos dados analisadas: *variância do Likert*, *LongString*, *rácio de Não-respostas* e *rácio do Tamanho das Respostas Abertas*. Finalmente, a exploração dos elementos de gamificação por parte dos sujeitos ficou aquém das expetativas, havendo uma média bastante baixa de items comprados e um valor muito alto de moedas não gastas. Dentro das limitações deste estudo é seguro afirmar que os elementos de gamificação não induziram envolvimento suficiente nos sujeitos de forma a produzir um aumento na qualidade nos dados. Assim, os potenciais benefícios da aplicação da gamificação a inquéritos em saúde não parecem sobrepor-se aos encargos exigidos na sua aplicação. Ainda assim, a aplicação móvel implementada está a ser utilizada para a recolha de dados no estudo do ISPUP, com todas as suas inerentes vantagens.

Palavras-Chave: inquérito em saúde, aplicação móvel, gamification, engagement, qualidade dos dados

Acknowledgements

I am deeply indebted to João Correia Lopes for his constant sharing of knowledge, experience and contagiously good mood, and to Gonçalo Gonçalves for his unflagging support, attention to detail and commitment to the highest standards. Special thanks to INESC TEC for believing in me and receiving me so welcomingly.

I am also thankful to the ISPUP team, Dra. Raquel Lucas and Dr. Francisco Fernandes, for their support and superior statistical advisory.

I could not have undertaken this journey without my family: my wife and son. Patrícia and Guilherme were patient and understanding when I was absent and, at the same time, always pushed me to reach beyond.

Raul Viana

"If you're not failing every now and again, it's a sign you're not doing anything very innovative."

Woody Allen

Contents

List of Figures

List of Tables

Abbreviations

6D	6 steps for Gamification framework
API	Application Programming Interface
GiB	Gibibytes
INESC TEC	Institute for Systems and Computer Engineering, Technology and Science
ISPUP	Public Health Institute of the University of Porto
JSON	JavaScript Object Notation
MDA	Mechanics, Dynamics, Aesthetics gamification framework
MVC	Model-View-Controller architecture
RAM	Random Access Memory
RAMP	Relatedness, Autonomy, Mastery, Purpose gamification framework
REST	Representational State Transfer
S-O-R	Stimulus-Organism-Response gamification framework
SDT	Self Determination Theory gamification framework
SEPIA	Studying Experiences of Pain In Adolescents
TSE	Total Survey Error
UML	Unified Modeling Language

Chapter 1

Introduction

Most public health studies rely on gathering data from the subjects [13]. Online surveys are the most prevalent study design choice nowadays. It is crucial to assess the quality of this data and ensure a high standard. Gamified surveys can be a tool to increase subject engagement and data quality.

1.1 Context

This dissertation reports on the work developed in the context of a study already underway, which encompasses the Public Health Institute of the University of Porto (ISPUP) and the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), among other partners associated with the field of rheumatology. The study, designated Studying Experiences of Pain In Adolescents (SEPIA), intends to collect data related to pain perspectives and quality of life from adolescents through surveys in a mobile application. This mobile application includes five different surveys to be completed sequentially and provides the possibility to present or not to present the subject with gamification elements. The subjects included in the sample are recruited from the "Geração XXI" [1] cohort study and also appointed by the "National Association of Patients with Childhood Arthritis and Rheumatism" [2].

INESC TEC works as a technological partner that designs and develops the mobile application and manages the data during the data collection phase.

[1] https://www.geracao21.com/pt/

[2] https://www.andai.org.pt/index.php?lang=pt

1.2 Motivation

Researchers in the public health field need to make decisions about the design of their studies involving data collection. Traditionally, these studies are based on telephone or paper interviews and these methods demand many human resources responsible for laborious and expensive digitisation work [34].

The choice to distribute the surveys through a mobile application mitigates these disadvantages but can also create new problems related to participant engagement and the quality of the data collected. When answering a survey on a mobile application, subjects do not have the supervision they would have in traditional telephone or in-person interviews, which can lead to a certain lack of commitment. Without the interviewer's pressure to finish the fill-out, the probability of dropouts can rise and make the survey unfeasible [50].

The introduction of gamification features in the mobile application can mitigate these disadvantages, making this study design process faster and the logistics and cost of dealing with data digitisation no longer a deciding factor. Gamification can possibly enhance user engagement and produce higher completion rates, higher amount of answers, as well as increase their richness.

Furthermore, the recent pandemic crisis brought new challenges to the data collection processes. Subjects are less willing to attend crowded places or interact with strangers, and participate in these studies. A completely remote approach with data collection made through an application distributed on the subject's mobile phones, can mitigate this post-Covid conjuncture.

1.3 Objectives

This dissertation's overall goal is to assess whether the inclusion of gamification features in health surveys, conducted via a mobile application, increases subject engagement and data quality when compared to a non-gamified version of the same application. To achieve this goal and draw reasoned conclusions, three main objectives will have to be accomplished:

- **Build a mobile application with and without gamification features** – which will be distributed among the selected subjects and will provide the means to collect the data;
- **Identify and use the appropriate gamification features** – which will affect the subject engagement and the data quality level;

- **Measure the differences in the subject behaviour and data quality with and without gamification** – extract information from the data collected, apply data quality measures and compare the differences

The accomplishment of these objectives will result in a mobile application capable of collecting data in health studies, particularly in the current SEPIA study. Also, it will make it possible to conclude the degree to which the gamification influences the survey's data quality and its applicability in these kinds of studies.

1.4 Document Structure

In addition to the current Introduction chapter, this dissertation has five more chapters.

Chapter 2 describes the theoretical framework and analyses the related work. In Chapter 3, the problem is described, a possible solution is presented, and its implementation details are described. Chapter 4 presents the achieved results and the calculations made to reach them. In Chapter 5, the different achieved results are correlated and their implications are discussed and put in context with the bibliography. The study limitations are also discussed, and future work suggestions are raised.

Finally, Chapter 6 concludes the dissertation and directs for the future work yet to be done.

Chapter 2

Background and Related Work

Last chapter presented this dissertation's context, motivation and objectives and described its general document structure.

This chapter will examine the theoretical background and the closest related works, putting several concepts into perspective. It will begin by analysing the current shifting tendency from paper and phone interview surveys to the mobile applications delivery design. Then it will present gamification as a way to mitigate the most prevalent online survey design disadvantages and describe its broad applications, and specifically survey gamification. It will continue exploring the survey data issues and its several measurement possibilities that allow assessing the survey gamification effect. Lastly, the statistical methods to analyse this effect quantitatively are revised. Finally, the closest works to this dissertation are put in perspective.

2.1 Mobile Surveys Validation

The word "survey" is often used to describe a method of gathering information from a sample of individuals. Surveys have a wide variety of purposes, but they also can be conducted in many ways — including over the telephone, by mail or in person. Nonetheless, all surveys collect information through standardised procedures so that every individual is asked the same questions the same way. The survey intent is not to describe the particular individuals who, by chance, are part of the sample but to obtain a composite profile of the population [55].

The transition from the traditional phone or paper surveys to online/mobile ones comes with a set of advantages, but also with particular set-backs [16].

Gwalteney et al. [22] performed a meta-analysis review of over 65 studies, comparing electronic and paper survey outcomes. They concluded that the differences were neither statistically nor clinically significant and that the two types are equivalent.

Despite that, several studies show a consensual agreement that the online surveys present the same or better results related to data quality when compared to more conventional methods already referred [16, 26, 34, 46]. Due to their characteristics, the online surveys are also mentioned to have great efficiency and usability [16].

2.1.1 Advantages and Disadvantages

The collection of data through mobile applications has advantages for the subjects and researchers. Subjects report them being very comfortable and accessible, allowing answers to be given at any time most convenient to the subject [34]. For the researcher, the advantages are the removal of all the logistics and organization of telephone interviews or the handling of paper surveys, thus removing the burden of digitising the data collected and the cost of these logistics operations [28]. Also, the possibility of implementing automatic push notifications offers the potential to provide consecutive reminders or other information. [16]. Lastly, the reduction in research costs and the efficient survey administration in terms of time and resource management [38] are also significant advantages to take into account. Other benefits of mobile survey collection methods are the absence of interviewer bias, quick or immediate follow-up, unrestricted geographic coverage and fewer processing errors [12].

The main challenges and disadvantages are that the subjects are left to themselves to complete the survey without pressure to finish, except for the notification reminders. Compared to a telephone interview, this is the main disadvantage because the telephone or in-person interviewer can make sure the survey is brought to completion [50]. Jibb et al. [34] conducted a systematic review and meta-analysis on studies comparing electronic data capture versus conventional data collection methods. Among the 24 analysed studies, Jibb et al. found that 37 % concluded that the conventional data collection methods reported superior data completion, showing that the unsupervised survey filling out could lead to data quality issues. Also, issues related to the technical component should not be ignored, like poor data signal strength or uneasiness using smartphones, which can lead to a lack of motivation [16].

Other disadvantages are truthfulness concerns because it is impossible to carry out compatible verification methods; ethical and legal concerns related to data privacy and data storage providers; and selection and sampling bias since info-excluded subjects can be part of the studied population [12].

2.2 Gamification

Games are a crucial aspect of human culture and society and naturally promote motivation and engagement [9]. A game refers to a structured play with rules, goals and challenges for entertainment [37] and gamification refers to the use of game elements in non-game contexts [52].

The gamification concept, in general, is interpreted in many different ways by different people. There are supporters [14, 32, 24, 34, 39, 58], and disbelievers [8, 36, 61] of its usefulness and its processes.

Therefore, and because of the widespread use cases, there are different characterisations of gamification. "The use of gameplay mechanics for non-game applications, particularly consumer-oriented web and mobile sites, to encourage people to adopt the applications" [63], "features of an interactive system that aim to motivate and engage end-users through the use of game elements and mechanics" [56] or "the attempt to change the way of thinking and using some 'gaming rules' to increase the interest of learners and to solve problems" [11], or even "using game-like mechanics in non-game applications to gain specific behavioural acquisitions" [11].

Later, Bozkurt and Durak [9] expanded these definitions, adding a motivation purpose and completing the previous definition: the process of using game elements in non-game contexts to increase users' motivation to provide more effective, efficient, engaging, enduring and entertaining experiences. Furthermore, Landers et al. stated that due to its intrinsic modularity, it is trivial to study the relations between gamification and its outcomes in different conditions, building a gamification science.

Hamari et al. [24] concluded in their study "Does Gamification Work? — A Literature Review of Empirical Studies on Gamification" that gamification usage provides positive effects, even though they are greatly dependent on the context and the users using it. More specifically, the quantitative studies tend to conclude that positive effects exist only in part of the considered relationships between the gamification elements and the studied outcomes.

2.2.1 Applications

Gamification intends to invoke intrinsic motivations such as reward, pride, ego and competitiveness rather than hedonic experiences like audiovisual content or economic incentives [23]. Gamification has been widely used in different contexts, from education, crowdsourcing, social networks, health and data collection [53].

In education, gamified applications which make available reports, instant broadcasts and two-way messaging systems provide a way of tracking student development easily and better [11].

In health care, gamification has been widely used to motivate patients to increase treatment adherence by creating a therapeutic, enjoyable environment that eases the burden of illness, physical or cognitive training or to motivate health behaviour change for all ages [10].

In a crowdsourcing context, the organisations attempt to make the work activity more like playing a game to provide motives for working other than just the monetary compensation [48].

The application of gamification elements in all these contexts aims to bring "design patterns" of games to increase motivation for specific behaviours framed in each context [53].

According to Darejeh and Salim [14] the main application of gamification in software or websites is to change user behaviour, enhancing user engagement. This engagement can be measured as the time users spend on the software, either entering content or using that content. Because of the abstract nature of the engagement concept, researchers using online surveys usually rely on indirect measures to quantify it, such as time spent on the survey and the number of words answered in response to free-text questions [25] or nonresponse rates and straightlining [45].

2.2.2 Survey Gamification

The quest to fully engage survey repondents, ensure higher quality data, and increase participation is a high priority for survey designers. The gamification of surveys is an emerging trend from such exploration [2].

Because of these widespread gamification definitions, usages and applications, there have been attempts to unify the survey gamification process. These attempts focus on defining a clear, reproducible design strategy, represented as gamification frameworks [47].

Numerous gamification frameworks have been developed in the recent years, from the best-known "framework for Success" (SDT), "Six Steps to Gamification" (6D), and "mechanics, dynamics, aesthetics" (MDA), to the simpler ones like the "GAME" or the surrogate ones like "Relatedness, Autonomy, Mastery, Purpose" (RAMP) or "Core Principles" [47].

Triantoro et al. [58] used the "Stimulus-Organism-Response" (S-O-R) framework to gamify an online survey with positive results in terms of enjoyment without compromising the integrity of the data collected. The S-O-R framework permits articulating the

relationship between the stimuli and different types of organismic states, while signalling theory allows characterizing at the same time the nature of the stimulus when it is presented in gamified systems [58].

Harms et al. [26] proposed a structured process of survey gamification based on the MDA framework, also with good results, mainly from the designer's feedback. The MDA framework provides a way to understand game elements as either mechanics, dynamics, or aesthetics. Mechanics describe the basic building blocks (data representations, algorithms, rules, interactive features) that make up a game. Dynamics refer to the resulting run-time behaviour over time. Aesthetics characterize a player's emotional response and experience [32].

Wyss [62] concluded that it is not always possible to apply a selected gamification framework as a whole due to discrepancies in social and user factors. Some studies opted to just partially use the selected framework [25, 45, 62]. The inclusion of a few gamification elements allows researchers to isolate its influence and draw more meaningful conclusions [49].

According to Oliveira and Paula [49], the more frequent researched gamification elements in the engagement and in the academic research areas are the avatar, rules, time restriction and badges.

The gamification of online surveys has the underlying goal of increasing respondent engagement [20]. Questionnaire filling is a traditional tedious experience, and online surveys have been criticized for their dullness, resulting in negative respondent behaviour such as speeding, random responding, premature termination, and lack of attention [26]. On the contrary, evaluations of gamified surveys have reported diverse psychological and behavioural benefits regarding user experience, motivation, participation, amount and quality of data [26].

The gamification of mobile surveys and the concern of its contribution to data quality is patent throughout the consulted bibliography. Consequently, despite the referred advantages of survey gamification, there are issues one must consider. One major issue is that survey gamification often involves using techniques such as rewarding a question, changing response format, and adding additional visual elements, all of which inherently affect traditional data quality measures, such as response times, straightlining, and length of open-ended questions [35].

Embedding a survey in a game-like context can also alter the meaning of the underlying construct intended to be measured by the instrument [35].

Thus, the survey gamification must be implemented with precautions, and the outcomes must be measured accurately.

2.3 Data Quality

Modern society has adopted the survey method as a principal tool to analyse itself, using it as a "telescope on society" [27]. Survey research is not an academic discipline, with a common language, a common state of principles and a well-organised professional reference group [19]. The field of survey research has evolved through the years from the contribution of somewhat independent and uncoordinated efforts of researchers trained as statisticians, psychologists, political scientists, sociologists, and public health specialists, among others. These bridges between the survey methods and bodies of theory have produced the knowledge about survey quality available nowadays [19].

Blasius [7] described survey data quality as an umbrella concept that covers three primary sources affecting the trustworthiness of any survey data: the study architecture, the institutional practices of the data collection agencies, and the respondent behaviours.

The total survey error (TSE) paradigm refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data [6]. Heeringa et al. [27] describes it as the collection of all sources of error that can influence the precision and accuracy of survey estimates of population parameters.

Decomposing the error in smaller and smaller components may facilitate handling and prevention. Biemer [6] proposes the TSE to be divided into sampling errors, sub-divided into sampling scheme, sample size and estimator choice and non-sampling error sub-divided into specification, nonresponse, frame, measurement and data processing. According to Biemer [6], measurement error and nonresponse are the more damaging type of errors regarding data quality. These errors may arise from interviewers, survey questions, interview factors and respondent behaviours. As the respondent's behaviour is one of the critical factors influencing errors and consequently data quality, shaping it could be an indirect form of managing the data quality.

Schacht [54] studied the influence of using game design elements in increasing data quality in web surveys. The experimental design comprises presenting subjects with one of two survey versions: non-gamified or gamified with badges and a meaningful story. Schacht [54] considered several indicators to assess survey data quality, which are listed in Table 2.1. Other studies with close methodologies and similar objectives also used similar data quality indicators [17, 44] and are also included in Table 2.1.

As survey gamification applies essential changes to the pre-formatted and already validated survey, it is essential to measure whether these changes produce the desired outcomes of increasing the engagement and if they don't produce unwanted ones, like changing the measuring target and the validation. Therefore, a set of data quality measurements

Table 2.1: Data Quality Indicators

Mavletova [44]	**Schacht [54]**	**Galesic [17]**
Completion rates	Voluntary continuing with the web survey	Response times
Response order effects	Item non-response	Item response rates
Non-substantive responses	Rounded answers to numeric questions	Length of open answers
Level of reporting of socially undesirable responses	Straightlining	Variability of answers to questions in grids
Length of open answers	Length of answers to open questions	–
Response time	Speeding	–
–	Consistency items	–

can evidence the survey gamification outcome: response times, length of open-ended answers and straightlining [35]. The response times are the time intervals taken to answer each question. Longer response times could suggest more attention and commitment. The same applies to the length of open-ended answers, with longer answers indicating a greater subject's willingness. Straightlining refers to the subject rushing through the survey, always choosing the same answer.

Other measurements can be mentioned, such as completion rates and non-substantive answers [44]. The completion rates are related to the number of completed surveys and the non-substantive answers, with subjects choosing more straightforward answers of type "don't know" or "cannot recall". Finally, the scale reliability and the scale validity, asserting if the survey measures what it intends and if it does so accurately [57].

2.4 Quantitative Analisys

The research field often requires going beyond the descriptive analysis of the data. Inferences about the qualities or characteristics of the population that the sample represents are recurrent. The use of inferential statistics allows concluding about a population based on the findings of a sample of that population [1].

Inferential statistics are based on probability theory, which examines repeated events and eventually establishes a pattern. Their results are used to find judgments on the probability of obtaining a set of results - the decision-making process of hypothesis testing. Therefore, unravelling the meaning of a study [1].

Table 2.2: Statistical Tests Examples, adapted from Adams and Lawrence [1]

Study design	**Parametric Test**	**Non-parametric test**
Correlation/No groups	Pearson's *r*	Ordinal data: Spearman's rho
Two independent groups	Independent-samples *t-test*	Nominal data: Chi-square Ordinal data: Mann-Whitney
Two related groups	Related-samples *t-test*	Nominal data: McNemar test Ordinal data: Wilcoxon test
Multiple independent groups	One-way ANOVA for independent groups	Ordinal data: Kruskal-Wallis test
Multiple related groups	One-way ANOVA for related groups	Nominal data: Cochran Q test Ordinal data: Friedman test

There is a wide range of statistical tests to choose from, each having its characteristics and use constraints. Nevertheless, the choice of the best test for each situation is critical [1, 21]. An interesting group of statistical analysis methods termed *survival analysis* is worth mentioning. Pohar and Stare [51] defined them as the study of a particular variable, representing the time between entry to a study and some event of interest. This method estimates the cumulative survival function. This function is the probability of being alive at time *t*, in case the final event is death [51]. Researchers usually associate survival analyses with long-term patient survival studies, such as oncology studies, comparing different cancer treatments [65]. However, according to Young et al. [65] survival analysis methods help analyse any time-to-event data. The event of interest, traditionally death, can be replaced with any endpoint that occurs at a particular time and can occur only once [65]. Therefore this event can be defined as the completion moment of the survey fill-out.

The statistical tests can be divided into several groups in different ways. Table 2.2, adapted from Adams and Lawrence [1], presents a simple and concise analysis of the most common statistical test choice constraints. The study design is the primary factor to account for, with the number of analysed groups and the relation between them defining what kind of tests could be applied. Then the normality of the dataset dictates the use of parametric or non-parametric tests, and finally, its type. Being metric, ordinal or nominal narrows the possible choices one step thinner. Data transformations can further add another layer to the decision process.

Parametric and non-parametric statistics, for most practical purposes, can be differentiated on the data distribution assumptions. Non-parametric statistics do not rely on assumptions about the shape or the form of the probability distribution from which the data were drawn [29]. Although non-parametric statistics have the desirable property of mak-

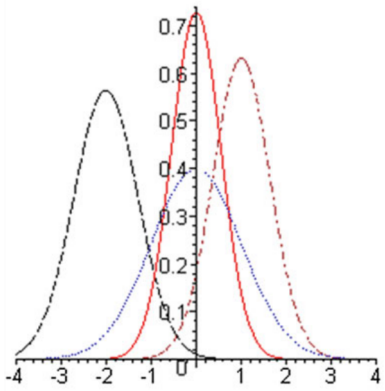


Figure 2.1: Plots of the Normal Distribution with Different Values of Mean and Standard Deviation, adapted from Ahsanullah et al. [3]

ing fewer assumptions and being very practical, they have a significant drawback [29]. This drawback constitutes a huge advantage to parametric statistics. It is related to their power, which reduces the Type I errors, meaning reducing the probability of rejecting the null hypothesis when it is true [31, 29]. Thus, parametric procedures should be chosen over non-parametric ones every time their assumptions are met [1, 31].

Ahsanullah et al. [3] refers to the normal distribution as one of the most important statistical distributions. It describes a family of functions of continuous probability distributions, having the same general shape and differing only in the mean and standard deviation. Figure 2.1, adapted from Ahsanullah et al. [3], depicts the plots of several normal distributions with different means and standard deviations.

The data's normality can be assessed through a graphical exploration based on histograms or box plots or through formal testing [64]. Of the different formal tests for normality available, Shapiro-Wilk test is considered the one with the best power across all types of distributions and sample sizes [5, 64].

As mentioned before, the choice of the statistical procedures depends on the study design and the data characteristics. Therefore, not surprisingly, the bibliography shows some inconsistency in the chosen statistical procedures, except for the almost unanimous agreement on the non-parametric approach. Studies analysing the gamification effects of surveys tend to use non-parametric tests to test the means, like Mann-Whitney and Wilcoxon. Harms et al. [25] used the Mann-Whitney test to compare independent groups that filled out two survey versions — gamified and non-gamified. Harms justified the Mann-Whitney choice with the non-normality of the data. Corey [13] used Wilcoxon test to compare independent groups that filled a public health online survey with gamified and non-gamified versions. Although Corey worked with a sample size of 91 subjects, the author does not justify the non-parametric choice, so it is likely that the data did not

follow a normal distribution.

Studies analysing the survey data quality tend to use the chi-square test arbitrarily. Goldammer [18] used the chi-square test to analyse several indices that could represent careless questionnaire responses with a total sample size of 360 subjects. Also, Mavletova [44] used the chi-square to compare data collected through mobile and personal computer surveys. Mavletova analysed completion rates, non-substantive responses and length of open answers. In both cases, the authors transformed the collected data, and that is the reason to chose the chi-square test. Finally, Heerwegh [28] compared some data quality indices between face-to-face and web surveys among 410 subjects. Heerwegh used the chi-square test in the data quality indicators and the Wilcoxon test to assess social desirability items.

2.5 Related Work

The closest work to consider is the Dores [15] Master dissertation. It was developed under the same research grant as the current work and had similar objectives. Both contribute to the ISPUP study, and although independent, both can be aggregated as a set because some of Dores' work transited to this dissertation, namely the mobile application and back-end bootstrap and early development.

While Dores [15] explored some different aspects not contemplated in the current work, like pain, pain management and pain evaluation, it also discussed the pain management mobile applications and their underlying surveys. Dores [15] focused on the initial bootstrap of the mobile application, the choice of the technologies and the gamification elements.

This work is the resumption of Dores [15] work, with the focus redirected to the survey gamification degree of engagement and its effectiveness, measured as an outcome of the data quality.

Corey [13] presented another closely related work, analysing the gamification of online surveys in public health data collection. From the premise that surveys have many issues, especially a large number of questions and take a long time to complete, Corey [13] then designs a study, introducing badges in online surveys to measure differences in completion rate and engagement. He concluded that gamification doesn't increase participant completion rates but also that participant engagement is higher in the gamified survey. Lastly, he found it inconclusive that the two survey versions' straightlining rates differ.

Corey's work conclusions may have limitations by introducing just one gamification element, namely the badges, which could be insufficient to introduce significant changes

in the two survey versions — gamified and non-gamified.

Another closely related study, "Web Surveys Among Children and Adolescents: Is There a Gamification Effect?" by Mavletova [45] proposed to measure the break-off rates, the error measurements and the nonresponse error in survey answers. Mavletova focused on children between the ages of 7 and 15. Her sample had 1050 participants, which were divided into 549 for the text-only survey, 472 for the visual enhanced one and 570 for the gamified one. The gamified survey elements included a narrative — travelling in the Antarctic and helping penguins; rules composed by a time limit and receiving points and rewards alongside visual feedback. Mavletova found that the gamified survey version produced no changes in the break-off and response variance rates. However, this version had a higher overall item nonresponse rate and completion time and a lower level of straightlining. She concludes that designing gamified surveys can be helpful to children and adolescents mainly because it was less burdensome and more accessible and enjoyable.

The above-referred studies represent the global inconsistency shown by the bibliography. Oliveira and Paula [49] in their systematic literature mapping, concluded that the quantitative outcomes of the survey gamification are very heterogeneous, with inconsistent results equally distributed between "significant", "good", "low" and even "negative". At the same time, the overwhelming majority concluded that the qualitative outcomes like "improved experience", "improved understanding" and "positive feedback" were significantly higher due to the gamification effect.

2.6 Summary

This chapter describes the theoretical framework of this dissertation's core themes – the mobile surveys, the gamification concept, the mobile surveys gamification, and the data quality.

The reviewed bibliography indicates that mobile surveys are the preferred design choice in public health studies but still have disadvantages. These disadvantages can be mitigated by applying gamification elements to the surveys. At the same time, survey data quality is an obvious public health researcher concern, and survey design choice must provide and assure data quality high standards. Finally, the most appropriate statistical tools must be applied to interpret the collected data correctly.

The next chapter will present the conceived solution and describe the implemented infrastructure and the study preparation.

Chapter 3

SEPIA – A Mobile Application for Health Care Surveys

The previous chapter brought in the theoretical background of this dissertation's central concepts, creating the foundation for the upcoming proposed solution. This chapter will present the proposed solution and describe the implementation phases. It will start by raising the problem statement and describing the mobile application that will address it. It will then continue with a more formal presentation of the infrastructure requirements, describing the actors, the user stories, the domain model and the technology stack used. The gamification elements introduced in the mobile application and the data quality metrics adopted are then presented and substantiated. Finally, the implementation phases are described, and the integration of the proposed solution with the ongoing ISPUP study is described.

3.1 Problem

As described in section 2.1, the ongoing shift from phone and paper surveys to online and mobile ones has to be done with orderliness. Nevertheless, adding gamification features to the mobile application could lead to an increase in data quality. This will help establish the mobile-based survey design with more confidence, mitigating the data handling burden. Comparing the usability metrics of the mobile application with and without gamification features will support/not support that:

> **A mobile survey with gamification features guarantees the same data quality as other methods of data collection through the stimulation of subject engagement**

The solution detailed in the following sections tries to address this question.

3.2 Proposed Solution

The conceived solution includes designing, implementing, and distributing a mobile application with and without gamification elements to collect survey answers.

This solution also encompasses the measurement and analysis of the collected data. Specifically, the quality data indicators that will possibly represent the differences in engagement and data quality between the gamified and non-gamified versions. This was accomplished by developing a mobile application and a back-end environment acting as a Representational State Transfer (REST) application programming interface (API). The mobile application aims to collect adolescents' and their caregivers' pain experiences through surveys.

To that end, the mobile application will present the subjects with five surveys with different questions, specially designed for adolescents and their caregivers, totalling ten different surveys. Both adolescents and their caregivers will have sequential access to their first four surveys. The moment a survey is completed, the next one becomes immediately available. The fifth and last survey stays locked until six months after the beginning of the first one. The survey's questions will use pre-defined types that can be re-used with different content: Likert scale, open answer, body chart, radio and checkbox, among others. Notifications will be set to trigger at specific time points defined by the ISPUP team. The caregiver will have access to his/her adolescent's progress information, namely the progress percentage in the current answering survey and the last time they were online. Only some adolescents will have access to the gamified version of the mobile application, which includes an avatar, quizzes, badges, a shop and in-app coins (these elements are described in Section 3.4).

Figure 3.1 shows several screenshots taken from the latest mobile application version. Figure 3.1a shows the login screen, Figure 3.1b one of the tutorial pages — the first avatar configuration, Figure 3.1c shows the timeline with the several surveys to answer and their status and finally, Figure 3.1d depicts a page from the first survey, with body gender choice question.

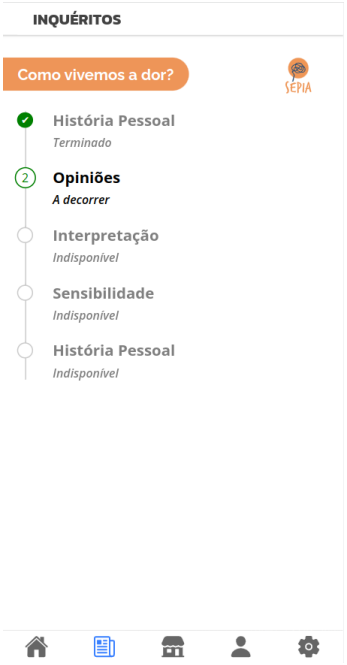
All these features are intended to be designed abstractly enough, so framing them effortlessly for different studies in public health or other data collection areas could be done quickly.

3.3 Requirements Analysis

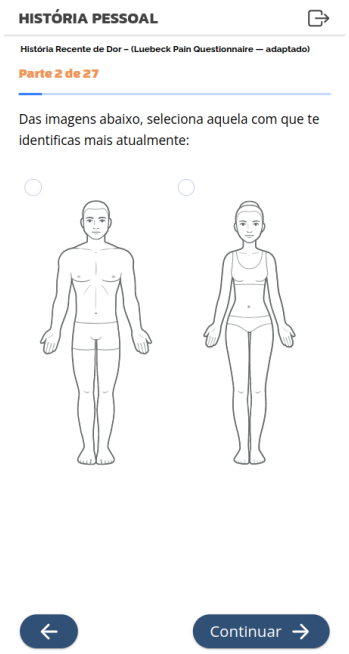
The requirements were gathered through several meetings with the ISPUP team. They have changed several times throughout the process, mainly the logic of the flow between

(a) Login Page

(b) Tutorial Page



(c) Timeline Page



(d) Survey Page

Figure 3.1: SEPIA Mobile Application

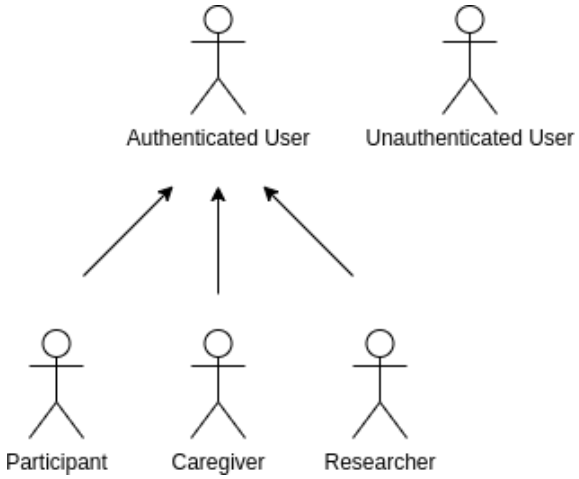


Figure 3.2: System's Actors

surveys and their internal structure. The first proposed survey flow logic was quite complex and enforced a minimum of seven days of delay between each survey for the first four surveys. The fifth survey would have a delay of six months from the completion moment of the first one. This structure was later dropped due to high drop-off concerns. It was replaced by a continuous sequential flow, without any delay, for the first four surveys. The opening of the fifth survey maintained the six months delay counting from the completion of the first survey. The solution prospect had to consider all these constraints and the user requirements. These latter are composed of several actors interacting with the system and how they interact with it, namely their user stories.

3.3.1 Actors

An actor[1] in the Unified Modeling Language (UML) specifies a role played by a user or any other system that interacts with the system. Actors may represent roles played by human users, external hardware, or other subjects.

The interaction with the system will be made through the different actors shown in Figure 3.2, each one with different permissions associated:

Unauthenticated User *Unauthenticated User* represents the visitor that will perform login in the system, becoming an *Authenticated User*.

Authenticated User *Authenticated User* will perform system log-out and hold account state information. The *Participant*, the *Caregiver* and the *Researcher* actors inherit all properties of the *Authenticated User* actor, which is a generalization of the other three.

[1] https://www.omg.org/spec/UML/2.5.1/PDF

Caregiver *Caregiver* and *Participant* represent study subjects and are actors of type *Authenticated User* who can interact with the mobile application features. *Caregiver* represents a caretaker study subject responsible for other subjects, interacts with the mobile application filling out surveys and controls his/her *Participant*'s progress.

Participant *Participant* represents study subject, interacts with the mobile application filling out surveys and takes advantage of the gamification elements.

Researcher Lastly, the *Researcher* represents the study researchers that have access to a web management portal, described in Section 3.6.

3.3.2 User Stories

User stories[2] were used to describe the software system solution features in an informal and natural language.

They are a popular method for representing requirements using a simple template such as *"**As a** (role), **I want** (goal), [**so that** (benefit)]"*. Their use is growing, especially in the context of agile software development [42].

Usually, user stories are created for each actor, describing the actions each one can perform in the system.

Tables 3.1, 3.2, 3.3, 3.4, and 3.5 hold the user stories for the different actors described in the previous section.

Table 3.1: *Unauthenticated User's* User Stories

Id	Name	Priority	Description
US00	Sign In	High	As an *Unauthenticated User*, I want to sign in so I can access my account

Table 3.2: *Authenticated User's* User Stories

Id	Name	Priority	Description
US01	Sign Out	High	As an *Authenticated User* I want to log out so I can keep my account secure

[2] https://www.omg.org/spec/UML/2.5.1/PDF

Table 3.3: *Participant's* User Stories

Id	Name	Priority	Description
US02	Main Page	High	As a *Participant*, I want to access the main page so I can have access to application functionalities
US03	Surveys	High	As a *Participant*, I want to be able to answer surveys so I can contribute with my data to the study
US04	Submit Survey	High	As a *Participant*, I want to be able to submit the answers so they can be saved
US05	Notifications	High	As a *Participant*, I want to be able to receive notifications so I can be remembered to finish the filling out of a survey
US06	Settings	High	As a *Participant*, I want to access the settings page so I can adapt it to my personal preferences
US07	Tutorial	High	As a *Participant*, I want to be presented with a walk-through so I can use the application easily
US08	Avatar	Medium	As a *Participant*, I want to have an avatar so I can personalize the application
US9	Buy Items	Medium	As a *Participant*, I want to be able to buy items in the shop so I can have more customization options
US10	Quiz	Medium	As a *Participant*, I want to answer quick quizzes so I can earn in-app currency
US11	Unlock Achievements	Medium	As a *Participant*, I want to be able to unlock achievements so I can earn in-app currency

Table 3.4: *Caregiver's* User Stories

Id	Name	Priority	Description
US12	Main Page	High	As a *Caregiver*, I want to access the main page so I can have access to application functionalities

Continued on next page...

Table 3.4 – continued from previous page

Id	Name	Priority	Description
US13	Surveys	High	As a *Caregiver*, I want to be able to answer surveys so I can contribute with my data to the study
US14	Submit Survey	High	As a *Caregiver* I want to be able to submit the answers so they can be saved
US15	Notifications	High	As a *Caregiver*, I want to be able to receive notifications so I can be remembered to finish the filling-out of a survey
US16	Settings	High	As a *Caregiver*, I want to access the settings page so I can adapt it to my personal preferences
US17	Adolescents Progress	Medium	As a *Caregiver*, I want to be able to see my adolescents progress so I can press them to finish the filling out
US18	Tutorial	Medium	As a *Caregiver*, I want to be presented with a walk-through so I can use the application easily

Table 3.5: *Researcher's* User Stories

Id	Name	Priority	Description
US19	Researcher Panel	High	As a *Researcher*, I want to access an admin panel so that I can use admin tools on accounts and data
US20	Create User Account	High	As a *Researcher*, I want to be able to create user accounts with different roles so I can add subjects to the study
US21	Change Password	High	As a *Researcher*, I want to be able to change a user's password so I can allow this user to start using the application
US22	Add Gamification	High	As a *Researcher*, I want to be able to choose if the created account has gamification elements so I meet the ratio requirements

Continued on next page...

Table 3.5 – continued from previous page

Id	Name	Priority	Description
US23	Remove Account	High	As a *Researcher*, I want to be able to mark accounts as deleted so I can satisfy a subject's specific account delete request
US24	Delete Account Data	High	As a *Researcher*, I want to be able to delete account associated data so I can satisfy a subject's specific delete data request
US25	Peek Progress	Medium	As a *Researcher*, I want to be able to access a dashboard so I can find out the data collection progress
US26	Check Gamification Ratio	High	As a *Researcher*, I want to be able to check the ratio of gamification/non-gamification accounts so I can maintain it under the specified requirements
US27	Export Data	Low	As a *Researcher*, I want to be able to export the gathered data so that I can further work with it
US28	Check Downloads	Low	As a *Researcher*, I want to be able to call app Stores API's so I can check the current number of installed apps
US29	Access Gamification Results	High	As a *Researcher*, I want to be able to access a gamification dashboard so I can measure the gamification impact
US30	Get Results Data	High	As a *Researcher*, I want to fetch study data so I can interpret the results

3.3.3 Domain Model

The solution's complete domain model can be found in Appendix A. It can be divided into three sub-domains: the user, the survey and the gamification sub-domains. The user sub-domain, shown in Figure 3.3 comprises the *User* super-class and its sub-classes, *Participant*, *Caregiver* and *Researcher* and the relation between *Caregiver* and *Participant*, modelled as the *Kinship* class. *Participant*, *Caregiver* and *Researcher* inherit all the *User* class attributes.

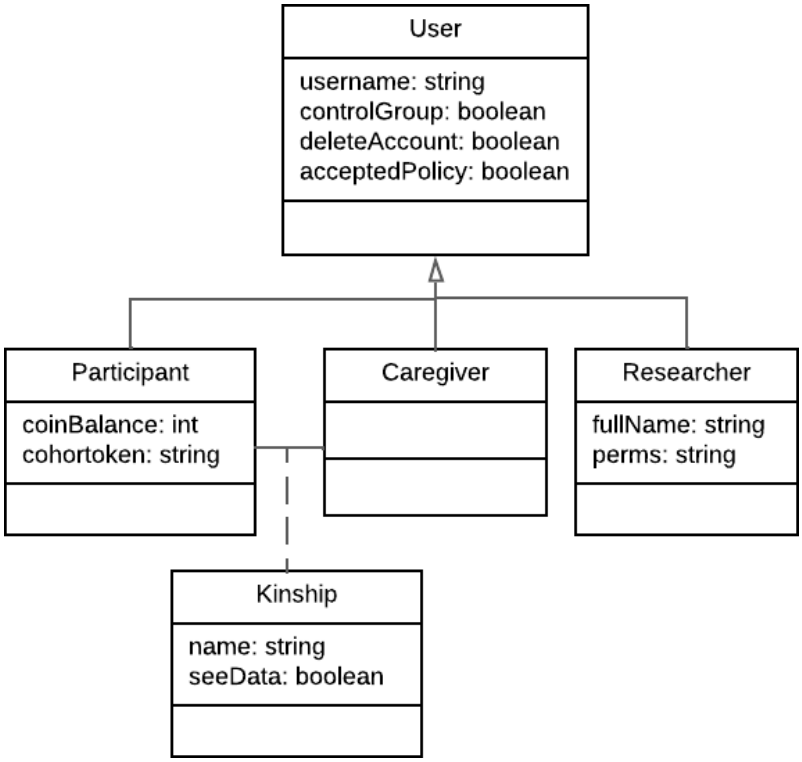


Figure 3.3: User Sub-domain

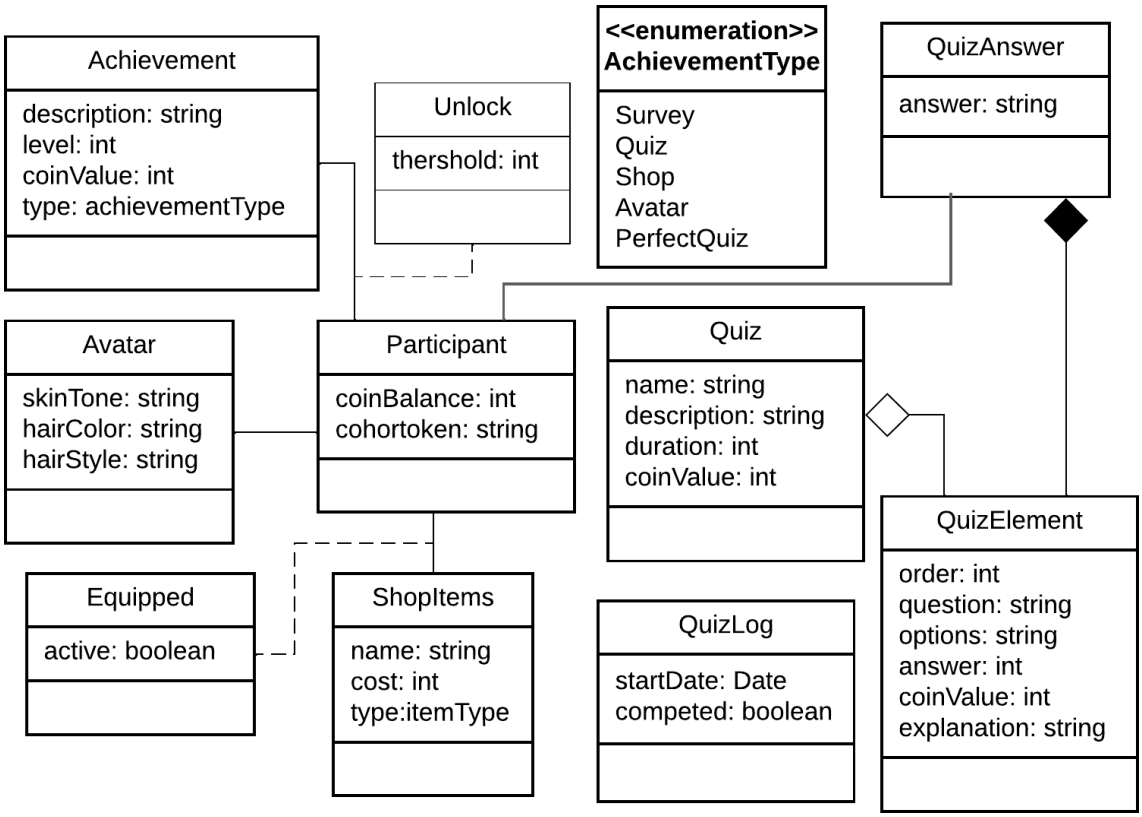


Figure 3.4: Gamification Sub-domain

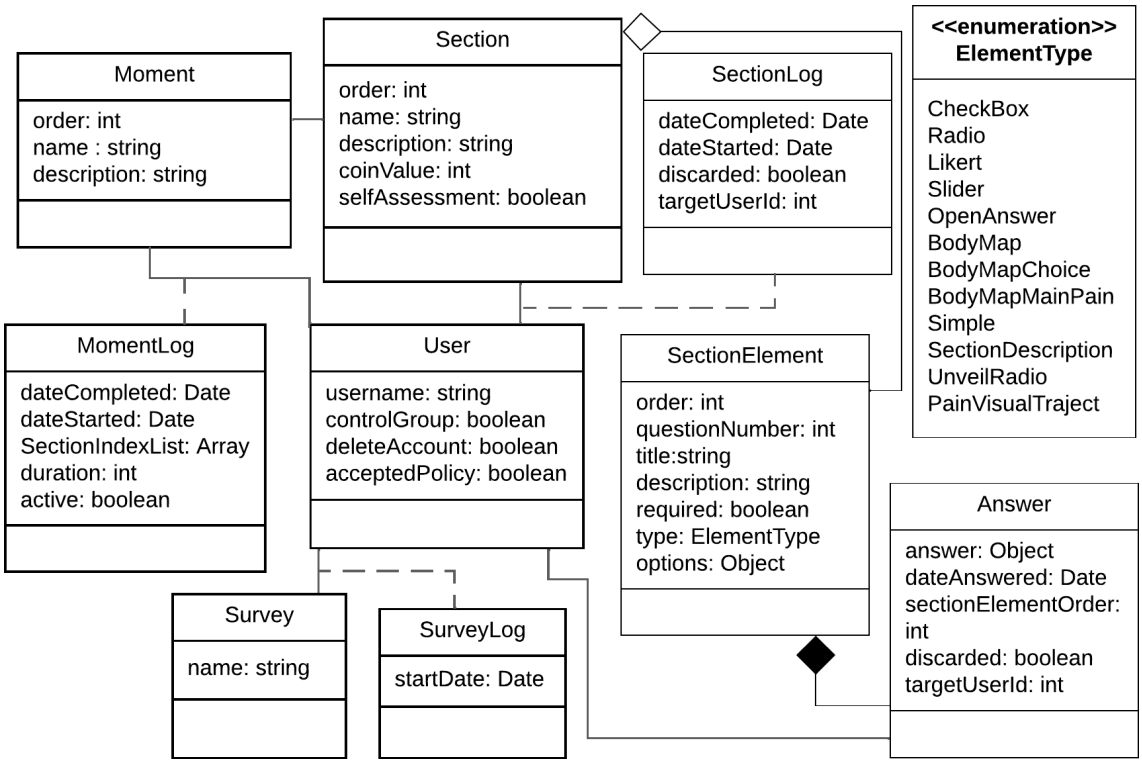


Figure 3.5: Survey Sub-domain

The gamification sub-domain depicted in Figure 3.4 incorporates all the classes related to the gamification elements. The *Achievements*, *Avatar*, *Quiz* and the *ShopItems* classes represent the different gamification elements of the mobile application. Each *Participant* achievement is modelled as the *Unlock* class, and each *ShopItem* present in the *Avatar* is modelled as the *Equipped* class. There is also an enumeration reference to the possible five achievement types that represent the available badges. The *Quiz* class is composed of *QuizElements*, which represent the *Quiz* questions, and each *QuizElement* will have the answers provided by the *User*. Each *User Quiz*'s progress is recorded in the *QuizLog* class.

The last sub-domain, shown in Figure 3.5, includes the survey classes, which models the surveys that form the study core. Each *Survey* is modelled as a *Moment*, which has *Sections* representing a moment page, and each *Section* comprises several *SectionElement* instances, each representing a single question. Each *SectionElement* will be associated to all the *Answer*s given by the users to that particular question. The progress in the *Section* is modelled in the *SectionLog* class. The *SectionElement* class represents the *Moment* questions and can have several pre-defined types enumerated in the *ElementType* enumeration.

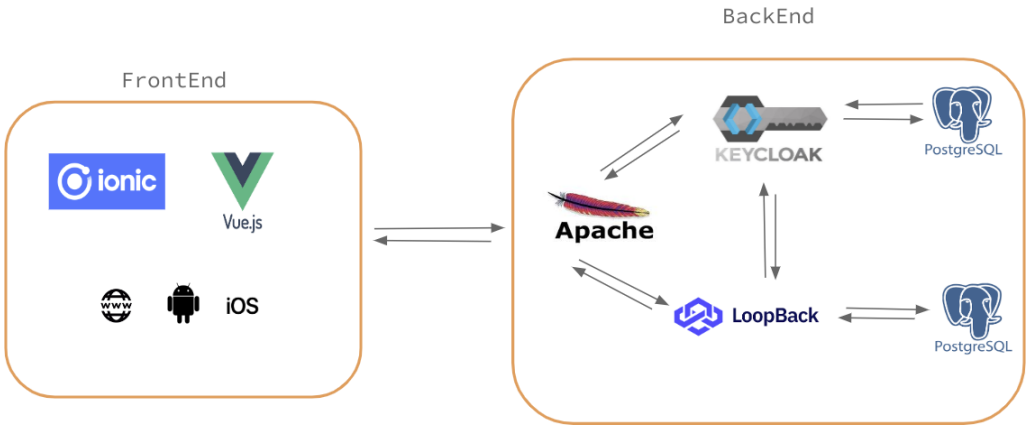


Figure 3.6: Technology Stack

3.3.4 Technologies

The solution was conceived according to the client-server model[3] and the model-view-controller (MVC) architectural pattern[4]. In the client-server model, a central server processes the client requests and persists the data when required. At the same time, the MVC architecture is commonly used for developing user interfaces dividing the related program logic into three interconnected elements. This is done to isolate internal representations of information, its presentation to the user and its acceptance from the user.

The client is developed with the *Ionic*[5] and *Vue.js*[6] cross-platform frameworks. This choice allows development simultaneously for the three platforms — *web*, *Android* and *iOS*. Alongside this major advantage, it also brings hot-reload, easy-testing and access to device-native functionalities [15]. Figue 3.6 illustrates the technology stack described.

The backend is implemented according to the concept of micro-services[7]. Each entity runs independently in its Docker[8] container. The backend server consists of an *Apache*[9] web server which is the only point of entry into the backend server. It acts as a reverse proxy and relays all incoming communication to the appropriate Docker container internally. That creates the only connection to the internet and distributes the different requests over the internal network.

[3] https://en.wikipedia.org/wiki/Client%E2%80%93server_model
[4] https://en.wikipedia.org/wiki/Model%E2%80%93view%E2%80%93controller
[5] https://ionicframework.com/
[6] https://vuejs.org/
[7] https://en.wikipedia.org/wiki/Microservices
[8] https://www.docker.com/
[9] https://httpd.apache.org/

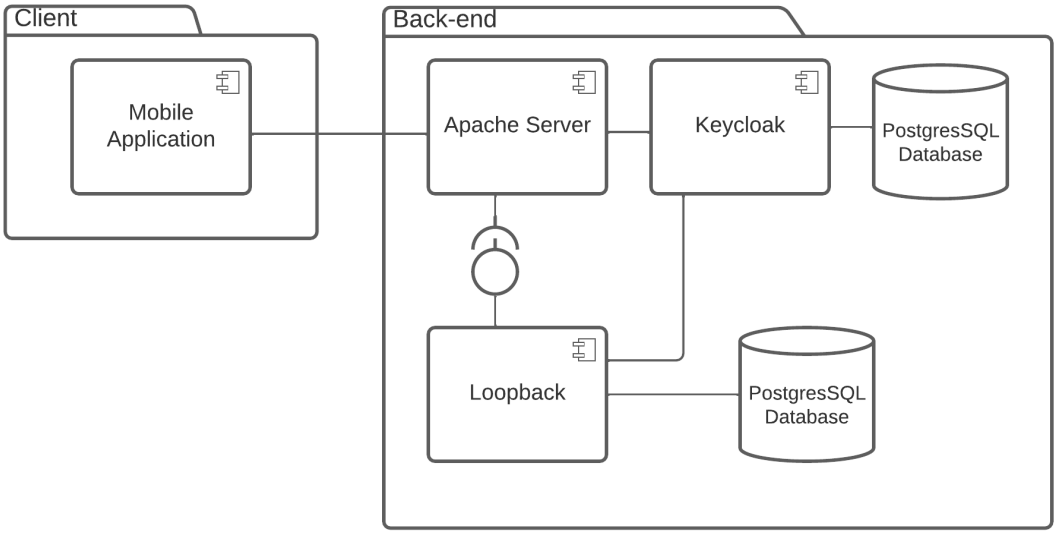


Figure 3.7: System's UML Diagram

This internal network has a service implemented with the Node.js Loopback[10] framework that provides an API, connected to a PostgreSQL[11] database. The authentication and request permissions are handled by a Keycloack[12] service which uses its own PostgreSQL database.

The Figure 3.7 shows a more formal system representation, using UML.

3.4 Gamification Elements

As described in Section 2.2.2, some authors apply just a few gamification elements to isolate their effects and measure them more precisely.

A set of gamification elements were selected following part of Xu's [63] research. The gamification elements chosen were an editable avatar, a badge system, a progress bar, in-app coins and items to buy with the coins. Jent and Jannet [33] concluded that using a progress bar could help enhance engagement. According to Oliveira and Pula [49], the avatar is a customization game element while badges fall in the reward classification, and both are often used throughout the bibliography. These gamification elements can influence users's behaviour [53].

The user can interact with the gamification elements by changing the avatar's hairstyle and skin colour as well as equipping clothes or accessories. These clothes or accessories

[10] https://loopback.io/doc/en/lb4/index.html
[11] https://www.postgresql.org/
[12] https://www.keycloak.org/

can be bought in the store with in-app coins earned by filling-out surveys and quizzes. A badge system is also available, recognizing the user's effort in filling out and interacting with the avatar and the shop. Link et al. [40] found that the highest reported indices of badges acceptance and usefulness were among users between 18 and 35 years old. At the same time, earning badges is also rewarded with in-app coins. There is also a progress bar showing users their progress filling out the survey, giving visual information about the length and the progress in the survey, enhancing the user motivation, especially in the longer ones [33]. The quizzes are composed of questions created by the ISPUP team, mainly focused on interesting curiosities extracted from previous ISPUP publications about chronic pain in adolescents. Examples of these questions are: "Which of the following is a risk factor for future pain?" or "Youngsters who report more pain have lesser physical pressure tolerance?".

Figure 3.8 shows a diagram of how the gamification elements interact with each other. In this figure, and marked in red with the number 1, is the mobile application main page. In the top bar it is possible to see the user's current in-app coin balance. This balance can be built by filling out surveys or quizzes, which can be accessed through the two cards right under the top bar. The third card, partially uncovered, represents the avatar's current state. Figure 3.9a shows a quiz page, with a correctly answered question. There is information about the quantity of in-app coins earned, and a text elaborates the correct answer at the bottom of the screen.

After earning some in-app coins, the user can navigate to the shop page, marked with the number 2 in Figure 3.8, and buy items from a wide range of accessories, shirts or pants. There are a total of fifty-five items distributed through the three categories. Figure 3.9b shows the shop's shirts portion in detail. Seven items are in the category of accessories, twenty-six in the shirts category, and twenty-two in the pants category. The average price of the items is 53,5 in-app coins. These items can then be used to equip the avatar on the page marked with the number 3 of the same figure. Beyond the customization with bought items, the avatar can also be customized by changing its hairstyle and skin colour. Figure 3.9c shows the hairstyle menu opened up, displaying the available coiffures.

Finally, there is a badge system that monitors all this activity and classifies the user in one of three levels, from beginner to expert, depending on the level of interaction, as detailed in Table 3.6. The badge progression details can be checked on the page marked with the number 4 in Figure 3.8. Every time a user achieves a badge, visual feedback is shown on the lower end of the screen. The badge achievement also grants in-app coins.

Table 3.6: Badge System

Badge Types	**Steps from Beginner to Expert**
Surveys	Fill out 1, 2 or 3 surveys
Quizzes	Complete 1, 2 or 3 quizzes
Shop	Buy 1, 3 or 5 items
Avatar	Equip 1, 3, or 5 items
Perfect Quizz	Answer correctly to all questions in 1, 2 or 3 quizzes

3.5 Data Quality and Engagement Measurements

Section 2.3 mentioned user behaviour as one of several factors contributing to the survey data quality. Several data quality measures can be applied to measure the gamification elements' ability to shape user behaviour and enhance data quality. Analysing the surveys created by the ISPUP team, the Likert-type questions stand out as the most recurrent type. Therefore, the answers to this type of question were analysed for their variability. Two complementary measures were adopted: individual response variance, as described by Arthur et. al [4] and individual invariability, as described by Goldammer et al. [18] and by Meade and Craig [46].

The individual response variability, or *Likert variance* rate, is the degree of different answers that the user gave throughout the surveys. On the contrary, individual invariability, or *Longstring* rate, is a measure of straightlining. The more straight answers with the same value the user gave, the highest the *Longstring* value will be. These two measures complement each other because *Likert variance* can detect types of straightlining that *Longstring* can not. For example, repeated patterns, like switching the answer from two adjacent answers (e.g., answering 3,4,3,4,3,4) [4].

Because the surveys had Likert questions with different scales of 3, 5, 6, and 7 answering possibilities, a simple method to standardise the scales was used in the corresponding Likert metrics calculations [30]. Also, according to Wyss [62] the Likert variables can be treated as continuous, as long as the Likert scale is seven or more categories; thus, the applied standardisation allows to treat the Likert variables as continuous. Furthermore, Mangiafico [43] proved that traditional nonparametric tests are good approximations for ordinal regression and can be used to analyse Likert data.

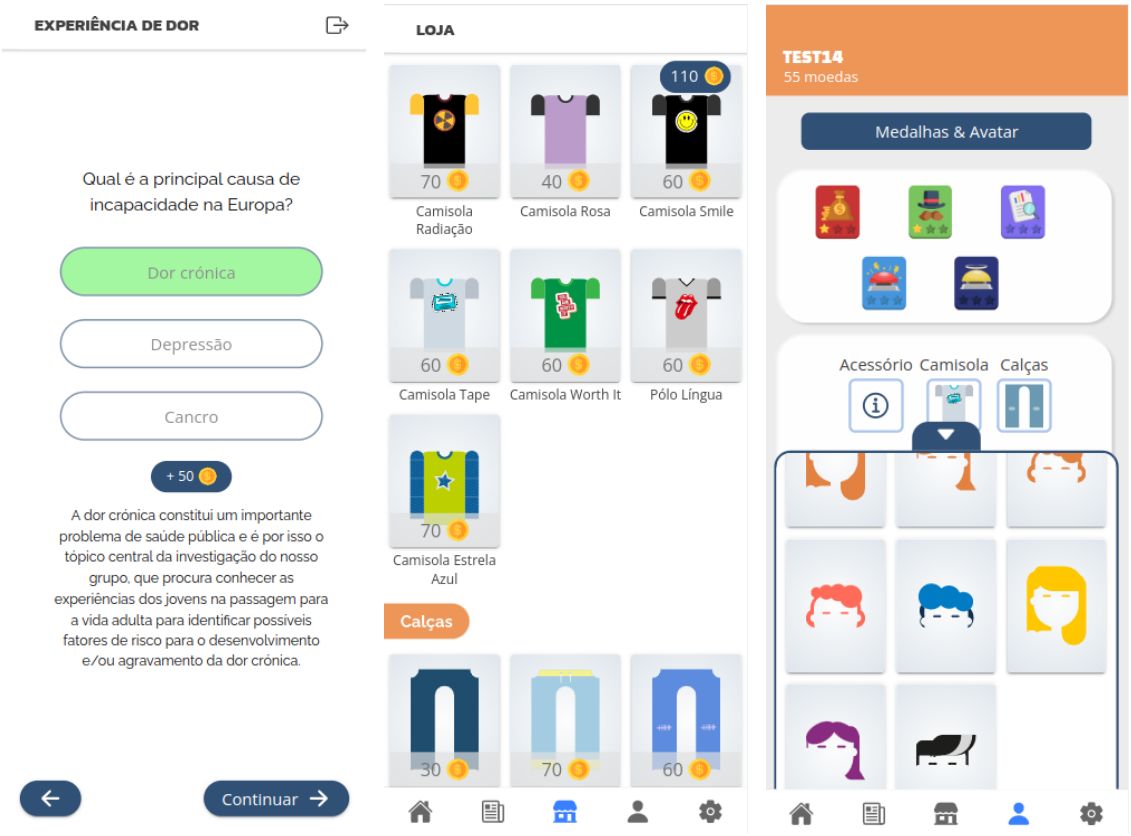
Another way of assessing user engagement is by analysing the length of the answers to open questions. This assessment was made using two metrics: the length of the answer to open questions and the ratio of non-response to these questions. This question type is suitable to measure engagement because, throughout the surveys, it is a non-mandatory type

Figure 3.8: Gamification Elements Integration

of question. According to Mavletova [44, 45] it is expected that the gamification group would have lesser values of non-response items and also lengthy answers to open questions, projecting more carefulness and commitment. Therefore, answers to open questions with no length are considered to the *Non-Response* rate calculations, and all the other answers to open questions are considered to the *Open Answer Length* rate.

Table 3.7 compiles the metrics used in this study to analyse the hypothetical effect of gamification on engagement and subsequent data quality.

Finally, the time to answer each of the four surveys was registered and analysed as another metric of engagement. Comparisons between gamified and non-gamified groups were made at each survey level and at a global level, considering all four surveys in conjunction. Bibliography tends to consider the gamified survey versions to take more time to answer due to a greater subjects' engagement and commitment. Mavletova [45] postulated that the gamified survey versions would produce longer answers to open questions, less straightlining and fewer non-respondent items, leading to longer answer times. Corey [13] measured the time to answer each question in his work, testing the engagement in straightlining. Lastly, Harms et al. [32] evaluated a survey gamification process and concluded that the gamified group spent more time in surveys.



(a) Quiz Page

(b) Shop Page

(c) Avatar and Badges Page

Figure 3.9: Gamification Elements Detailed

Table 3.7: Data Quality Metrics Adopted

Designation	Measurement
Likert variability rate	Intra-user variability score within the Likert type answers
Long String	Intra-user longest straight answer rate
Non-Item Response rate	Rate of empty answers to open questions
Open Answer Length	Length of the non-empty answers to the open questions

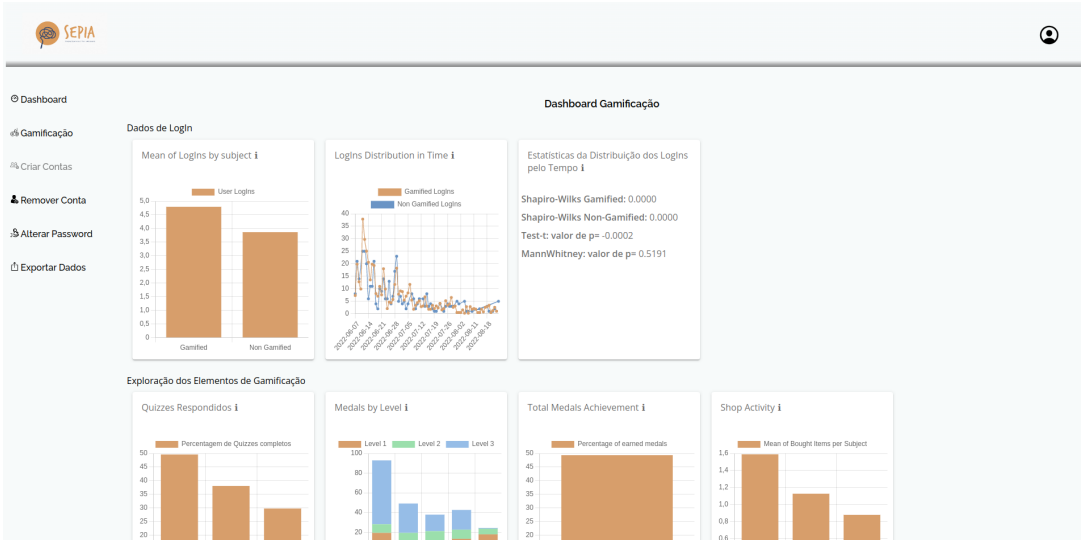
3.6 Methodology

The preparation for the beginning of the study involved some intermediary steps. The mobile application and the researcher portal had implementation cycles until the last release, right before the data collection phase began. During this time, both had parallel development and several test rounds. These were intended to make sure the mobile application release would have the requested usability, and correctness metrics and the whole system's stability and scalability were as desired.

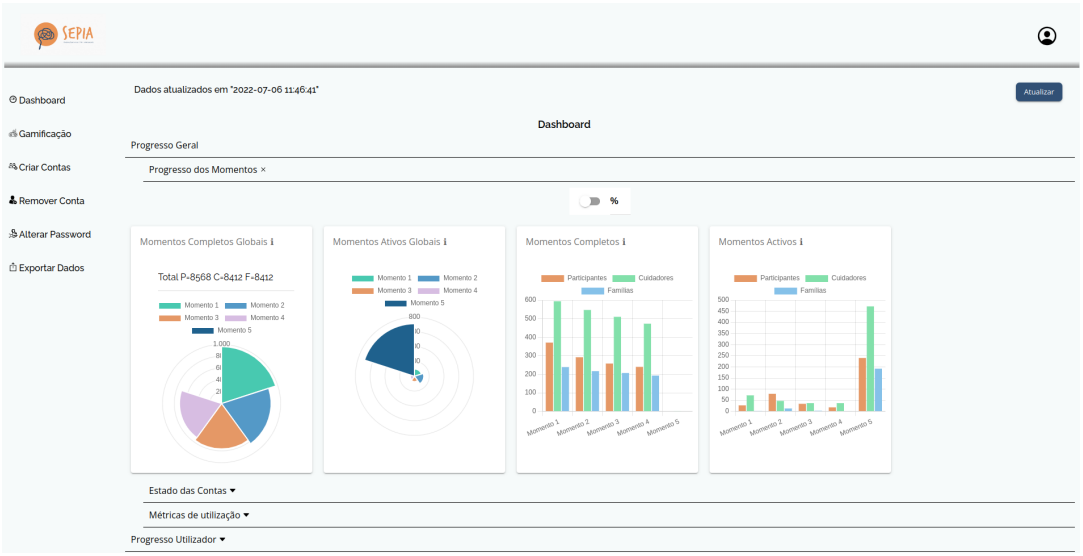
The first testing cycle began applying an internal test round, with test accounts to the mobile application distributed through the INESC TEC and ISPUP project teams. This test round led to the discovery of several survey typos and some bugs, which were corrected until the next step. Also, the implementation of the surveys flow was adjusted at this point as requested by the ISPUP team. The first four surveys can be answered continuously, and the last one can only be answered six months after the first one ends.

The next step included a pilot study and a usability survey. New testing accounts were distributed to a group of twenty young teenagers and their parents. This sample was composed of outside acquaintances of both the ISPUP and INESC TEC teams and did not belong to the final sample. In the end, a usability survey was also distributed to assess the usability, ease of use and the overall opinion about the mobile application among the pilot study users. The summary of this usability survey is available in Appendix B on page 61. The main conclusions drawn were that the mobile application is easy to learn, use and understand and that the gamification elements added some interest and were also easy to understand and use.

At the same time, the researcher portal was also being implemented, with the requirements presented by the ISPUP team. The researcher portal was designed with usability in mind. The fixed sidebar layout was the best solution, presenting navigability to every page on the left side of the screen. The portal dashboard can be found in Figure 3.10b. The researcher can create accounts, delete accounts, change accounts' password, export



(a) Researcher Portal's Gamification page



(b) Researcher Portal's Dashboard

Figure 3.10: Researcher Portal

data and access the dashboard to assess the data collection progress through this portal.

After discussing with the ISPUP team and getting their approval and interest, a new gamification page was added to the researcher portal, shown in Figure 3.10a. This gamification page initially aimed to request pertinent data from the server and compute all the relevant outcomes, showing them in a dashboard format. These outcomes were intended to be included in the results of this work. However, as there was a poor choice of advanced Javascript statistic libraries, precisely a library that implemented the log-rank test, it was

decided to perform all the statistical analyses using the R environment[13] and Rstudio[14]. So, the charts were calculated directly within the researchers' portal gamification page, and the statistical metrics were calculated with the Rstudio tool.

ISPUP team developed an algorithm that divided the process of informing and inviting elements among the "Geração XXI" cohort into different groups. Despite that, the first wave of participation was anticipated to be of several hundred subjects accessing the mobile application simultaneously, causing an unpredictable load on the server. To adress this issue, stress tests were conducted on the server, simulating access of a maximum of two hundred simultaneous requests, resorting to the Apache JMeter[15] testing application. This load test revealed a bottleneck on the Keycloak instance and several seconds of response delay in the latest processed requests. Based on these conclusions, the total server capacity was upgraded from 2 to 12 processing cores and from 2 to 16 Gibibytes (GiB) of random access memory (RAM) momentarily, for the first two months of the data collection cycle. Also, the Keycloak instance was doubled, and a load balancer between the two instances was implemented on the Apache server.

At this point, everything was ready, and the data collection started on 07/06/2022. Although ISPUP plans to extend the data collection phase for roughly one year, this study results analysis focuses on the data collected from day one to 24/08/2022. This limitation to approximately two months of collected data was due to timeline constraints.

3.7 Summary

This chapter presented details about the implemented solution. User requirements comprise the actor's description and the required user stories, the description of the domain model and the chosen technologies stack. The implemented gamification elements and the chosen data quality measurements were also presented. Finally, several implementation iterations were described alongside details about some decision processes.

The next chapter will detail and describe the obtained results.

[13] https://www.r-project.org/
[14] https://www.rstudio.com/
[15] https://jmeter.apache.org/

Chapter 4

Results

The last chapter presented the research problem and detailed the proposed solution, describing the mobile application and the supporting infrastructure. It also described what metrics would be extracted from data and their meaning. This chapter characterises the sample and presents the results achieved using those metrics. It starts by describing the subjects' characteristics and how they were gathered. Then presents the results for the login quantities for each survey and total. The answering time patterns results are composed of each survey's answering time and survival rates. The gamification exploration results traverses the degree of completed quizzes, the shop activity and the earned badges. Finally, the data quality results are presented, composed of *Likert variance*, *LongString*, *Non-response* and *Open Answer Length* rates.

4.1 Introduction

The obtained results are drawn from a subset of the whole data ISPUP will collect. The data collection phase is ongoing and has a planned duration of approximately one year. Therefore, the extracted subset represents the collected data from the data collection phase beginning – 07/06/2022 until 24/08/2022. This date was chosen, balancing the latest possible date and the necessary time to process and analyse the results inside the time constraint window. Nevertheless, these almost two months allow the subset to hold enough data to reveal the trends and the degree of causality. Furthermore, the quantitative analysis was done with a significance level of 5 %. The results extracted from this subset allows to describe the subjects's behaviour under three dimensions: the usability metrics, the degree of interaction with the gamification elements and the measures of the data quality. These three dimensions are different representations of the subject engagement and the degree of the gamification influence.

4.2 Sample Characterization

The SEPIA study uses the "Geração XXI" cohort sample. ISPUP has used this cohort sample in several health studies throughout the years.

The ISPUP's "Geração XXI" cohort started in 2005 with 8495 recruited families. In this cohort, a family comprises a caregiver and, at least, one newborn. This implies that these newborns are, at present, adolescents, having between 16 and 18 years old. In June 2022, 8146 were still electable for the SEPIA study because they had the contact data up to date, and none of the family members had been institutionalized. These 8146 electable families were randomized with an 80/20 ratio, with 80 % of the families being placed in the gamified group and the remaining 20 % in the non-gamified group. Only the adolescents were exposed to the gamified version of the mobile application.

The ISPUP team took care of the SEPIA study presentation and invitation to all the potential subjects. Of the total electable 8146 families, 5757 had a valid email address and were contacted via email, while the rest were contacted through a phone message. This first contact was to present the upcoming SEPIA study and to invite them to read and fill out a "Lime survey" questionnaire containing a more in-depth study description, an informed consent and a link to download the mobile application. On 24/08/2002, the closing date of of this dissertation data collecting phase, 1204 families had completed the Lime survey questionnaire. Belonging to these 1024 families, only 580 adolescents downloaded the mobile application, made the first login and accepted the in-application informed consent. Of these 580 adolescents, 461 belonged to the gamified group and the remaining 119 to the non-gamified group. The results were retrieved from the data produced by these 580 adolescents.

Since the data collection phase began, ISPUP has been persisting in contact with the families that do not give any feedback or that have stopped in some of the intermediate steps by phone calls, email newsletters, and phone message reminders.

By design, and due to privacy regulations, ISPUP does not share each subject identifying data with INESC TEC, so it is impossible to link it with the collected data, precluding any other sample characteristics description.

4.3 Log In Quantity

The number of logins of each subject was recorded throughout the study. In the scope of this study, login represents not only the actual login, introducing credentials on the login page but also the main page loading. This is due to the expiration date defined in the subject's assigned token. To prevent the tedious credential presentation process

Table 4.1: Logins Quantitative Analysis

Metric Designation	**Shapiro-Wilk Gamified** α **= 0,05**	**Shapiro-Wilk Non-gamified** α **= 0,05**	**Wilcoxon Rank Sum** α **= 0,05**
User Logins	$p < 0.001$	$p < 0.001$	$p = 0.0017$

every time the subject uses the mobile application, the JavaScript Object Notation (JSON) Web Token used to identify the users in their requests was set with a highly extended expiration date. Therefore, after the initial credentials presentation process on the login page, every time the user loads the mobile application or awakes it from the idle state is presented with a newly loaded main page, bypassing the login page. Thus, the login events represent the number of times the user utilised the mobile app. Figure 4.1 shows the average number of logins registered in each study group — gamified and non-gamified. Concomitantly, Figure 4.2 shows the adjusted number of logins on each day of the elected data collection phase for the two groups. These two figures suggest a higher number of logins in the gamified group than in the non-gamified with the gamified one having a mean login number of 4,8 and the non-gamified one 3,8.

The trend suggested in Figure 4.1 and in Figure 4.2 is supported by the quantitative analysis. After assessing the two, gamified and non-gamified logins distributions' non-normality, the Wilcoxon rank sum test showed that these two distributions are statistically different within the used confidence level. Table 4.1 presents the statistical details. Thereby, there is enough evidence to state that the number of logins made by the subjects in the gamification group was higher than those made by the non-gamified group subjects. This indicates a greater engagement toward the mobile application, leading users to return more often. As gamified systems provide more user enjoyment than traditional systems do [58, 59], it is safe to state that the differences in login behaviour were due to the gamification elements.

4.4 Answering Time Patterns

The answering pattern was analysed through two perspectives: the total time each user spent answering the surveys and survival rates for each survey. Figure 4.3 shows the average time in minutes spent by each group to complete the different surveys. It appears that, except for the fourth survey, users in the gamified group spent more time filling out the surveys. Although this trend is present, the quantitative analysis shows that there is

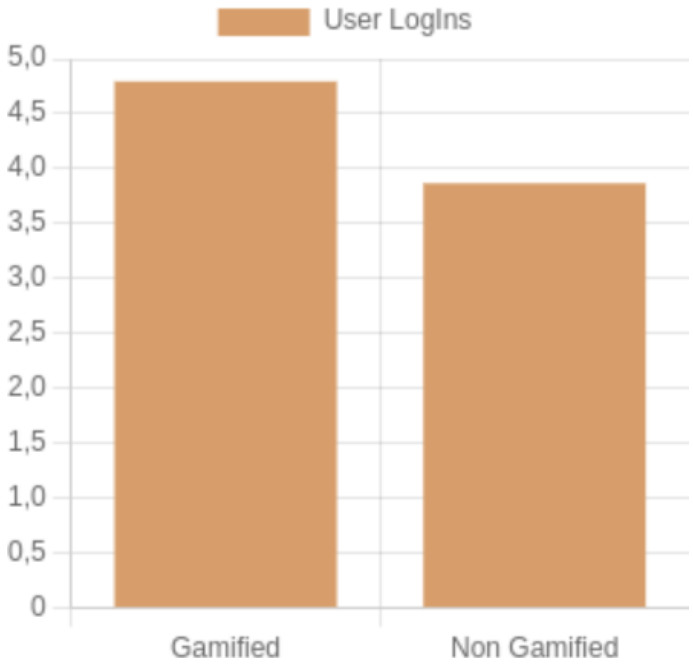


Figure 4.1: Mean User Logins by Study Group

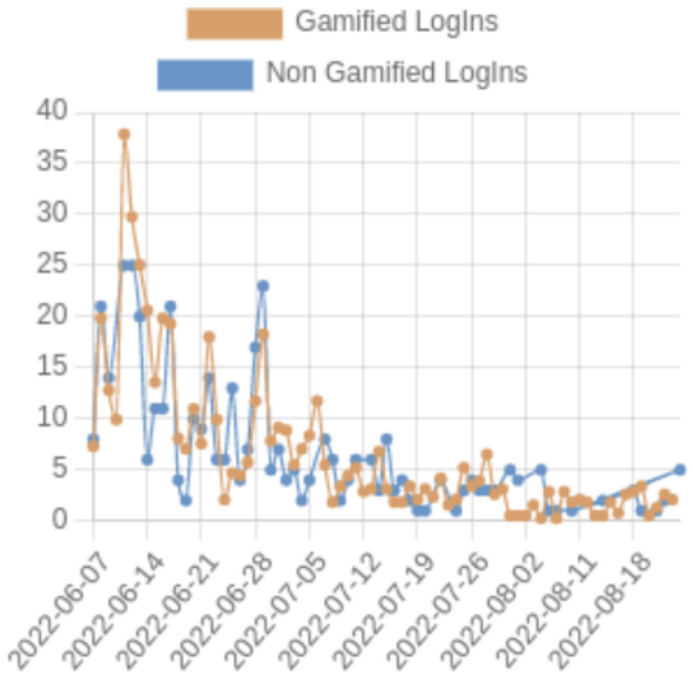


Figure 4.2: Adjusted Logins Distribution in Time by Study Groups

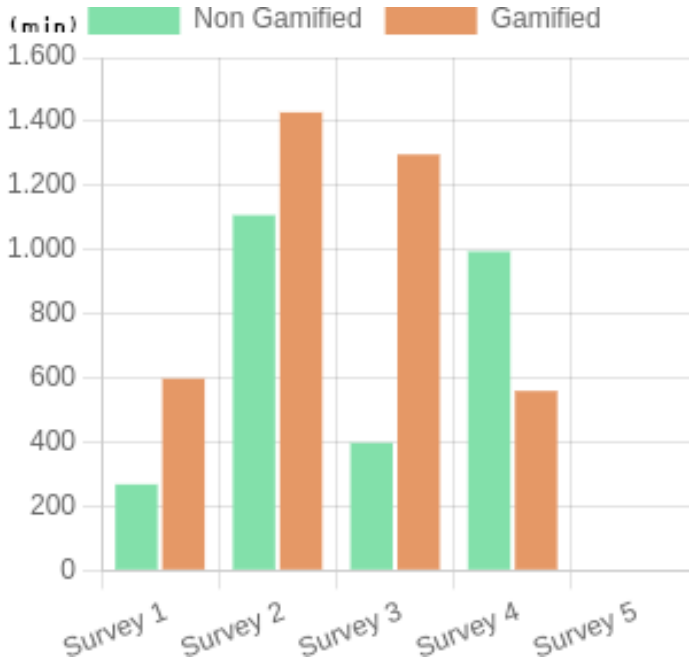


Figure 4.3: Surveys Answering Times by Study Group

no statistically significant difference between the survival rates of the gamified and non-gamified groups in every survey. The Log-Rank test was used to compare the survival rate distribution of the gamified group with the survival rate distribution of the non-gamified group in each survey. No statistically significant evidence exists to state that they differ for every survey. Table 4.2 displays this analysis detailed results.

Despite the previous establishment of the same filling time between the two groups, and without undermining it, the analysis of the survival rates plots in Figures 4.4, 4.5, 4.6 and 4.7 reveal two main trends in the second, the third and the fourth surveys. These figures hold the plots for the first four surveys. The vertical axis represents the probability of "surviving", and the horizontal axis represents the time in hours. At $t=0$, all plots start with 100 % because none of the subjects filled out the survey yet. As time passes, the probability decays as subjects achieve the event of interest. For all plots, $t=0$ represents the moment the user finished the tutorial and $t=730$ represents 730 hours after, exactly one month later. All plots show a seemingly vertical line in the first hours because many subjects filled out the surveys at the beginning of the study. The gamified group is represented by orange colour and the non-gamified one with the green colour. The first trend is that the gamified group takes more time to complete the surveys. The second is that the non-gamified group completes the surveys in greater quantity. These trends are denoted by the orange line representing the gamified group that keeps above the green one (more time to complete the survey) and never crosses the green one, becoming lower (fewer subjects completing the survey).

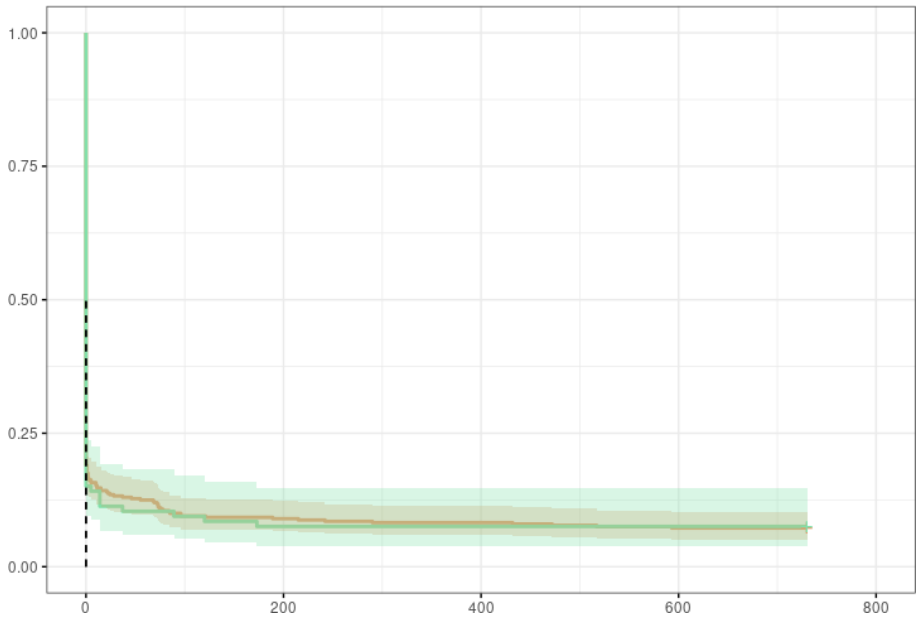


Figure 4.4: Survey 1 Survival Rate

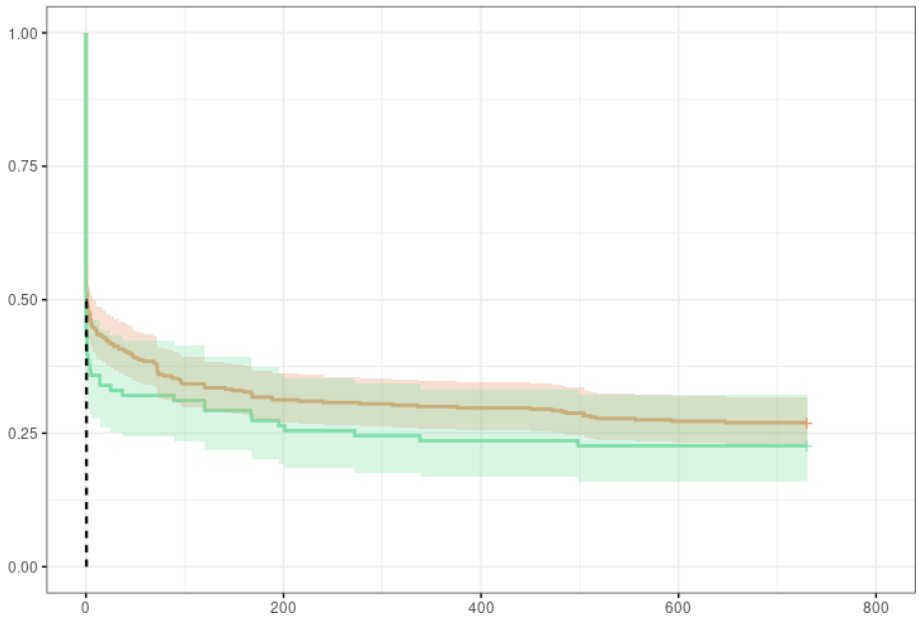


Figure 4.5: Survey 2 Survival Rate

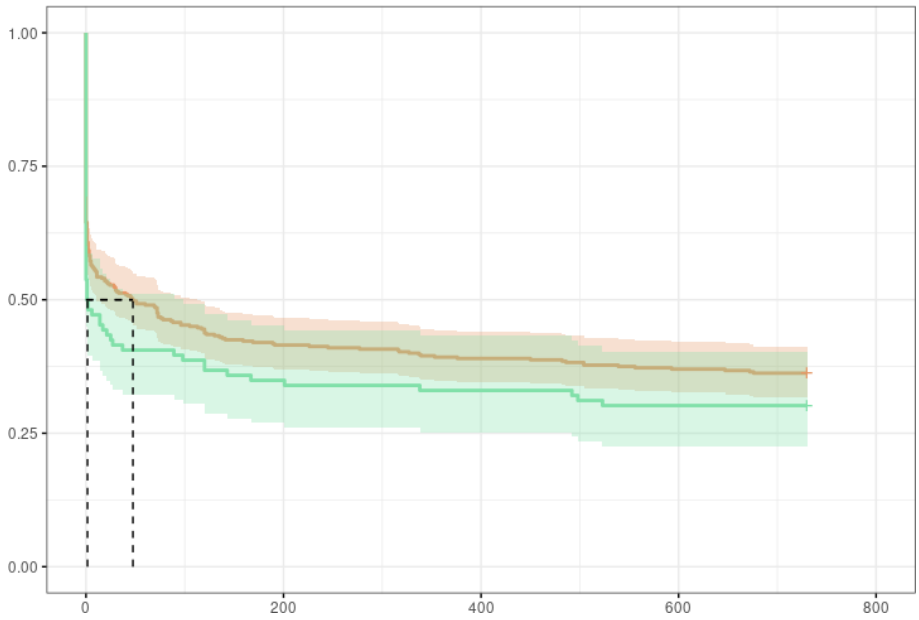


Figure 4.6: Survey 3 Survival Rate

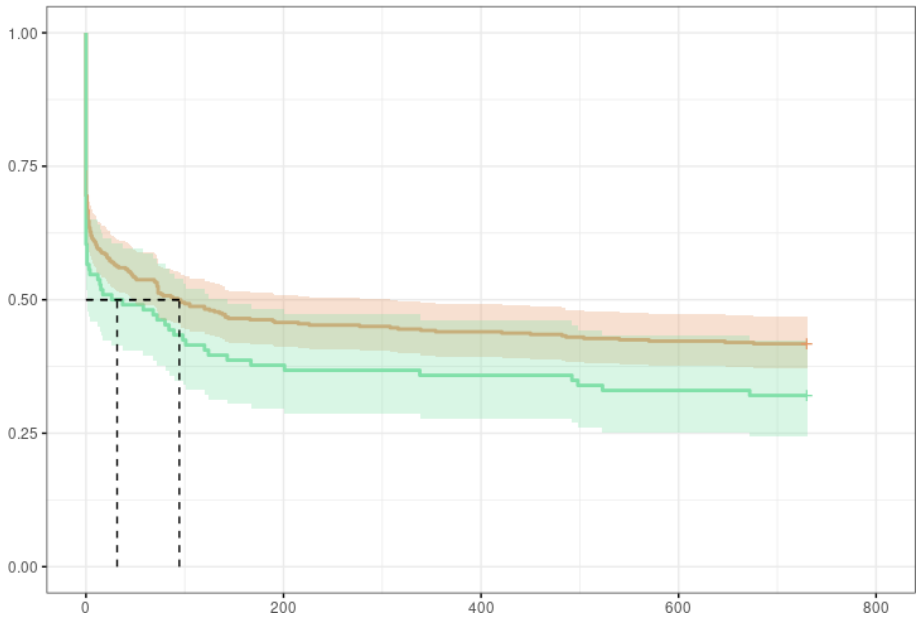


Figure 4.7: Survey 4 Survival Rate

Table 4.2: Survival Rates Analysis

Survey	**Log-Rank Test** α **= 0,05**
1	p = 0.86
2	p = 0.25
3	p = 0.20
4	p = 0.07

4.5 Gamification Exploration

Three behaviour dimensions were analysed to assess the subjects' interaction degree with the gamification elements. The percentage of completed quizzes, the extent of buys in the shop and the distribution of the achieved medals by their type.

The number of completed quizzes was analysed to determine subjects' interest in this particular gamification element. Figure 4.8 shows the percentage of quizzes completed at the time of the completion of the fourth survey. This particular point in the study timeline was chosen because it is the last subject's iteration until now, except for the fifth survey which is blocked for six months after the completion of the first survey, and no subject has reached it yet. This particular point provides a reasonable time for the subjects to explore the gamification elements. The first quiz has a completeness value of 50 %, the second quiz has a value of almost 40 % and the third of 30 %. There is a descendent pattern of consecutive lesser completed percentages from the first until the third quiz because they can only be accessed in order. These percentages represent a reasonably middling engagement measure for this particular gamification element.

The number of bought items represents the extent of interaction with the shop page. Figure 4.9 on page 42 portrays the average number of items bought by subjects in each of the three item categories at the time of the completion of the fourth survey. At this point, subjects have filled out the first four surveys and are waiting for the last survey to become available. The average number of purchased items is 1.6 for shirts, being 1.2 for pants and 0.9 for accessories. These values are below the expected, showing that, on average, each subject only bought between one and two items. More interestingly, Table 4.3 details that, on average, subjects still have a coin balance of 593 in-app coins at this moment. This amount is enough to buy approximately 90 % of the full range of item options.

Finally, the earned badges throughout the study can reflect the subject's degree of acceptance with the gamification as a whole. Figure 4.10 displays the percentage of earned medals in for each type. The levels are stacked because levelling up implies losing the database's current level and swapping it to the next one. For example, the "Complete the

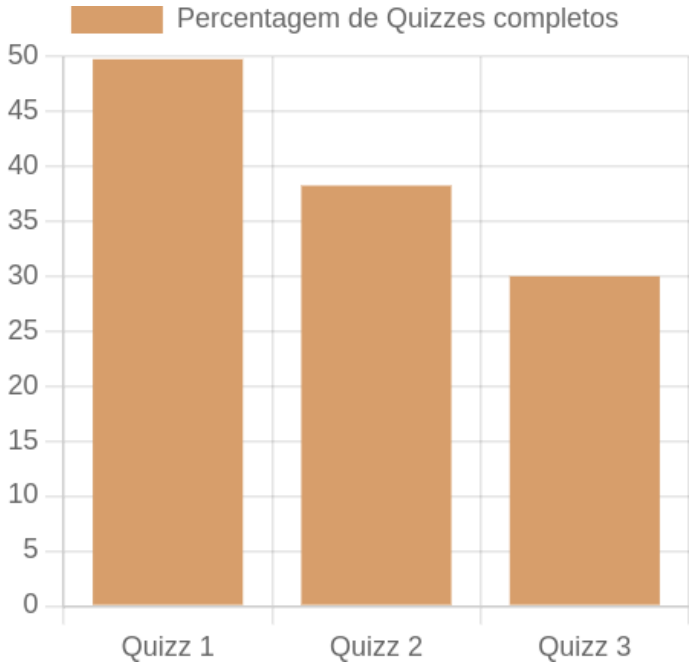


Figure 4.8: Percentage of Completed Quizzes

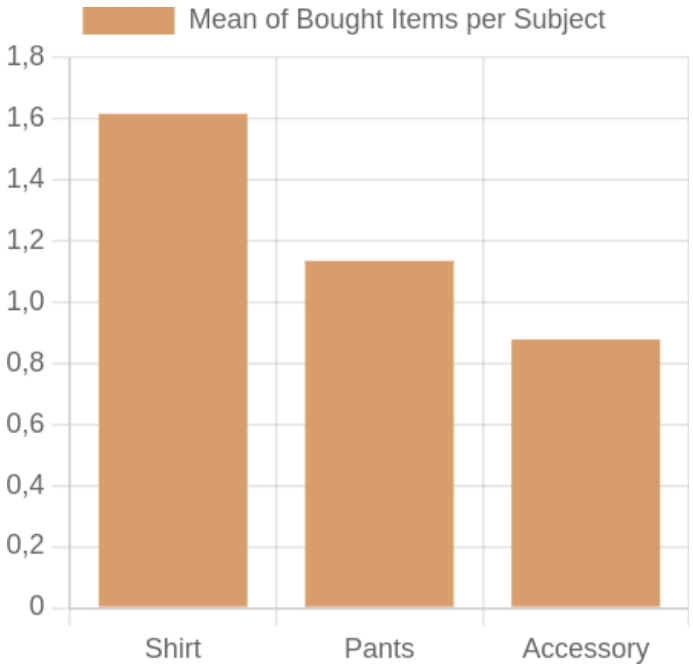


Figure 4.9: Average of Bought Items by Subject

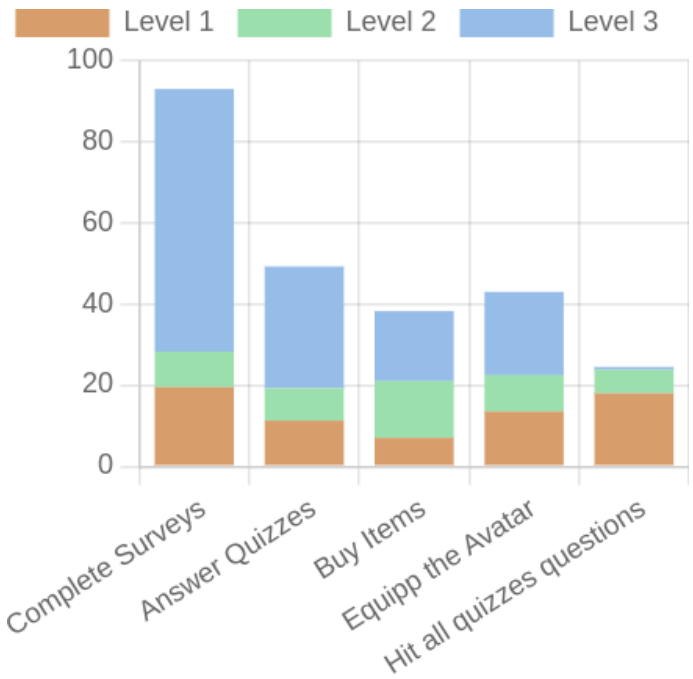


Figure 4.10: Medals Achieved by Type

surveys" bar shows 20 % of the subjects in level 1, in orange. This means that 20 % of the subjects only filled out the first survey. In green, roughly 9 % of the subjects are in level 2. This means that 9 % of the subjects already filled out the first and the second survey, currently holding the level 2 badge. Lastly, in blue, approximately 64 % of the users filled out the third or the fourth survey and therefore are in level 3, as detailed in Table 3.6. Thus, the second, third and fourth bars, with the corresponding titles of "Answer Quizzes", "Buy Items", and "Equipp the Avatar", represent the subjects' interaction with the badge system. Most subjects achieved level 3 on these three badges, indicating a fairly good engagement rate. On the contrary, roughly over 50 % of these three available badges were not earned, meaning that many subjects did not interact with these gamification elements. This value is higher than expected. The last bar, "Hit all quizzes questions", lacks representation in the level 3 badge. This is expected, as it is improbable to correctly answer all quizzes' questions without a profound knowledge of the ISPUP's recent year's publications or massive luck.

4.6 Data Quality Metrics

The data quality metrics results are summarized in Table 4.4. Contrary to expected, none of the studied metrics suggested differences between the gamified and the non-gamified groups.

Table 4.3: Shop Activity Detail

Description	**Mean Value by Subject**
Coin Balance	593
Number of Purchased Items	1.2
Coins Spent on Items	63.57

The Shapiro-Wilk test produced enough evidence to consider the non-gamified group Likert variance to have a normal distribution, but not the gamified group. Therefore, as one of the distributions is non-normal, a non-parametric test had to be applied, and the Wilcoxon Rank Sum test was chosen to compare the equality of means between two independent samples. The result allows concluding that there are no statistically significant differences between the Likert variance distribution in gamified and non-gamified groups.

The same process was applied to the "Open Answer Length" metric. This metric revealed non-parametric distributions within the two groups. Furthermore, the distributions between the two groups were considered not statistically significantly different.

The "Non-Response" metric was not tested for normality because its data had a binary format. Thus, a contingency table was created with the two groups' data and applied a Chi-square test with Yates' correction. The result shows that there is no evidence statistically significant to assume that the proportions differ between the two groups.

Finally, the "LongString" metric was also tested for normality and showed a non-normal distribution for both groups. Although the Wilcoxon Rank Sum test presents evidence that the distributions between the two groups are not statistically significantly different, the "p-value" calculated is close to the pre-defined significance level. Therefore, even though there is no statistically significant evidence that the straightlining behaviour is different in the two groups, there is a trend. Moreover, this trend has the exact tenor as expected, as the gamified group mean is lower than the non-gamified group mean (**m=1.82**$\pm$1.92 vs **m=2.13**$\pm$1.66) which represents less straightlining in the gamified group.

4.7 Summary

This chapter described the achieved results throughout the three proposed dimensions. The achieved results for login and answering response time behaviours, gamification elements interaction and answering quality were presented. The majority of these results were not in line with the expectations.

Table 4.4: Data Quality Results

Metric	**Shapiro-Wilk Gamified Group α = 0,05**	**Shapiro-Wilk Non-gamified Group α = 0,05**	**Test Used**	**Test Result α = 0,05**
Likert Variance	$p < 0.001$	$p = 0.712$	Wilcoxon Rank Sum	$p = 0.544$
LongString	$p < 0.001$	$p < 0.001$	Wilcoxon Rank Sum	$p = 0.070$
Non-Response	–	–	Chi-square with Yates' correction	$p = 0.135$
Open Answer Length	$p = < 0.001$	$p < 0.001$	Wilcoxon Rank Sum	$p = 0.203$

The next chapter will relate the different results, contextualize them with the bibliography and indicate possible future work.

Chapter 5

Discussion

The last chapter described the achieved results in its defined dimensions and stated that most of them fell outside the expected patterns.

This chapter will relate the different measurements to each other and with the existing bibliography. It will start to analyse the implications of the higher logins number in the gamified group, the low level of exploration of the gamification elements and the lack of evidence on the effect of the gamification on the data quality measures. Then it will mark out the study limitations and relate them with the findings. Finally, it presents possible directions for future works based on the present findings and limitations.

5.1 Results Analysis

Results presented in section 4.3 showed a statistically significant difference in the number of logins between the gamified and the non-gamified group. This means it is safe to conclude that the gamified group had an increased level of logins during the analysed timespan. Xu [63] refers to this behaviour as re-engagement. As the login value is statistically significantly different between the two groups is possible to conclude that the gamification elements enhanced the re-engagement behaviour of users in the gamified group. In their bibliography meta-analysis, Looyestyn et al. [41] concluded that most of the analysed studies that investigated time spent on software as the engagement outcome reported positive effects. Among the studies considering engagement in terms of the number of occasions visited, the majority reported positive effects. As more logins and more time spent on the mobile application increases the probability of filling out more surveys, these are significant findings.

Unfortunately, the remaining results do not follow these findings and are not in line with what was expected. As discussed in Section 2.3, the gamification elements are expected to produce survey answering with more engagement and less carelessness. This could possibly lead to longer completion times, as stated by Mavletova [45]. However, neither the total time to respond nor the survival rates for each survey showed statistically significant differences between the two groups. These findings could indicate that although the gamification elements influenced the login behaviour, the survey filling out was not affected by them. One possible explanation for this unexpected result is that the gamification elements were designed outside the surveys, with no direct connection besides the completion payment with in-app coins and the correspondent badge. The only element introduced inside the surveys was a progress bar. Keuch and Zhang [35] warned that some gamification changes could alter the meaning of the underlying construct intended to be measured by the instrument and even its validity and the surveys were core instruments of ISPUP's study and should not be changed. Nevertheless, some other gamification elements could be applied, enriching the connection between the gamified narrative and the survey fill out without compromising the survey's integrity. These could be a leaderboard, one of the Wiss's [62] proposed elements, more specifically a leaderboard for each survey, ordered by the percentage of progress among all subjects. Additionally, there could also be some social interaction elements. Hamari [23] concluded that the social elements are essential for creating engaging gamification services.

Although not statistically significant, the survival rate plots in figures 4.4, 4.5, 4.6 and 4.7 presented two trends, one very unexpected and the other naturally expected. The first one is that the non-gamified group tend to have a survey completion rate higher than the gamified one. There is more non-gamified completed surveys than gamified ones. It was expected that the gamified group would have a higher adhesion, mainly because it was already shown that the gamified group had a proven higher number of logins, more in-app time, and exposure to the gamified elements. There is no explanation for this trend, except that the gamification elements being so secluded from the surveys could influence the subjects to interact with them, relegating the survey filling to secondary importance. The second one is that the gamified group took more time to complete the surveys than the non-gamified group. Keush and Zhang [35] concluded that more than 60 % of the sudies analysed in their bibliographic review showed an augmented response time in the gamified survey versions. This was expected [41, 13] because it is expected that the engagement stimulates more responsible fill-out behaviour and more complete answers to open questions. Regardless, the remaining results, namely the gamification elements exploration and the data quality metrics, do not sustain this claim since they do not present statistically significant differences between the two groups. Even so, these two trends

align with the results of Harms et al.al [32]. They also found a lower response rate for the gamified group, and at the same time, more time was spent in the surveys in comparison with the non-gamified group. Harms et al. proposed an explanation for this behaviour, postulating that the gamified survey version "may have caused polarized reactions among participants, turning those away who did not approve of the chosen design". This can be in line with the results for the gamification elements exploration that will be discussed next.

Survival rates show a trend of more time spent completing the surveys. However, at the same time, there are no statistically significant differences in the data quality indicators, and the gamification exploration fell behind expectations. Since the data quality indicators reveal no statistically significant differences between the two groups it is safe to say that the longer time to complete surveys trend shown in the survival rate plots does not derive from the gamification effect. Furthermore, the gamification elements did not produce sufficient engagement to influence different response patterns between the two groups or different non-mandatory answer quantities and lengths. The metrics "Likert Variance", "Non-Response", and "Open Answer Length" lack statistically significant differences and fall in line with the results of Mavletova[45], Harms et al. [25] and Guin et al. [20]. Keush and Zhang [35] found in their bibliographic review that 50 % of the analyzed studies reported longer answers in the gamified conditions compared to the standard surveys. However, all of these studies reframed the questions as part of the gamification process. Therefore, they concluded "it is not clear whether longer responses reflected better data quality". These data quality results are consistent with the results found for the answering time behaviour. Since there are no differences in the answer length, the number of skipped answers and variability between the two groups is also expected to find no differences in the survey completion times. If the engagement in the gamified group drove no extra answering carefulness and commitment, the answering time pattern is the same in the two groups. Lastly, the "LongString" showed no statistically significant difference between the two groups, but the value calculated was very close to the confidence level previously stipulated. This can reveal a trend to different "LongString" rates between the two groups that could be related to the higher completion times trend shown on survival rate plots. Then, different "LongString" rates could imply different careless answering behaviour and higher data quality in the gamified group. Nonetheless, these are just trends, and no solid conclusions can be drawn from them.

Finally, the exploration of the gamification elements fell short of what was expected. The interaction with the available gamified elements was below expected, possibly contributing to the inferior engagement achieved. As presented in Section 4.5, on average, subjects bought approximately only one item of each category and at the time of the fourth

survey completion had a significant balance of 593 in-app coins. This amount of in-app coins is sufficient to buy nearly all the inventory. These findings suggest that the gamification elements did not produce the expected interest levels to create an engagement strong enough to change subjects' behaviour to influence the data quality and answering time pattern. Mainly two factors can contribute to these circumstances.

The first one could be the flow of the surveys. Initially, the ISPUP team had requested that the surveys have a minimum available time of one week. Even if the subject had filled out the survey, the next one would only be available at the end of the pre-defined time. During the implementation process, this strategy was changed to a continuous flow, without time limitations except for the fifth survey, mainly to avoid drop-outs. Link et al. [40] concluded that gamification works better with a long-term panel survey than a one-time survey. This is the case observed in this work. Since subjects had no time constraints, the majority filled out the first four surveys in a row, which can be considered a one-time survey design, for all practical purposes. Maybe, if the surveys had the previously referred time constraint, subjects would have had more time and opportunity to boost the gamification elements exploration and, consequently, engagement. Moreover, previously in this section, evidence was presented to support that users belonging to the gamified group are more likely to return to the mobile application (i.e., re-engagement, as defined by Xu [63]). Based on these two observations, is possible to postulate that the effect of gamification could have been more significant, had the long-term design of the study been chosen. With the long-term design, subjects would be more exposed to the gamification elements, and its effects could be more prevalent on the outcomes.

The second is the relation between the sample's age and the chosen gamification elements. Triantoro et al. [59] concluded that the user personality moderates the effects of gamification on engagement. The gamification elements implemented and their narrative could be too childish for the sample's age group. Even though Krath et al. [37] referred that social comparisons in the form of leaderboards can have different, and even opposite, effects in different samples, possibly social gamification elements would have been more suited. As referred above, the interconnection between the gamification elements and the survey fill-out could have been enhanced with leaderboards and social elements. These elements could also elicit more complex personality traits, like pride, competitiveness and social status, which could be more suitable for the sample's age group.

This study's findings follow in line with many other studies throughout the bibliography [44, 45, 26, 25, 20, 60]. These studies also found none or minimal contribution from gamification to user engagement. Despite that, these studies found a significant relationship between gamification and the users' motivation to complete the surveys. As Guin et el. [20] concluded: "it is possible to create a more enjoyable survey experience through

gamification, but the hoped-for benefits in engagement did not follow". The motivation and the enjoyability of filling out the surveys are another, more subjective, gamification consequence. The establishment that answering time patterns and data quality levels are not influenced by the gamification effect is not enough to preclude the survey gamification. Motivation and enjoyability can also be important to the outcomes of the survey [54]. The admeasurement of the gamification relation with motivation and enjoyability could be the key factor in establishing it as the default survey design pattern in health studies.

5.2 Limitations and Future Work

Despite its contributions, this study has several limitations. These limitations are related to the sample characteristics, the type of gamification elements applied, how they were applied, and the type of measures made.

The restricted sample group age can limit the generalization of the results since the subjects are all within an age interval of two years and are adolescents. This sample group age can also block the degree of engagement induced by the gamification elements. This is because the gamification elements were possibly not eliciting the best personality traits for the sample group age.

Another limitation is that a significant number of gamification elements were used. This has the possible advantage of enhancing the gamification effect but also removes the possibility of detaching the different gamification elements and measuring their effects separately. The complete segregation between the gamification elements and the survey environment could have led to a lesser influence from the gamification elements in the answering behaviour.

The measurement of the answering time patterns and the data quality could not be enough to conclude about the standardisation of gamification inclusion in health studies. Measuring motivation and enjoyability could be equally important to establish that.

Finally, the single survey-like study design was probably not the best solution to measure the effectiveness of the gamification elements.

These limitations, alongside the study design issues, should guide future work focus. This focus could reflect mainly on the aspects described below.

Further studies, designed as a survey panel, extended through a substantial amount of time are required to fully understand gamification effect on the outcomes. Another study design could also be tested, applying a different gamification element to sub-sets of the gamified group and comparing its effects between them. This would lead to the possibility

of dealing with each gamification element separately and measuring its effects individually, possibly ranking them by their effect power on the outcome. Different gamification element types should also be tested concerning their group age adequacy and relation to the survey filling out environment. If different gamification elements were used, more adequate to the sample characteristics, possibly different outcomes could emerge. In the same way, if the relationship between the gamification elements and the survey fill-out were tightened, possibly even introducing some elements inside the survey, the outcome could be different.

Finally, another type of measurement should be added. Motivation and enjoyability can also play a key role in the gamification wide range of effects, and it should be measured with more valuation.

5.3 Summary

This chapter perambulated the different perspectives built up from the results and related them to each other and the relevant bibliography.

Also enumerated the several study limitations and pointed out possible paths for future works.

The next chapter will present the study conclusions.

Chapter 6

Conclusions

This study's results clearly demonstrate that gamification had a re-engagement effect, increasing the number of logins. It also reveals that gamification did not affect the answering time patterns and survey data quality.

Within this study's limitations, it is safe to say that there was no visible advantage in applying gamification to surveys in this particular health study because the benefits did not overcome the gamification process burden.

Nevertheless, more investigation is needed, mainly in the motivation and enjoyable survey filling-out domain, because they are an engagement expression that can make the survey gamification process worthwhile in health studies.

Regarding the previously defined objectives, it can be concluded that all of them were accomplished. The mobile application was implemented according to the requirements presented by the ISPUP team and the ones poped-in urged by this study design. The ISPUP team is currently using this mobile application to collect data for their SEPIA pain perspective study. Some subjects were exposed to the gamification elements while others were not, and all the collected data were analysed and computed, which allowed the drawing of reliable conclusions.

References

[1] E. Lawrence. A. Ann. *Research methods, statistics, and applications*. SAGE Publications, Los Angeles, 3rd edition, 2018.

[2] Maree Ackehurst and Rose-Anne Polvere. *Increasing survey engagement through gamification and the use of mobile devices*. 2016.

[3] Mohammad Ahsanullah, B M Golam Kibria, and Mohammad Shakil. *Normal and Student's t Distributions*. Atlantis Press, Tampa, 2014.

[4] Winfred Arthur, Ellen Hagen, and Felix George. The Lazy or Dishonest Respondent: Detection and Prevention, jan 2021.

[5] Yap Bee Wah and Nornadiah Mohd Razali. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(November):21–33, 2011.

[6] Paul P. Biemer. Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5):817–848, 2010.

[7] Jörg Blasius and Victor Thiessen. *Assessing the Quality of Survey Data*. SAGE Publications Ltd, 2012.

[8] Ian Bogost. "Gamification Is Bullshit" — The Atlantic. https://www.theatlantic.com/technology/archive/2011/08/gamification-is-bullshit/243338/. Accessed: 2022-03-03.

[9] Aras Bozkurt and Gürhan Durak. A systematic review of gamification research: In pursuit of homo ludens. *International Journal of Game-Based Learning*, 8(3):15–33, 2018.

[10] Stéphanie Carlier, Dries Coppens, Femke De Backere, and Filip De Turck. Investigating the influence of personalised gamification on mobile survey user experience. *Sustainability (Switzerland)*, 13(18), 2021.

[11] Eser Çeker and Fezile Özdamli. What "gamification" is and what it's not. *European Journal of Contemporary Education*, 6(2):221–228, 2017.

[12] Tung Zong Donald Chang and Nicole Vowles. Strategies for improving data reliability for online surveys: A case study. *International Journal of Electronic Commerce Studies*, 4(1):121–130, 2013.

[13] Alexander Corey. *Gamification of Online Surveys for Public Health Data Collection*. Master thesis, University of Guelph In, 2017.

[14] Ali Darejeh and Siti Salwah Salim. Gamification Solutions to Enhance Software User Engagement—A Systematic Review. *International Journal of Human-Computer Interaction*, 32(8):613–642, aug 2016.

[15] Diogo Dores. *Mobile Application for Reporting Chronic Pain Experiences*. Master's thesis, Universidade do Porto – Faculdade de Engenharia, 2021.

[16] Florian Fischer and Sina Kleen. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: Scoping review. *Journal of Medical Internet Research*, 23(1), 2021.

[17] Mirta Galesic and Michael Bosnjak. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2):349–360, 2009.

[18] Philippe Goldammer, Hubert Annen, Peter Lucas Stöckli, and Klaus Jonas. Careless responding in questionnaire measures: Detection, impact, and remedies. *Leadership Quarterly*, 31(4):101384, 2020.

[19] Robert M Groves. Opinion Research – Research on Survey Data Quality. *The Public Opinion Quarterly*, 51(May), 1987.

[20] Theo Downes Le Guin, Reg Baker, Joanne Mechling, and Erica Ruyle. Myths and realities of respondent engagement in online surveys. *International Journal of Market Research*, 54(5):613–633, sep 2012.

[21] Nalika Gunawardena. Choosing the correct statistical test in research. *Sri Lanka Journal of Child Health*, 40:149–153, 2011.

[22] Chad J. Gwaltney, Alan L. Shields, and Saul Shiffman. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health*, 11(2):322–333, 2008.

[23] Juho Hamari. *Gamification. Motivations & Effects*. Doctoral thesis, Aalto University School of Business, 2015.

[24] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does Gamification Work? — A Literature Review of Empirical Studies on Gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences*, January(6-9):3025–3034, 2014.

[25] Johannes Harms, Dominik Seitz, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. Low-cost gamification of online surveys: Improving the user experience through achievement badges. *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 109–114, 2015.

[26] Johannes Harms, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. Gamification of online surveys. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 565–568, New York, NY, USA, oct 2014. ACM.

[27] Steven G. Heeringa, Brady T. West, and Patricia A. Berglund. *Applied Survey Data Analysis: Overview*. 2020.

[28] Dirk Heerwegh. Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1):111–121, 2009.

[29] Tanya Hoskin. Parametric and nonparametric: demystifying the terms. *Ctsa.Mayo.Edu*, pages 1–5, 2010.

[30] IBM. Transforming Different Likert Scales to a Common Scale. https://www.ibm.com/support/pages/transforming-different-likert-scales-common-scale. Accessed: 2020-05-12.

[31] R. Peck. C. Olsen. J. Devore. *Introduction to Statistics and Data Analysis*, volume 59. Thomson Brooks, Belmont, third edit edition, 2008.

[32] T. Grechenig J. Harms, S. Biegler, C. Wimmer, K. Kappel. Gamification of Online Surveys: Design Process, Case Study, and Evaluation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9298:V–VI, 2015.

[33] Sophie Jent and Monique Janneck. Using gamification to enhance user motivation: The influence of gender and age. *Advances in Intelligent Systems and Computing*, 601:3–10, 2018.

[34] Lindsay A. Jibb, James S. Khan, Puneet Seth, Chitra Lalloo, Lauren Mulrooney, Kathryn Nicholson, Dominik A. Nowak, Harneel Kaur, Alyssandra Chee-A-Tow, Joel Foster, and Jennifer N. Stinson. Electronic data capture versus conventional data collection methods in clinical pain studies: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 22(6), 2020.

[35] Florian Keusch and Chan Zhang. A Review of Issues in Gamified Surveys. *Social Science Computer Review*, 35(2):147–166, 2017.

[36] Jan H.G. Klabbers. On the Architecture of Game Science. *Simulation and Gaming*, 49(3):207–245, jun 2018.

[37] Jeanine Krath, Linda Schürmann, and Harald F.O. von Korflesch. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, 125(July):106963, 2021.

[38] Nojin Kwak and Barry Radler. A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics*, 18(2):257–273, 2002.

[39] Richard N. Landers, Elena M. Auer, Andrew B. Collmus, and Michael B. Armstrong. Gamification Science, Its History and Future: Definitions and a Research Agenda. *Simulation and Gaming*, 49(3):315–337, 2018.

[40] Michael W. Link, Jennie Lai, and Kelly Bristol. Not so fun? the challenges of applying gamification to smartphone measurement. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8520 LNCS(PART 4):319–327, 2014.

[41] Jemma Looyestyn, Jocelyn Kernot, Kobie Boshoff, Jillian Ryan, Sarah Edney, and Carol Maher. Does gamification increase engagement with online programs? A systematic review. *PLoS ONE*, 12(3):e0173403, mar 2017.

[42] Garm Lucassen, Fabiano Dalpiaz, Jan Martijn E.M. van der Werf, and Sjaak Brinkkemper. The use and effectiveness of user stories in practice. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9619, pages 205–222. Springer, Cham, 2016.

[43] Salvatore S. Mangiafico. R Handbook: Introduction to Likert Data. https://rcompanion.org/handbook/E_01.html. Accessed: 2022-06-02.

[44] Aigul Mavletova. Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review*, 31(6):725–743, 2013.

[45] Aigul Mavletova. Web Surveys Among Children and Adolescents: Is There a Gamification Effect? *Social Science Computer Review*, 33(3):372–398, 2015.

[46] Adam W. Meade and S. Bartholomew Craig. Identifying careless responses in survey data. *Psychological Methods*, 17(3):437–455, 2012.

[47] Alberto Mora, Daniel Riera, Carina Gonzalez, and Joan Arnedo-Moreno. A Literature Review of Gamification Design Frameworks. *VS-Games 2015 - 7th International Conference on Games and Virtual Worlds for Serious Applications*, 2015.

[48] Benedikt Morschheuser, Juho Hamari, and Jonna Koivisto. Gamification in crowdsourcing: A review. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2016-March:4375–4384, 2016.

[49] Kayque W.R. Oliveira and Melise M.V. Paula. Gamification of Online Surveys: A Systematic Mapping. *IEEE Transactions on Games*, 13(3):300–309, 2021.

[50] Chathurangi H. Pathiravasan, Yuankai Zhang, Ludovic Trinquart, Emelia J. Benjamin, Belinda Borrelli, David D. McManus, Vik Kheterpal, Honghuang Lin, Mayank Sardana, Michael M. Hammond, Nicole L. Spartano, Amy L. Dunn, Eric

Schramm, Christopher Nowak, Emily S. Manders, Hongshan Liu, Jelena Kornej, Chunyu Liu, and Joanne M. Murabito. Adherence of Mobile App-Based Surveys and Comparison with Traditional Surveys: ECohort Study. *Journal of Medical Internet Research*, 23(1), 2021.

[51] Maja Pohar and Janez Stare. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81(3):272–278, 2006.

[52] R. Khaled. L. Nacke S. Deterding, D. Dixon. From Game Design Elements to Gamefulness: Defining "Gamification". *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011: Preface.*, pages 9–15, 2011.

[53] Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69:371–380, 2017.

[54] Silvia Schacht. Web Survey Gamification – Increasing Data Quality in Web Surveys by Using Game Design Elements. *Proceedings of the Twenty-Fifth European Conference on Information Systems (ECIS)*, 2017(July):2907–2917, 2017.

[55] Fritz Scheuren. What is a survey? American Statistical Association Alexandria, 2004.

[56] Katie Seaborn and Deborah I. Fels. Gamification in theory and action: A survey. *International Journal of Human Computer Studies*, 74:14–31, 2015.

[57] Roger Tourangeau, Hanyu Sun, Ting Yan, Aaron Maitland, Gonzalo Rivero, and Douglas Williams. Web Surveys by Smartphones and Tablets: Effects on Data Quality. *Social Science Computer Review*, 36(5):542–556, 2018.

[58] Tamilla Triantoro, Ram Gopal, Raquel Benbunan-Fich, and Guido Lang. Would you like to play? A comparison of a gamified survey with a traditional online survey method. *International Journal of Information Management*, 49:242–252, 2019.

[59] Tamilla Triantoro, Ram Gopal, Raquel Benbunan-Fich, and Guido Lang. Personality and games: enhancing online surveys through gamification. *Information Technology and Management*, 21(3):169–178, 2020.

[60] Susanna Warnock and J. Sumner Gantz. Forum: Gaming for respondents: A test of the impact of gamification on completion rates. *International Journal of Market Research*, 59(1):117–138, 2017.

[61] Jamie Woodcock and Mark R Johnson. Gamification: What it is, and how to fight it. *Sociological Review*, 66(3):542–558, 2018.

[62] P A Nanjari Wyss. Gamified Surveys : what game mechanics are driving people ' s motivation ? pages 1–50, 2019.

[63] Yongwen Xu. Literature Review on Web Application Gamification and Analytics. Technical Report April, Department of Information and Computer Sciences – University of Hawai'i, Honolulu, 2012.

[64] B. W. Yap and C. H. Sim. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, 2011.

[65] Young K., Menegazzi J., and R. Lewis. Statistical Methodology: IX. Survival Analysis. *Academic Emergency Medicine*, 6:244–249, 1999.

Appendix A

Domain Model

Figure A.1 shows the solution domain model UML diagram.

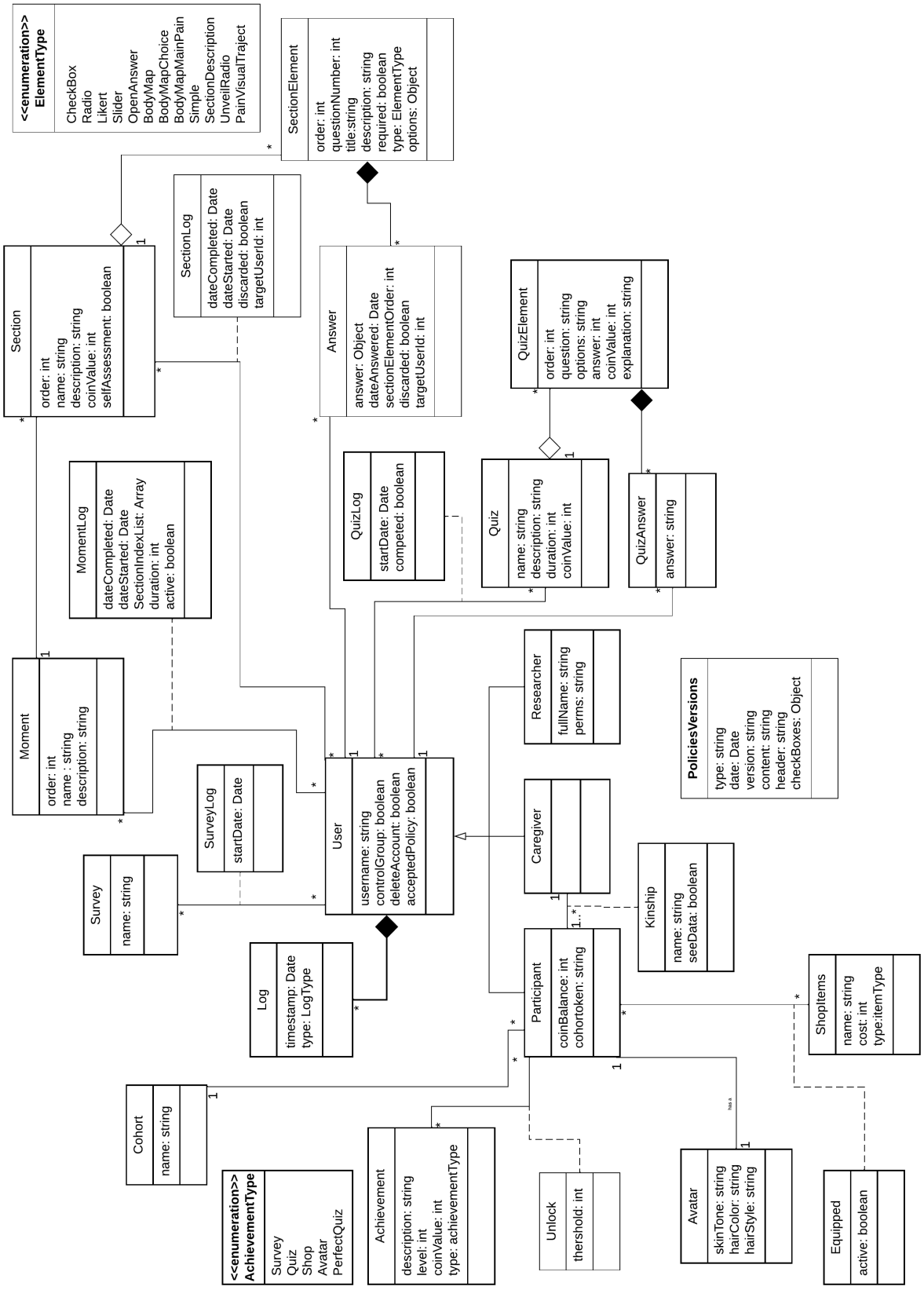


Figure A.1: Solution Domain Model

Appendix B

Pilot Study Usability Survey Results

Figure B.1 on page 63 shows the post-pilot phase survey results for each question referenced in Table B.1. The scale varies from 1 — Strongly Disagree, to 5 — Fully Agree.

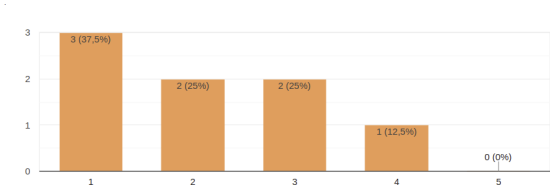
Table B.1: Post-Pilot Phase Survey Questions

	Question	Image
1	I think that the application is unnecessarily complex	Image B.1a
2	I think that the application is easy to use	Image B.1b
3	I think I would need help to fully utilize this application	Image B.1c
4	I think that the different application elements are well integrated	Image B.1d
5	I think the application shows inconsistencies	Image B.1e
6	I felt confident using the application	Image B.1f
7	I think it was an easy process to learn how to use the application	Image B.1g
8	I needed much time to get familiar with the application functionality.	Image B.1h

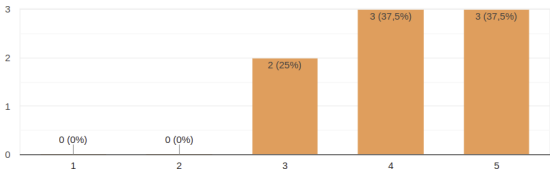
Continued on next page...

Table B.1 – continued from previous page

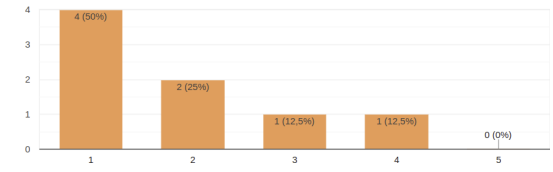
	Question	**Image**
9	I think it was necessary to think hard before answering some questions	Image B.1i
10	I think the questionnaires are too long	Image B.1j
11	I think the store is an easy-to-use element	Image B.1k
12	I found the three quizzes interesting	Image B.1l



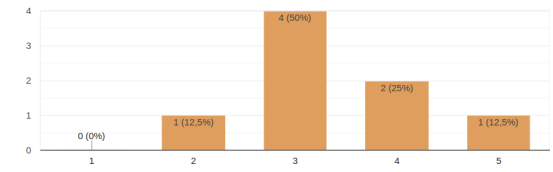
(a) Question 1



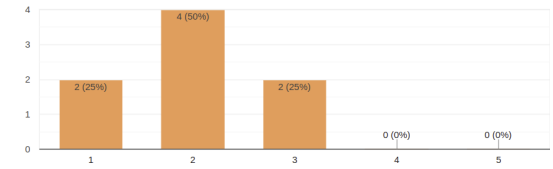
(b) Question 2



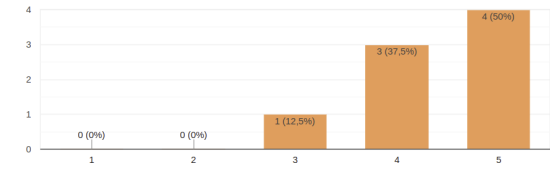
(c) Question 3



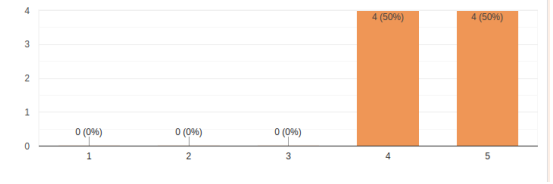
(d) Question 4



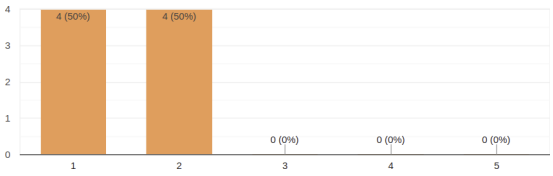
(e) Question 5



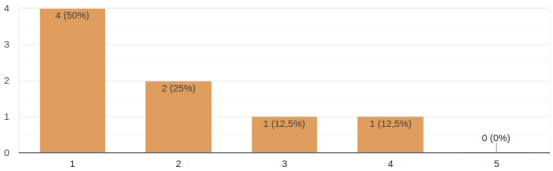
(f) Question 6



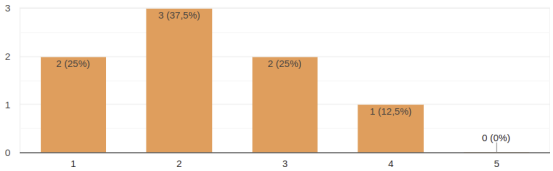
(g) Question 7



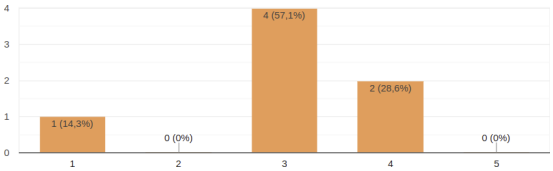
(h) Question 8



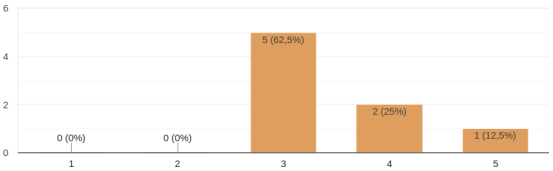
(i) Question 9



(j) Question 10



(k) Question 11



(l) Question 12

Figure B.1: Pilot Survey Results