SAARLAND UNIVERSITY

Faculty of Mathematics and Computer Science
Department of Computer Science
MASTER THESIS

Investigating the Effect of Choice between Fixed Gamification Interventions in a Gamified Online Survey

submitted by
Swathi Krishnaraja
Saarbrücken
March 2021

Advisors:
Dr. Pascal Lessel
German Research Center for Artificial Intelligence
Saarland Informatics Campus
Saarbrücken, Germany

Reviewers:
Prof. Dr. Antonio Krüger
German Research Center for Artificial Intelligence
Saarland Informatics Campus
Saarbrücken, Germany

Dr. habil. Helmut Horacek
German Research Center for Artificial Intelligence
Multilinguality and Language Technology
Saarbrücken, Germany

Saarland University
Faculty MI – Mathematics and Computer Science
Department of Computer Science
Campus - Building E1.1
66123 Saarbrücken
Germany

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Pascal Lessel for giving me the opportunity to write my master's thesis under his guidance. A huge thanks to him for providing me with a great deal of support and patient assistance throughout the thesis, and for always providing me with the most competent feedback, and pushing me in the right direction. Your insightful feedback has truly shaped my thinking, and taught me valuable lessons that I could take away for the rest of my academic career. Thanks for being my biggest inspiration.

I would like to thank and convey my sincere gratitude to my supervisor Professor Antonio Krüger for giving me the opportunity to perform my master's thesis under his supervision at the German Research Center for Artificial Intelligence. Your invaluable feedback during the presentations has truly been influential in shaping my research. In addition, a special thanks goes to my second reviewer Professor Helmut Horacek for reviewing this thesis.

I would like to thank doctoral researcher Maximilian Altmeyer for all his support during the initial stages of thesis, by providing valuable insights during the thesis discussions, and for helping me with research materials.

Furthermore, I thank my family and friends for all their encouragement, moral support, and motivation throughout the thesis. Above all, my deepest gratitude goes to my fiancé, Gowtham, for his immense love and support.

Last but not the least, I would like to thank each and every participant for taking part in my study.

Abstract

Gamification is considered to be the application of game design elements to non-game contexts. Gamification has already proven to be beneficial in surveys for improving user motivation, performance, and the overall quality of responses. However, not all studies prove the above-mentioned benefits. This is mainly due to the development of systems based on a one-size-fits-all approach i.e. by providing the same game elements to all user types. But there is a need to understand that people's preferences are a sum of different characteristics and that each user exhibits different attitudes towards gamification. Therefore, in this thesis, we employ the simplest forms of customization. Customization is an approach that allows users to make selections and set preferences in a system. This is done with the intent of giving users the ability to control certain aspects of the system and provide them with a sense of fulfillment and autonomy. Studies from psychological theory and research suggest that giving people the ability to choose, increases their intrinsic motivation, perceived control, task performance, and overall life satisfaction and happiness. In order to investigate the effectiveness of providing choice, we employ it in the context of surveys, as surveys are considered to be one of the most important tools to make inferences about an entire population, people's attitudes, perceptions, intents, habits, awareness, experiences, and characteristics. But the main problem with online surveys is identified to be about maintaining user engagement and motivation. Thus, providing customization in gamified online surveys, allows us to realize the needs of users and provides a sense of control in choosing their preferred survey version. In this thesis, we conduct two experiments: one experiment to understand the effectivity of providing the choice to a user to either enable or disable a fixed game element. The second experiment is conducted to understand the effectivity of providing choice between two different gamified versions of a survey, but without the option to disable gamification. For the first experiment we had N = 80 participants, and for the second experiment we had N = 60 participants, assigned to the different study conditions we had. Our results mostly show no significant differences between different survey versions suggesting that survey experience can only be improved by having interesting and compelling questions, along with relevant gamification applied to the survey. Overall, users are curious and interested to try different gamified survey versions. These results helps us to conclude that overall gamification is considered to be effective in the context of surveys unless appropriate and relatable game elements are applied to the survey theme. Overall, we conclude from the studies conducted that providing choice in choosing the preferred gamification in a survey is perceived positively by users.

Contents

Chapter 1
Introduction

Gamification is increasingly becoming pertinent [7], not only in games, but also in a variety of domains ranging from personality assessment [38, 39], to image tagging [22, 26], and market research [6]. Several studies have indicated that the application of gamification is effective only when the gamification matches users' individual preferences and personalities [37]. However, mapping user personality onto game design elements is quite hard. To address this situation, we explore two simple and effortless forms of **customization**, which allows users' to change their preferences with a single click of a button - providing the ability to either enable or disable gamification, or the ability to choose between two fixed gamified interventions. In this thesis, we conduct two experiments to examine the effect of having choice and how this choice in turn affects users' performance and intrinsic motivation.

In this chapter, we give a short overview on customization of interactive systems, and how customization could influence the system's perceived effectiveness. Then we highlight a few studies that show the importance of providing choice and autonomy to users' and how it could help in obtaining positive effects on users' experience and performance. Later, we give a brief overview of the context in which we would like to investigate this study. i.e. gamified online surveys. In section 1.2, we present the specific research goals that we would like to achieve, and also discuss the concrete research questions that were extracted from our research goals. In section 1.3, we conclude by providing an overview on the remaining parts of this thesis and the overall thesis structure.

1.1 Motivation

Customization is an approach that allows individuals to make selections and set preferences in a system [32]. This is done with the intent of giving users more control over the user experience. Additionally, customization is a way to personalize the interactive systems to suit individuals of different capabilities and dispositions [28]. In general, customization involves activities such as moving items around an interface to

reflect user's priorities, selecting the content, layout, functionality, or altering colors and other design aspects that are appealing to the users [32]. This approach of customizing based on one's preferences is considered as a means to empower users [32]. Schöbel et al. [33] identifies a number of studies from psychological theory and research, demonstrating that giving people the ability to choose, can increase their intrinsic motivation, perceived control, task performance, and overall life satisfaction and happiness. However, there are a few drawbacks with customization: increased user effort, and too many choices can often be misleading. To address this problem, we employ the simplest form of customization, with the aim of minimizing the user effort, and yet be able to achieve the benefits of customization. To investigate different forms of choices, we conduct two experiments, each aiming at providing minimal customization options.

1. **Experiment 1: Enabling or Disabling Gamification**
   Investigating the effectiveness of customization by allowing users to either enable or disable gamification. This allows us to target the users who dislikes gamification. This experiment is conducted with the intent to identify if providing users with the choice to disable gamification exerts positive user behavior, thereby improving their performance and intrinsic motivation.
2. **Experiment 2: Selecting from Fixed Game Configurations**
   Investigating the effectiveness of customization by providing users with two different versions of the survey each containing different game elements. Here, the users will be given the freedom to choose their preferred version (or game configuration). This experiment is conducted with the intent to identify if allowing users to choose their own game configurations exerts positive motivational outcomes accompanied by a sense of autonomy.

Thus, this thesis aims to investigate the influence of simple customization on user's behavior and perception.

1.1.1 Customization and Gamification

Gamification is defined as the application of game design elements such as badges, narratives, points, levels, leaderboards, progress, unlockables, etc, in non-game contexts [7]. Gamification has been actively employed in a variety of domains (A few examples are listed here: Market Research [6]; Sports and Leisure activities [14, 15]; Recycling [19]; Manufacturing [20]; Image Tagging [22, 26]; Gastronomy [29]). The concept of gamification is promising as it is considered to be one of the most efficient techniques [25] to motivate consistent participation and longterm engagement with a product or service. A gamification approach called "top-down gamification" has been employed in majority of applications, where users' are provided with a fixed gamification system i.e. all the game elements present in the system are fixed and cannot be altered. However, it is important to understand that one-size-fits-all approach does not work effectively with respect to gamification, as different game elements support different user types and personalities [37]. One approach to overcome this problem is to tailor gamification according to each user type. This can be achieved by either personalizing or customizing a system. Personalization allows the system to adapt itself automatically based on people's likes and preferences [28]. This approach is done by analysing player types or personality traits, and thereby altering the system. However, personalization approaches

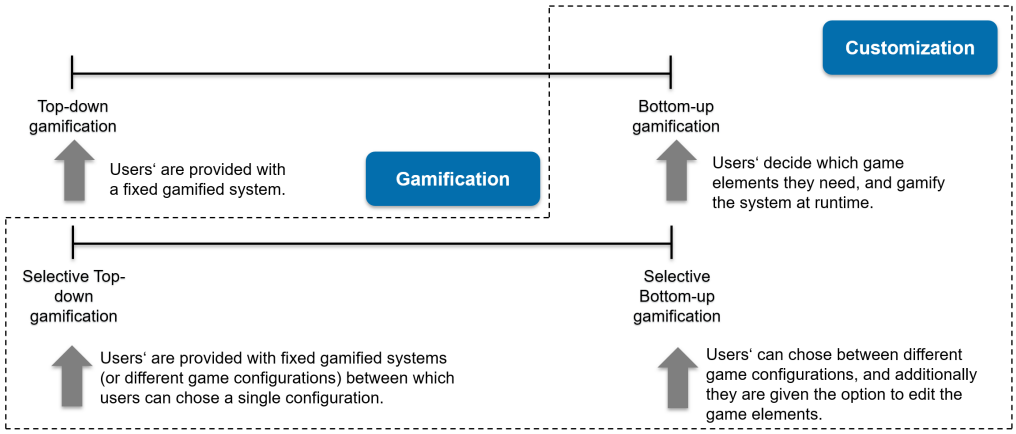


Figure 1.1: Overview of gamification and degrees of customization.

tend to fail in some situations and provides inappropriate setups for some users [5]. On the contrary, customization is an approach that allows users to adapt the game design elements and game setup according to their preferences at runtime [28]. This approach provides the flexibility to adapt the system and allows user to make their own choices.

However, there are different degrees of customization options that one can provide to a user. Figure 1.1 depicts the different forms or levels of customization. A system's degree of customization can be determined based on the amount and complexity of the configuration options. As the number of options to configure a system increases, so does the level of complexity. Such a system, that requires high user effort, can be categorized under high form or high degree of customization. On the contrary, when there is limited choice (or fewer options) for the user's to customize the system, it is referred to as a system with low degree of customization. Bottom-up gamification, selective top-down gamification and selective bottom-up gamification (see Figure 1.2) are few customization approaches that allow users to define their own gamification setup with certain restrictions based on the system's design. The different degrees of customization will be explored in detail in section 2.1.

Bottom-up gamification: Bottom-up gamification allows users to decide whether they want to use gamification at a system's runtime. If the user chooses to use gamification, they are given the ability to adapt all available game elements in the system and combine them as they see fit.

Selective top-down gamification: Selective top-down gamification is a type of customization where users have the option to select between different "top-down" defined game configurations. This allows the users to choose the game configuration they want, without having the option to edit or add a new configuration.

Selective bottom-up gamification: Selective bottom-up gamification is a type of customization where users have the option to select between different "top-down" defined game configurations. Additionally, they have the option to edit or adjust the

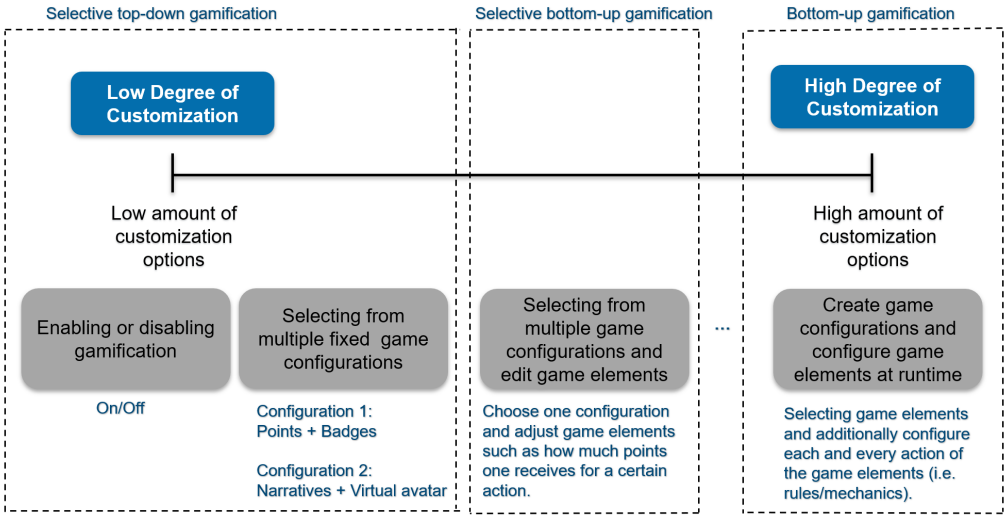


Figure 1.2: Detailed overview of different degrees of customization and their respective categories in gamification.

game elements that are present within the selected game configurations.

Previous studies [21, 22, 34] has shown that allowing users' to customize their system can lead to high levels of self-efficacy, freedom, and motivation. While complex forms of customization yields significantly positive results, they are problematic when it comes to the effort needed from the user to configure the game setup [20, 21, 34]. Additionally, providing too many options to configure leads to "choice overload" - a condition where users feel overwhelmed with the range of choices provided, and thereby resulting in detrimental effects. This served as a driving factor for us to investigate whether positive effects can be achieved with minimum amount of customization options. To answer this question, we chose the context of gamified online surveys (see subsection 1.1.3, where we present two simplified forms of customization to users with a bare minimum of configuration options. i.e. In experiment 1, users can either enable gamification from a fixed gamification setup, where the system is already designed with certain game elements, or users can disable the provided gamification setup. In experiment 2, users can select between two fixed gamification setups, where each of the gamified setup contains different set of game elements. However, in this experiment, users are not provided with the option to disable gamification. Investigating these two approaches helps us to explore the simplest form of customization and also enables us to gain understanding on how having a choice affects users' behavior in a gamified online survey.

1.1.2 The Effect of Choice and Autonomy in Gamified Interventions

In this section, we discuss the aspects that led us to choose the selective top-down gamification for developing our survey platform. As discussed previously, there are a number of positive effects (self-efficacy, control, freedom) of providing choices to user's in customizing their system. However, an increased user effort might lead to analysis

Figure 1.3: The effect of choice in three experimental studies: gourmet jams, chocalates, and essay assignment topics.

paralysis [33].

Current research [10, 16, 21, 33] shows that, as the number of options to modify the system increases, so does the level of complexity of the decision itself. Iyengar & Lepper [16] conducted three experimental studies with gourmet jams, chocalates, and essay assignment topics. In each of the three experimental studies, users were either given with a limited array of 6 choices, or an extensive array of 24 to 30 choices (see figure 1.3). Findings from this study show that people tend to enjoy the task, and feel satisfied only when provided with less number of options (i.e. 6 choices). It is also notable that people performed better in such limited-choice contexts. Haynes [10] conducted a similar experiment, where a limited array of 3 choices were compared with an extensive array of 10 choices to choose a prize. This experiment included an additional factor, that limits the time allocated to make a decision (limited vs. extended decision time). Results show that participants felt dissatisfied, regretful, difficulty and frustrated in making a decision with 10 choices within the given time period. However, participants in the limited choice (i.e. 3 options) condition experienced higher satisfaction and tend to enjoy the task more compared to participants with extensive choice or set of options. Both these studies [10, 16] indicate that choice overload might eventually lead to frustration, dissatisfaction, and disinterest in the task, resulting in decreased task performace. Schöbel et al. [33] posits that people feel most confident in their decisions when they understand the available options and can comfortably compare and evaluate them. In other terms, providing autonomy and freedom in choosing, along with minimum possible effort is merely not enough, but important.

Considering the above-mentioned benefits of offering limited choice, we, in this thesis, focus on examining the effectivity of providing a simple choice in a gamified context.

1.1.3 Gamification in Online Surveys

In order to investigate the effect of the simplest form of customization, we chose the context of gamified online surveys. Surveys are considered to be one of the most important tools to make inferences about an entire population, people's attitudes, perceptions, intents, habits, awareness, experiences, and characteristics [27]. However, the main problem with online surveys were identified to be about maintaining user engagement and motivation. In order to overcome this problem, gamification was

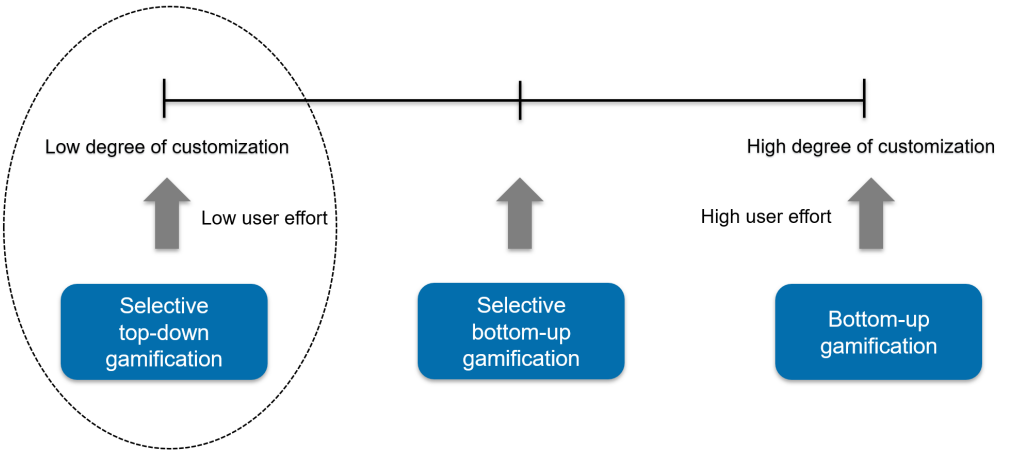


Figure 1.4: Spectrum showing the levels of customization. The gamified survey platform of our thesis will be developed based on selective top-down approach.

introduced in online surveys [24, 29, 38, 39]. Even though, some studies showed that gamification in online surveys improved user performance and user experience, not all studies showed positive effects [24, 29]. This is mainly due to the fact that not all users prefer gamification in a system, and the preferences and likes of all users are not the same. As a measure to overcome this issue in gamified online surveys, we introduce two simple forms of customization, that requires minimum effort from the user, to choose their preferred survey version.

The main focus of this thesis is to examine the impact of choice in the presented gamified system. In the past, different degrees of customization has been investigated in the areas of games and gamification related research. This thesis can be seen as a contribution to a context (gamified online survey), that has not been investigated with this viewpoint before.

1.2 Research Goals

As already introduced in subsection 1.1.1, customization requires a certain amount of effort from the user to make modifications and adapt the system according to one's needs and preferences. The main goal of this thesis is to minimize the user effort in customizing and have a closer look on the influence of customization on users' behavior and perception, in terms of self-efficacy, feeling of control, and freedom.

The overarching research question of this thesis is that, which effects does choice have in gamified interventions. In order to investigate this from different angles, we divided the study into two parts: one - to investigate the effectiveness of providing choice that allows users' to either enable gamification or disable gamification in the system, and a second study - to investigate the effectiveness of choice that allows users' to choose

between different gamified interventions consisting of different game configurations (or game elements), but without the option to disable gamification.

The following points summarizes the concrete research questions of this thesis:

Experiment 1: Enable or Disable Gamification

(a) Does the choice to enable or disable gamification in a survey positively affect user's psychological outcomes?

(b) Does the choice to enable or disable gamification in surveys affect user's behavioral outcomes?

Experiment 2: Choice in Gamified Interventions

(a) Does the choice of choosing between different game configurations in gamified surveys positively affect user's psychological outcomes?

(b) Does the choice of choosing between different game configurations in gamified surveys affect user's behavioral outcomes?

Both **Experiment 1** and **Experiment 2** measure users' behavioral outcomes (i.e. how customization has influenced a users' performance in answering the survey) and users' psychological outcomes (i.e. how customization has affected users' overall survey experience and satisfaction). Hence, in this thesis, we investigate the influence of two different customization approaches towards gamified systems. In order to study this effect, a between-subject experimental design was chosen for every study. The development of the gamified survey platform with the above-mentioned customization options was part of the work for this thesis.

1.3 Outline

In this section, the structure of the remaining parts of this thesis is outlined. The next chapter gives an overview of the **Related Work** conducted in the context of this thesis. It includes existing research on customization and different degrees of customization options in gamification, the effect of choice and autonomy in gamified interventions, and present works on gamified online surveys. It also discusses the variables measured in previous research on gamified online surveys. The **Study Platform** chapter briefly describes the methods, and technologies used for the implementation of the gamified survey platform. Additionally, it gives an overall understanding of the application development and schema structure. The **User Study** chapter gives a brief introduction to the two experiments conducted in this thesis. A detailed explanation on the experimental procedure and experimental design is presented, and the different aspects considered for the development of our gamified survey platform is discussed. Subsequently this chapter presents the overall experimental design methods and hypotheses for our study. Additionally, this chapter summarizes the obtained results for the two experiments conducted in this thesis. In the final chapter **Conclusion**, the conclusion of the thesis is discussed in detail including the contributions, limitations and further recommendations for future work. The **Appendix** and **Bibliography** can be found at the end of this thesis.

Chapter 2
Related Work

This chapter gives an overview of the research conducted in context of this thesis and is structured as follows: In section 2.1, we briefly introduce customization and the different degrees of customization options that are possible within gamification. This section, also helps us to understand how different forms of customization minimizes or maximizes user effort. In subsection 2.1.2, we take a closer look at the existing works on choice and autonomy and we discuss the different effects choice gives to an individual, and why it is important. This section also briefly discusses the impact of providing freedom of choice to users and how it is interconnected with one's behavioral and psychological outcomes. Afterwards, in section 2.3, we present various studies that investigate different game design elements in an experimental study context i.e. gamified online survey. Also, we discuss the motivational power of game elements and the specific effects of gamification on psychological need satisfaction. Lastly, in section 2.4 we discuss the variables that were measured in previous studies, both behavioral and psyhchological variables, as this is an integral part in identifying the dependent variables that we will be using in our study.

2.1 Degrees of Customization in Gamification

As illustrated in subsection 1.1.1, customization is important with respect to gamification, as one-size-fits-all approach of fixed gamification systems does not necessarily work well with all user types and at all situations. In order to be effective, gamification should be tailored to users' expectations and individual preferences. However, adapting game elements to individuals preferences can be a complex task, as users can have different preferences towards gamification and different motivations for performing a task. In this chapter, we will elaborate on different degrees of customization, where every user receives certain customization options that allows them to adapt and adjust the system as they see fit.

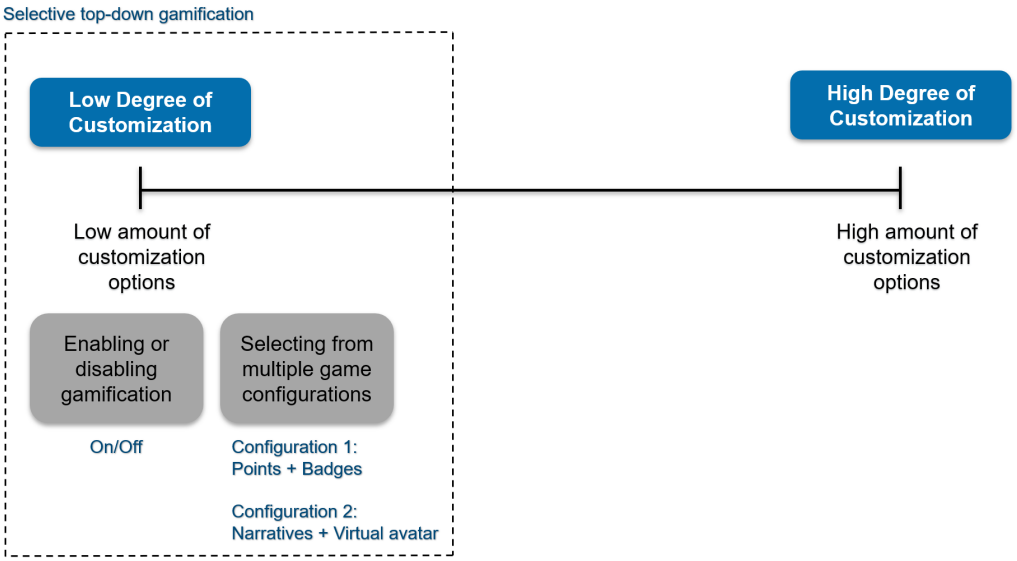


Figure 2.1: Example of low degree of customization.

2.1.1 Low Degree of Customization in Gamification

A customization approach that requires minimum effort from the user to set up the game configuration in a system is referred to as a system with lowest degree of customization (see Figure 2.1). In this section we provide examples of previous work that contains minimal amount of customization options.

To better understand low degree of customization in gamification, let us first understand the concept of top-down gamification, and selective top-down gamification. These two types of gamification requires limited amount of effort from the user to experience or setup gamification.

Top-down gamification: Top-down gamification is an approach where users are provided with a fixed game configuration i.e. users will not have the choice to remove or disable the game configuration that has been applied to the system. Additionally, top-down gamification does not allow users to edit or change the existing game elements that are present in the system.

Selective top-down gamification: Selective top-down gamification is a customization where users have the option to select between different "top-down" defined game configurations. This allows the users to choose the game configuration they want, without having the option to edit or add a new configuration.

The work of Lessel et al. [22] serves as an example of lowest degree of customization in gamification, where the authors provide users with the ability to enable or disable gamification. The authors conducted their experiment in the context of an image tagging task. The image tagging task works in the following way: Firstly, an image of an abstract painting will be displyed to the participant for 5 seconds. Secondly, the participant has

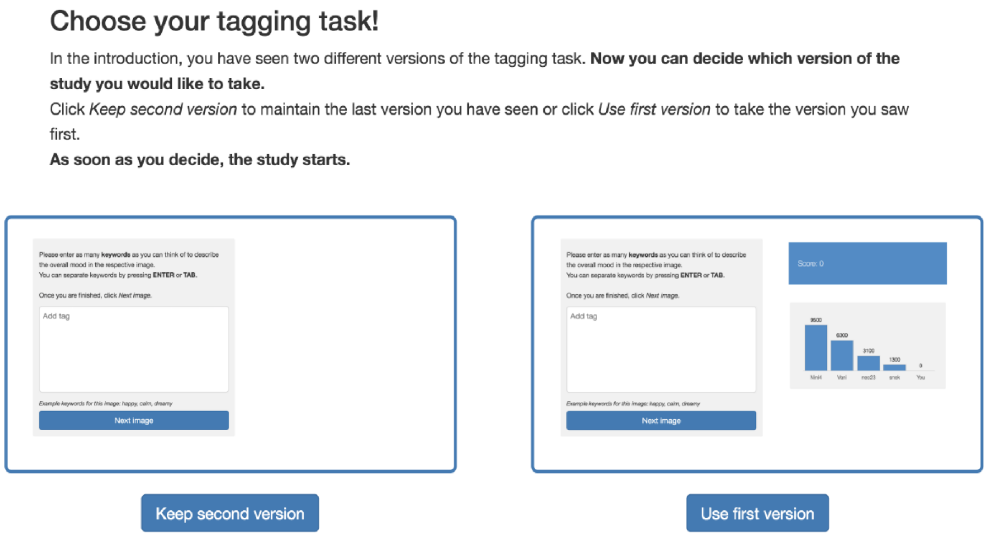


Figure 2.2: The two versions of image tagging task offered to the users for choosing their preferred version. Non-gamified image tagging task (left) and Gamified image tagging task (right).

to provide any number of meaningful tags for each of the displyed image. Both, the number of generated tags, as well as the quality of tags were used for analysis.

The experiment comprised of a total of 77 participants, each of them being assigned to three different conditions namely gamification, no gamification and choice in gamification. Participants in the gamification condition were adopted with a gamified version of the image tagging task, where they experienced different game elements such as points and leaderboard. This condition allows participants to receive 100 points for each of the generated tag, irrespective of the tag quality. Additionally, a participant could compare their performance on a leaderboard. In the no-gamification condition, the participants were provided with the image tagging platform without any game elements i.e. the participant will have to generate tags based on the given abstract, and will not be awarded with any points for performing the activity. The third condition of the study involved offering choice to the participants (see Figure 2.2), where one can choose whether the gamification should be kept active or not. This was done by allowing the users to experience both the versions (gamified and non-gamified) through a guided tour. After this, the users will be directed to a page where the users can select the desired version of the image tagging platform (see Figure 2.2) through a single click of a button i.e. either enabling gamification, or by disabling gamification (on/off). The participants who chose the option of enabling gamification, will be given with a gamified image tagging setup consisting of points and leaderboard. Similarly, the participants who chose to disable gamification will be provided with a non-gamified image tagging platform with no game elements.

The authors derived from the study that, the participants in the gamification condition provided significantly more tags compared to the participants in no-gamification condition.

Figure 2.3: The four reward systems used in the *Cafe Flour Sack* game.

These findings were in line with the previous study that was conducted by Mekler et al. [26] (we do not go deeper into the study conducted by Mekler et al., as it is not relevant to the topic of customization, but rather of pure gamification). This suggests that participants are motivated by gamification and the usage of game elements. Furthermore, the participants in the choice condition also performed significantly better in generating more tags and with same quality, than the participants who were provided with fixed gamification setups. This suggests that, even the simplest form of customization can motivate participants to perform better in an image tagging task by generating more tags.

Hence, this is an example of the simplest form of customization which takes into consideration the choices of the user, and yet minimizes the overall user effort to attain the desired customization. Furthermore, this form of customization is especially beneficial for participants who does not like gamification, and therefore has the ability to disable gamification.

2.1.2 Moderate Degree of Customization

A moderate degree of customization is when the customization options allow users to alter some of the game elements or gamification related attributes, with a moderate user effort.

Selective bottom-up gamification: is a type of customization where the users have the option to select between different "top-down" defined game configurations. Top-down defined game configurations are those systems that already contains certain game elements. In this type of gamification, users choose one game configuration and adjust each of the game elements that are present within that game configuration.

Siu and Riedl [34] conducted a study in which they created a cooking themed game called *Cafe Flour Sack*, that asks players to classify cooking ingredients for potential recipes. The game provides four different reward systems or reward categories (see Figure 2.3)

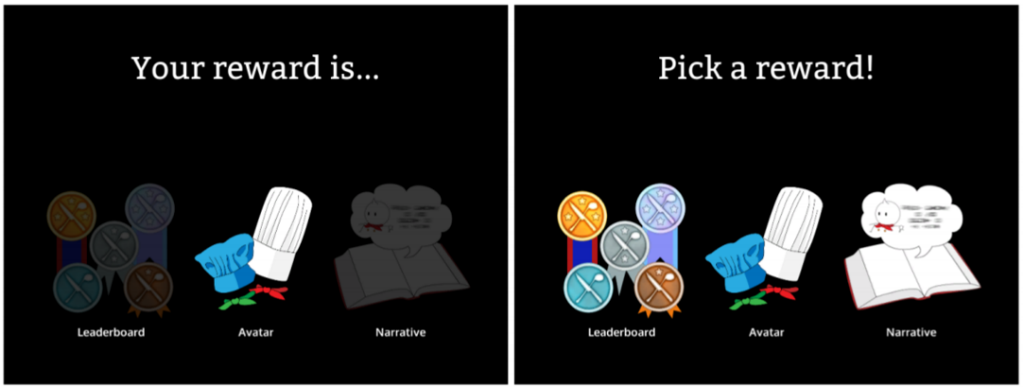


Figure 2.4: *Cafe Flour Sack* Left: The reward is assigned randomly to the player. Right: The players can choose their own reward from the given categories.

for players to interact with: global leaderboard (to display ranks of players relative to other players), customizable avatar (a 2D avatar chef can be customized using digital items), unlockable narratives (can unlock short stories set within the game), and global progress tracker (can track their overall contribution to the tasks relative to other players). Depending on the condition, participants were either able to select their own rewards (see Figure 2.4) from the reward categories (choice condition), or they were randomly given with a reward (random condition). The results were evaluated based on two factors: task completion (i.e. correctness of completed tasks, number of completed tasks, timing of completed tasks), and player experience (i.e. reward preference, perception of choice, duration of play, and boredom). Results show that players in the choice condition had significant effects on task performance, by completing the tasks faster and with high correctness, and similar player experiences. Although, this type of customization does not allow users to modify any of the rewards further, it gives the ability to choose one's own reward, ultimately leading to better performance and motivation.

2.1.3 High Degree of Customization

An example of high degree of customization in gamification involves higher user effort to set up different customization options in a system during runtime. i.e. the system allows the user to set up their own gamification configuration, by choosing the game elements they want and additionally setting the rules and mechanics of each of the game elements. This type of customization requires a greater amount of effort from the user.

Bottom-up gamification: Bottom-up gamification allows users to decide whether they want to use gamification at a system's runtime. If the user chooses to use gamification, they are given the ability to adapt all available game elements in the system and combine them as they see fit.

In order to see the validity of "bottom-up gamification", Lessel et al. [20] conducted an experiment using an imaginary (non-prototypical/no concrete realization) task management application. This was done with the intent to assess users expectations without biasing

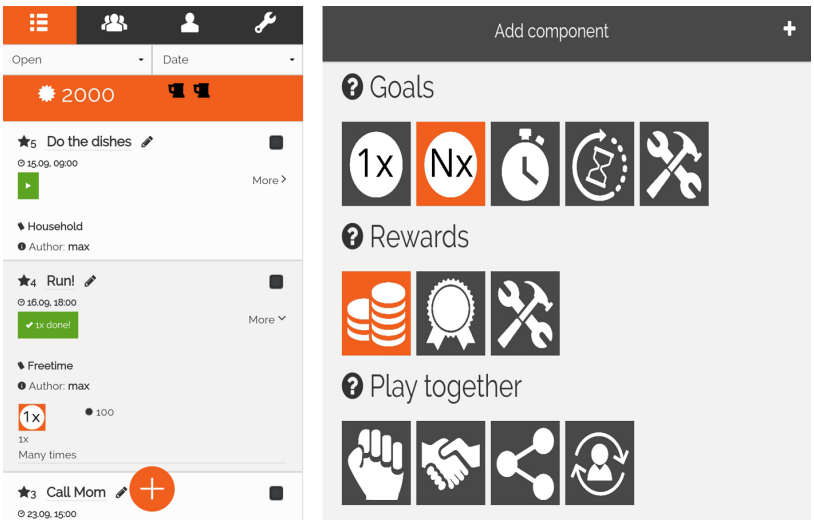


Figure 2.5: The *BU-ToDo* app. Left: Main task interface. Right: Game element configuration screen.

them with a prototypical realization, and also allows to get an idea of how users react to self-tailored gamification. Through an online questionnaire, participants were asked about their demograohics, gaming affinity, their experience with gamification, and how they motivate themselves to do day-to-day unpleasant tasks, and how they perceive "bottom-up" gamification in different contexts (participants were explained with an abstract idea of how bottom-up gamification functions, what are game elements and how these game elements work). Participants were then provided with four scenarios: *cleaning*, *piece work*, *exercising*, *energy saving* and asked to indicate how they would motivate themselves in those situations, and how they could use the game elements to keep themselves motivated. Results show that participants who were open to motivate themselves in private/work life using an app, were significantly open towards using the bottom-up gamification concept. However, there were mixed expressions towards the perception of game elements, and towards using motivational app in their daily life, highlighting that there could be individual and contextual differences in the perception of gamification. This suggests that following a bottom-up gamification approach can be beneficial and reasonable, as it accounts for individual differences in defining or selecting one's own customization.

As a follow-up study of the validated bottom-up gamification concept, Lessel et al. [20] conducted an experiment using a task management application named "BU-ToDo" (see Figure 2.5), a mobile web app that was created by the authors, offering bottom-up gamification elements. The BU-ToDo app offered the flexibility to customize the game elements, thereby providing a certain degree of freedom in the amount of customization options.

Figure 2.5 (left) depicts the task overview screen of the app. It shows an overview of all the tasks that was created by the user. A user can create a new task by providing a name, description, category, priority, and a due date. Additionally, tasks can be set to reoccuring, and reminders can be added and configured. For every task, users can decide which

game elements they want to use for that particular task, allowing them to use different combinations of game elements for each of the tasks created. Figure 2.5 (right) depicts the game configuration screen where game elements are categorized into three groups: *Goals, Rewards, Play together*. Game elements under the *goals* category include number of times a task needs to be solved (1x, Nx), timer element that specifies a time frame, and also allows to specify how long a task should be performed. *Rewards* category includes points, that can be achieved when a task is completed, badges indicating achievements. Both the categories allowed users to create self-defined goals and self-defined rewards. The *Play together* category offers certain social features that allows users to invite friends, compare their rewards, work on tasks together, or assign tasks to each other. For every task that a user creates, they must select atleast one goal and atleast one reward, and for those selected game elements, they can configure further depending on the configuration options that are available for the corresponding element.

In order to investigate this over different degrees of customization, the authors chose a crowd-sourcing setting, where the participants were requested to perform as many microtasks as possible, in the domain of receipt capturing. Figure 2.5 depicts the game element selection screen. Here, the users decide themselves, on what, when and how the game elements will appear for each of the to-do tasks that they create themselves on the task managing application.

The prototype was tested in a twelve-day-long user study allowing bottom-up gamification tasks such as adding users daily tasks and selecting appropriate game elements for rewarding or motivating themselves. Figure 2.5 depicts the main screen containing daily tasks such as *do the dishes*, *run*, etc,. Commonly-used game elements (see Figure 2.5) such as receiving points, achievements, self-defined rewards, progression, goals, feedback, social recognition and competition through leaderboards were identified and some of them were used during the prototype testing. The study suggests that users are interested to spend more effort and thought into customizing their own gamified environment.

2.1.4 Wrap-Up

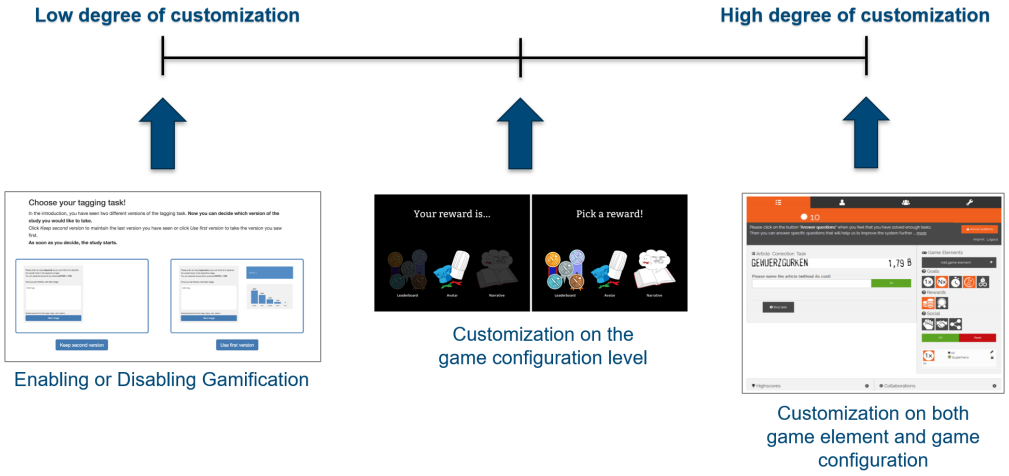


Figure 2.6: Forms of customization (Left: Lowest degree of customization (from an *Image Tagging* Platform); Middle: Moderate degree of customization (from *Cafe Flour Sack* game); Right: High degree of customization (from *BU-ToDo* app)).

All the above-mentioned studies (see Figure 2.6) suggest that providing users with the ability to choose from the offered customization is crucial and motivating for users', thereby pushing them towards performing better in the tasks, in different contexts and situations. However, the success of different degrees of customization depends on the type of task and on whether the user is willing to invest time and effort to setup a gamified environment.

By drawing outcomes from different spectrums of customization, it allows us to situate our work under the lowest degree of customization option. Our thesis aims to investigate on a single, elementary customization option: enable gamification or disable gamification (experiment 1), and choosing between two fixed gamified interventions (experiment 2). This is done with the motive to minimize the user effort in designing the application environment, yet giving the advantage to choose their preferred gamified environment.

2.2 Choice and Autonomy

Given the positive effects of freedom in choosing one's own customization, an increased user effort might lead to analysis paralysis[2]. Current research [10, 16] tests the phenomena of choice overload by creating tasks that requires decisions involving multiple options (example 3 options vs. 10 options) [10]. Haynes [10] conducted an experiment to assess the effects of the number of alternatives provided to the users. For this purpose, the author designed a study that presents participants with descriptions of prizes. The participant has to choose one prize from the number of alternatives given. Each participant were presented with different sets of choice - a moderate choice set containing 3 options versus a larger choice set containing 10 options. The participants were allowed to make their decision within a certain amount of time, say 2 minutes versus an extended decision time of 5 minutes. The results of the study reveals that participants tend to enjoy the task more when presented with more number of choices, however the particiants felt the decision making process to be very difficult and frustrating compared to the participants who had a fewer set of options to choose from. This eventually led participants with more number of options to be less satisfied with their final decision than participants with a limited option set. Similarly, Iyengar & Lepper [16] conducted three studies in different contexts: jam selection, essay topic selection, and chocalate selection. For each of the study, participants were either presented with a limited-choice of 6 options, or an extensive-choice of 24 options, without having any time constraint (i.e. the participants can take their own time to choose from the options presented). The authors concluded that despite the higher initial enjoyment in the extensive-choice set, participants proved more dissatisfied and regretful of their choices. Moreover, when participants were presented with a limited-choice set, they were more likely to buy the product (in jam selection and chocalate selection studies), and were more likely to produce quality work (in essay writing). As the number of options for the users to modify increases, so does the level of complexity of the decision itself. This behavior will eventually lead to frustration, disssatisfaction, and disinterest in the task. However, Schöbel et al. [33] posits, that people feel most confident in their decisions when they understand the available options and can comfortably compare and evaluate them. In other terms, providing autonomy and freedom in choosing according to the user's preferences with less user effort is merely not enough, but important.

2.2.1 Wrap-Up

From the above-mentioned studies, we derive that people who are given with the freedom of choice (with minimum number of options) are more intrinsically motivated to perform better in the tasks given, and feel more satisfied about their decision. This shows that people feel inherently more intrinsically motivated, as their need for autonomy is satisfied. In order to investigate this phenomena of choice overload, we in this thesis, provide simplest and minimum viable choice set to users, and investigate the effectiveness of providing choice in the context of gamified online surveys.

[2]The feeling of dissatisfaction when users are provided with an abundance of choices. In other words, called as paradox of choice.

2.3 Gamification in Online Surveys

In our thesis, we would like to focus on investigating the effect of gamification in online surveys. Online surveys have been one of the target areas for gamification efforts [38] and has been predominantly used in a variety of domains (Sports and Leisure Activities [15, 14]; Recycling [19]; Image tagging [22, 26]; Market Research [6]; Manufacturing [20]; Gastronomy [29]). Surveys are considered to be one of the most important tools to make inferences about an entire population, people's attitude, perceptions, intents, habits, awareness, experiences, and characteristics [27]. Granello & Wheaton [11] explains the main problem of this quantitative tool, to be about maintaining users engagement (response rate) and the validity of the data (data quality). As a measure to improve these factors gamification was introduced in online surveys. Gamification, in general, is defined as the application of game elements to non-game contexts [7]. Deterding et al. [8] formulates the term 'gamification' as the use of game design elements, as a means to improve the user experience and user engagement in non-game services and applications. Incorporating gamification in surveys has been considered to be an effective method for improving user motivation, performance and satisfaction [11, 27]. Adding to it, Deutskens [9] highlights the fact that offering game elements such as prizes has proven to be beneficial in achieveing higher survey response rates. Following this, we briefly look at the work of several researchers and how they proposed to add game elements to online surveys to increase user's motivation, participation, perceived fun, engagement, etc,.

2.3.1 Gamified Online Surveys: Sports Domain

Harms et al. [15] investigated the effect of gamification on online surveys through a sports and leisure activities survey. The authors compared a non-gamified survey with a gamified survey to analyze whether gamification had any impact on the participant. The gamified survey consisted of a single game element namely badges (cf. Figure 2.7), which will be unlocked by the participant based on certain pre-defined rules. Figure 2.7 depicts 10 badges containing name and description of the achievement, which can be obtained by the participants while answering the surveys.

The survey was conducted as a between-subject study design, with a total number of 126 participants, distributed to gamified (N = 66) and non-gamified (N = 60) respectively. Results were calculated based on the affect scores (I-PANAS-SF), user experience (AttrakDiff2), subjective ratings (Likert-type questions), respondent behavior (completion rate, time spent on the survey, number of words in a free-text question, speeding, strightlining, empty answers.), and on qualitative feedback (positive and negative comments). There were significant differences in the scores for hedonic quality - simulation and attractiveness between the gamified (higher and better) and non-gamified surveys. Based on the subjective ratings, participants tend to prefer the gamified survey higher than the non-gamified survey. The qualitative feedback also suggested that participants perceived the gamified survey with achievement badges to be more interesting and fun, with a majority of comments (22 positive and 4 negative comments) on the achievement badges. When we have a closer look at the evaluation results, it reveals an improved psychological outcome (better user experience, higher preference, positive qualitative feedback), but insignificant differences in behavioral

Figure 2.7: Achievement badges designed for the gamified survey

changes. In a nutshell, gamification in surveys is perceived by users in a positive manner by experiencing more fun, higher willingness to use the survey, and willingness to recommend the survey, thereby suggesting that game elements act as a means to improve user experience.

As an extended version of the above research work [15], Harms et al. [14] designed a gamified survey containing many elements and compared it with a non-gamified survey in the context of sports and leisure activities. The gamified survey consisted of game elements such as avatar creation (avatars visual appearance based on demographics data provided by the users), free exploration (navigate through the four sports disciplines within the survey), questions answered through microgames (cf. Figure 2.10 soccer, javelin throwing cf. Figure 2.11, long jump, and sprint), and feedbacks were provided to the users by awarding coins (see left corner of Figure 2.10), customizations for their avatar (cf. Figure 2.8), medal ceremony (cf. Figure 2.9). The introduction of these game elements resulted in a highly game-like appearance for the gamified survey. Similar to the previously discussed study, this study also compared the gamified survey with a non-gamified one with a total of 60 participants, equally distributed across the two surveys. According to the self-rated user experience, there are significant differences in users perceived fun, and more participants prefer gamified surveys over non-gamified survey, and are most likely to recommend the survey. Qualitative results shows that respondents positively commented on the novelty, variety, and interactivity of the gamified survey, and in addition to that they found the gamified survey to be playful and fun. Notably, the graphics and customizable avatar received positive comments, and intrigued the users interest in such a way that the users were willing to play even after finishing the survey. On the contrary, participants complained that the gamified survey took much longer to answer compared to the non-gamified survey. On the flipside, participants of the non-gamified survey found the survey to be boring, but that it was easy to use and easy to answer. Apart from the psychological outcomes of the gamified survey, behavioral outcomes such as participants response rate[3], time spent on the survey, and amount of plain-text answers did not find any significant differences. The authors explains the lower response rate to be a reflection of polarized reaction towards

[3]Response rate, also known as completion rate, is the number of people who answered the survey divied by the number of people in the sample.

Figure 2.8: Shop to spend the rewarded coins

Figure 2.9: medal ceremony as a thank-you page

Figure 2.10: Microgame soccer for Single-choice questions

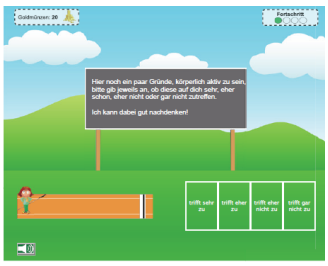


Figure 2.11: Microgame javelin throwing for Likert questions

the questionnaire, as some users felt uncomfortable (in terms of duration, complexity, controls, responsiveness of individual games) with the design of the gamified survey.

Both the studies conducted by Harms et al. [14, 15] indicates that the low-cost gamified survey with a single game element, as well as the gamified survey with multiple game elements, both shows same psychological and behavioral outcomes. Despite the high effort invested in designing survey with multiple game elements, similar primary outcomes were obtained by the survey consisting of a single game element. The authors describe this method as a good Return-on-Investment (ROI)[4], with the key performance indicators (KPIs) as the cost or effort invested in gamification of the survey. The study proves that using a single game element is well-suited for low-cost method for survey gamification. This study motivated us to use a single game element for our thesis, and additionally driving us to investigate on the effectiveness of individual game elements within the context of surveys. As their study proved that providing badges are beneficial, we would like to further examine whether providing users with a choice to customize their environment with a single click of a button, and decide whether they would like to answer the survey in a gamified setup or a non-gamified setup. As a result, we would like to situate our work in the domain of low-degree customization with low-cost gamification.

[4]ROI is a performance measure used to evaluate the efficiency of an investment or comparison between the efficiency of a number of different investments.

2.3.2 Gamified Online Surveys: Personality Assessment Domain

Triantoro et al. [38] investigates the effect of gamification in online surveys by using the stimulus-organism-response (S-O-R) framework[5] and signaling theory[6]. The authors designed two versions of the survey - a traditional survey with Likert scales, and a gamified survey with game mechanics such as personalization, time constraint, and scoring mechanism. Each of these game elements were operationalized with the respective game elements and purposes:

1. Avatar - An avatar can be personalized by choosing the gender, hairstyle, face features, and complexion.
2. Timer - A timer appeared when participants were answering a set of aptitude questions.
3. Points - Points were assigned based on the number of correctly answered questions.

Both survey versions intended to measure users' personality traits, and therefore were designed with the help of Big Five personality assessment instrument. The Big Five trait taxonomy serves as a means to capture data, based on the personality traits: agreeableness, conscientiousness, extraversion, neuroticism and openness to experience. Additionally, this study also evaluated the effect of gamification on user's cognitive and affective reactions such as enjoyment and attention. From a total of 694 participants, 392 participants were assigned to the traditional survey, and 302 participants were assigned to the gamified survey. Results show that the extrinsic mechanisms such as rewards (points), does not affect the cognitive reactions (i.e. attention towards the survey), but it does affect the affective reactions (i.e. enjoyment associated with the survey). At the same time, the effect of intrinsic mechanisms, such as constraints (timer), is highly influential on the cognitive and affective reactions. With respect to self-presentation mechanisms (such as avatar), there exists a significant effect on both cognitive and affective reactions. Moreover, the authors argue that allowing the players to control their own presentation creates opportunities to become more invested in the game environment.

2.3.3 Gamified Online Surveys: Gastronomy Domain

Prott et al. [29] decided to investigate the effectiveness of gamification in online surveys within the context of gastronomy and catering industry[7]. For the empirical study, the authors created two questionnaire versions, a classical (cf. Figure 2.12) and a gamified questionnaire (cf. Figure 2.13).

Figure 2.12 depicts the start page of the classical questionnaire, consisting of a simple design, with plain colors, sharp buttons and zero animations. In contrast, Figure 2.13 shows the gamified version of the questionnaire, consisting of additional animations, vibrant with colors, several illustrations and rounded edges for the buttons. The seven different game elements incorporated in the gamified questionnaire were:

[5]The S-O-R framework allows articulating the relation between the stimuli and different types of organismic states.

[6]Signaling theory allows to characterize the nature of the stimulus when it is presented in gamified systems

[7]https://itell.solutions (last accessed on 15 June 2020)

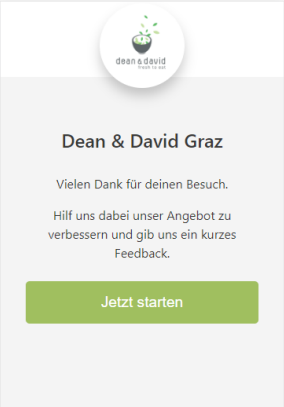


Figure 2.12: Start page of the non-gamified questionnaire. (Translation of picture text: Thank you for your visit. Help us improve our offer and give us feedback.)

Figure 2.13: Start page – Gamified questionnaire. (Translation of picture text: You are one of the few chosen to fill out the gamified questionnaire. Collect skill points, badges and much more. Take your chance!!!).

1. Narrative - Narrative, in the form of a story gives additional meaning to the activity being performed and it has been scientifically proven that the human brain processes information much more easily, when it is embedded in a context [35].
2. Avatar - Avatars are visual representations of players within the gamified environment. Usually the avatars are created by the player to adopt or create a unique identity [30].
3. Points - Points are rewarded for successful accomplishment of specified activities within the gamified environment, and they serve to numerically represent a player's progress [30].
4. Badges - Badges are visual representations of certain achievements. Earning a badge can be dependent on a specific amount of points or on particular activities within the game. Badges has shown to exert social influences within players if they are rare and hard to earn [30].
5. Progress bar - Progress bar is a good indicator for estimating the time it takes to complete a certain task in an application [1].
6. Instant feedback - Instant feedback is the immediate feedback of the system during or after an activity by users. Users get information on how they are progressing in the system and can thus learn very quickly how to improve themselves in a gamified process [1].
7. Glowing choice - Glowing choice helps users to move forward in an activity that uses visual representations to refer to the next action. For example, glowing choice for pre-generated text responses in a free-text answer field acts as an incentive to deliver a text response or a longer text response [29].

Figure 2.14: Final Page with feedback on progress. (Your today's achievement. You have answered 8 out of 9 questions. You have collected 40 gold talers. You needed 97 seconds to complete the questionnaire. Your current progress. The king has received 2 feedbacks from you. Your treasure chest contains 45 gold talers. Your effort earned 2 medals.); translation of picture text.

At the end of the survey, participants were provided with an overall achievement list in the form of a narrative (cf. Figure 2.14).

The authors reported the main findings of the research to be a significant increase in free-text responses and an increase in the overall survey completion time. Results show that there was a tendency for more longer responses to the free-text questions with an average number of characters used being 15.23 for the classical and 44.53 for the gamified questionnaire. However, the authors argue that the increase in completion time could be the result of using multiple game elements, especially narratives. Also, it is unclear about which game elements from the seven game elements used, motivated participants during the gamified version. Therefore, in our thesis we would like to investigate each game element seperately as it is evident from this study that identifying the game elements that triggers positive user behaviors could help us to evaluate the effect of the particular game element and also understand whether the context of the survey is also a determining factor for the choice of game elements.

2.3.4 Gamified Online Surveys: Market Research Domain

Keusch et al. [17] conducted a meta-analysis on 14 studies in a variety of contexts, that experimentally compared gamified surveys with a non-gamified one. The most often used game mechanics or motivational affordances in the chosen papers were challenges (9 studies), specific goals/objectives or rules (7 studies), back stories, themes or narratives (7 studies), and rewards (6 studies). The authors provided a review of the overall findings regarding the impact of gamification on psychological outcomes, validity, and behavioral outcomes related to non-response and measurement error. Authors reported that the

psychological outcomes were positive in the gamified surveys. Out of the 14 studies, 2 studies reported that users found the gamified survey to be more interesting, 3 studies reported more fun, 5 studies reported that the users enjoyed the gamified version more. Only a few studies (4 studies) reported mixed opinions on the user behavior, for example, users exerted extreme reactions (both positively and negatively) to open-ended questions. The authors argue that the behavioral outcomes such as non-response error and careless responses varied with experiments, as each used different game elements in their study. 7 studies included narratives, stories, themes, out of which only 3 studies showed fewer break-off rates, and other 4 studies found no significant differences. The author identifies the reason behind this break-off rate to be a possible disconnect with the survey topic and the narratives used. They also identify an increased completion time in gamified surveys (6 studies), longer answers to open-ended questions (3 studies), more straightlining in traditional surveys than in gamified surveys (3 studies). The authors argue whether the differences in answers were a result of gamification or merely because of the manipulation of the question format (i.e. changing the wording of the questions) and questions layout (in the form of drag and drop or sliding bar), which is not necessarily related to gamification.

While most of the studies from the review suggests a positive effect of using game elements on psychological outcomes such as fun, interest, enjoyment, there are conflicting data when it comes to behavioral outcomes such as completion/break off rates, open-ended questions, straightlining. Inorder to study the impact of particular game elements on behavioral outcomes, it is neceesary to evaluate each element individually. Our thesis aims to choose single game element for each survey, and investigate which game element fits better for the chosen context.

An example showing that users not only enjoy the gamified surveys, but also outperforms in providing quality data can be seen in the study conducted by Mavletova [**?**]. The study was conducted among children and adolescents between the age group of 7 - 15 years in the context of market research. The survey comprises of three versions: text-only survey with plain text-based questions, a visual survey which comprises of animations, and a gamified survey consisting of game elements such as narrative story, rules, challanges and rewards. The four game elements used in the survey were meant to fulfil the following purposes:

1. Narrative - A screenplay of the gamified survey which motivates respondents to achieve goals.
2. Rules and goals - Predefined set of rules and goals are displayed for the respondents for identifying what should be achieved and how it can be done.
3. Challenges - Challenging tasks or quests which require some elements of skill and effort to maintain the respondents interest.
4. Rewards - As a driver of engagement among respondents.

The author identifies the following positive effects of using gamification in online surveys:

- The results show that gamified survey outperforms the other versions with respect to careless responses, especially there is a lower level of straightlining[8] answers

[8]Straightlining is when the users select the same options in a likert scale; Middle-point responses refers to the neutral responses such as 'I don't know' answers, 'no opinion' answers.

and lower level of middle-point[7] response styles. This concludes that the gamified survey produces lower degrees of careless responses.

- Users requested help less often, and reported that the questionnaire was easier and more enjoyable especially for younger children.

In conclusion, the findings of the study provides evidence for a positive gamification effect in the surveys among children. However, one limitation is that since four game elements (narratives, rules, challenges, and rewards) were used in the study, the authors were unsure about which particular game element rendered the positive effects in data quality.

A study by Cechanowicz et al. [6] investigates the effects of gamification on participation and data quaity in the context of market research domain. Three versions of the survey were designed, each involving different levels of gamification (cf. Figure 2.15. A plain survey, a partially gamified version with interactive game mechanics, and a fully-gamified version with game elements such as theme, reward, challenge, and feedback) were designed to conduct a comparative study between the versions. The study involved tasks in three areas of branding and advertising namely: Image identification task, slogan matching, and five-second quiz. Every task used different game elements that paired well with the task. The following list of game elements were used:

1. Theme - Gamified survey took the theme of a trivia game show, comprising of virtual cash prizes, cheering sound effects, a market trivia banner and an introductory message that read 'Welcome to the Game show'.
2. Reward - For every completed question with the correct response, points were rewarded.
3. Challenge - Challenge in the form of a timer was added, which directly affected the number of points earned on a correct response. i.e. It affects the points - 100% if the timer has expired, 200% if more than half of the timer had expired, 300% if less than half of the timer had expired, and so on.
4. Feedback - For every correct response, points are revealed in the form of a visual animation, and for every wrong response, the correct answers are displayed to the participants. In addition, participants were told how many keywords they used and the total number of correct keywords they could have used which serves as a driving factor to encourage participants for longer responses.

The study which involved 600 participants, showed that gamification approach towards survey, led to significantly higher levels of participation (i.e. people completed significantly more questions in the gamified survey), and that the motivation increases with the addition game elements(cf. Figure 2.15). In addition, the study also revealed that the positive effects of gamification remained consistent across age, gender, and prior game experience, indicating that gamification can improve motivation across all demographic factors.

The authors argue that the use of different game elements could also impact the survey responses. For example, the addition of timer in the gamified survey made participants

[7] Middle-point responses are neither-nor responses which adds neutral value to survey evaluation.

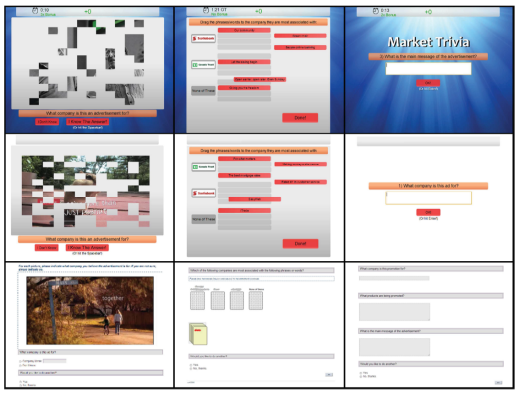


Figure 2.15: Full Game (top row), Partial Game (middle), and Survey (bottom) versions of the Image Identification (left column),Slogan Matching (centre), and Five-Second Quiz (right) question types. Source: [6]

to rush towards the completion, which inturn resulted in shorter responses in free-text questions. This supposedly indicates that a wrong choice of game element will result in a lack of response quality and quantity.

2.3.5 Wrap-Up

As we can see from the above discussed papers, a number of studies has shown that adding game design elements in surveys enhances the response rate and involvement among users [6, 15, 29]. Triantoro et al. [38] also presents three important mechanisms while including gamification to surveys, as extrinsic (rewards), intrinsic (constraints), and self-presentation mechanisms. The findings also highlights the fact that allowing users to control their own presentation creates opportunities to adapt to the game environment. In contrast, there are still a few studies identified by Keusch et al. [17], that gamifying surveys have no (not always beneficial) significant influence on users psychological and behavioral outcomes, such as user experience, motivation, engagement, participation, satisfaction, enjoyment and the amount and quality of data. Despite these contradicting results, Hamari et al. [13] suggests that the effects of gamification were strongly influenced by users and context. Therefore, while designing the gamified surveys, it is essential to consider the game design elements that are suitable for the context. Another study that supports this claim is [6], where the choice of game element (example: timer) lowered the quality of the respondent's data. Often, there also seems to be a disconnect between the story of the game and the game element used. An important point [14] to note is to check whether a single game element is enough in the chosen context, as it helps to save the time taken to design the survey as well as the cost of redesigning or reusing the game elements.

From the above discussion, we define our thesis to be allowing users to control their presentation of the surveys, thereby creating an opportunity to the user to choose from the presented gamified survey versions, with a minimum possible effort i.e. single-click of a button. Taking it further from here, we would like to investigate the impact of individual game elements on user's behavioral and psychological outcomes.

2.4 Variables Measured in Gamified Online Surveys

From the related work, we identified a few variables that measures certain aspects determining the quality and the quantity of the gathered survey data. Table 2.1 show a list of papers that investigates the effectiveness of gamification in online surveys in a variety of domains. Most of the studies [14, 15, 29, 38, 39] evaluated the effectiveness of providing gamification in online surveys, by comparing a gamified online survey with a non-gamified online survey. A few other studies [6, 12, 23, 24, 31] evaluated the effectiveness of gamification, by comparing three different survey versions i.e. a non-gamified survey, a decoratively visual survey containing animations and visually appealing user interface design (including graphics, animation, color, background) and a gamified survey containing game design elements and certain game mechanics.

Article	Domain	N	Game elements used	Survey Versions
1	Market research	1007	Narratives, avatar, rewards	Non-gamified, decoratively visual, functionally visual, gamified
2	Branding and advertising	644	Narratives, reward, timer, challenge, progress elements	Non-gamified, decoratively visual, fully gamified
3	Sports	139	Badges	Gamified, non-gamified
4		60	Narratives, avatar, points, medals	Gamified, non-gamified
5	Children and adoloscents	1050	Narratives, rules, challenges, rewards	Non-gamified, decoratively visual, gamified
6		737	Narratives, rules, challenges, rewards	Non-gamified, decoratively visual, gamified
7	Meta-analysis	-	-	-
8	Psychology	340	Narratives, badges	Non-gamified, decoratively visual, gamified
9	Personality assessment	694	Avatar, points, timer	Gamified, non-gamified
10	Gastronomy	150	Narratives, points, badges, avatar, progress elements	Gamified, non-gamified
11	Personality assessment	694	Narratives, avatar, challenge, timer, points	Gamified, non-gamified
12	Market Research	397	Feedback, points, progress, time pressure, increasing challenge (levels), ranking (leaderboard), story, avatar	Non-gamified, Gamified version 1 (No correct answers), Gamified version 2 (Humans with correct answers), Gamified version 3 (Monsters with correct answers)
13	Market Research	709	Rules, miniquests, feedback, story, competitive elements	Non-gamified, Gamified
14	Market Research	220	Rules, timer, visuals, game-like question wording	Non-gamified, Gamified

Table 2.1: A list of papers that conducted experiments with gamified online surveys in different domains.

After conducting a literature research on 'gamification in online surveys' (see Figure 2.16), we derived at the dependent variables that could be measured. Hence, we have broken down the variables into the following two categories: behavioral and psychological. Behavioral variables are those that measures the task performance and the overall quality of the survey data. Psychological variables are those that measures the subjective experience of the participants in terms of enjoyment, user experience, and usability.

2.4.1 Behavioral Variables

Table 2.2 show variables that were used in previous studies, to measure the performance of a user during the survey session. The variables that had same meaning were grouped together into a single dependent variable. If a particular variable (eg: dropout rate) measures the inverse of the actual dependent variable (i.e. survey completion rate), we

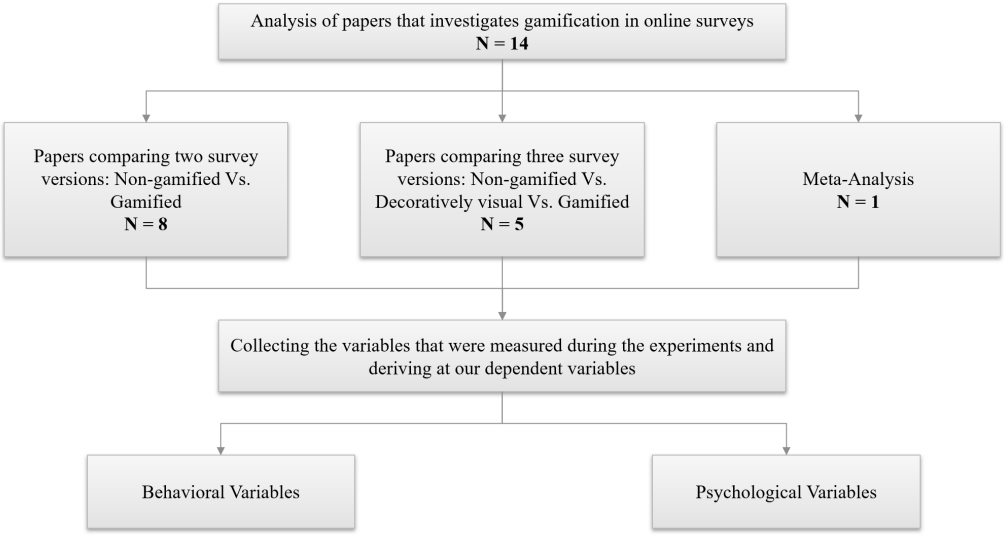


Figure 2.16: A diagramatic representation of the literature analysis process for deriving at our dependent variables.

coded it negatively to ensure for the correctness of the grouping (see Table 2.2).

1. **Survey completion rate:** can be identified by measuring the number of people who entered and successfully completed the survey.
2. **Item non-response rate:** can be identified by measuring the number of optional questions not answered.
3. **Words in free-text answers:** can be measured by counting the number of characters used in open questions. In our survey questionnaire, questions: 10, 11, and 14 are free-text questions.
4. **Degree of careless responses:** can be identified by keeping track of the following items:
    a **Straightlining:** is considered to be a careless response or a negative behavior, and is identified when the participant answers a block of questions in such a way that visually a straight line is created. Therefore, this behavior can be measured by identifying consecutive blocks of answers.
    b **Speeding:** is identified when a participant answers the survey questions as quickly as possible, by randomly clicking the answers, without reading the text carefully, or without thinking sufficiently for an appropriate answer. Such a behavior can be measured by counting the time spent on each page. A threshold value was set for every survey page by counting all the words in a question and the possible answers and by multipling it with 200ms [15].

Behavioral variables measured in different studies	Grouping of Behavioral variables
Survey completion rate Response rate Return rate	Survey completion rate
Premature termination Dropout rate	- Survey completion rate
Item non-response rate	Item non-response rate
Survey participation Number of optional questions answered	- Item non-response rate
Number of characters/words in free-text answers	Characters/words in free-text answers
Straightlining	Straightlining
Speeding	Speeding
Conflicting answers	Conflicting answers
Middle point responses	Middle point responses
Don't know answers	Don't know answers
Duration Survey completion time Total time spent on the survey	Survey completion time

Table 2.2: A list of behavioral variables collected from different studies. Left: Behavioral variables. Right: Grouping of the behavioral variables that holds similar meaning and interpretation. (A negation (-) denotes that the variable represent contrary meaning to that of the chosen variable.)

c **Conflicting answers:** measured by comparing similar questions. In our survey questionnaire, we do not include questions to measure conflicting answers. Hence, this variable is excluded from our study.

d **Middle responses:** comparing whether there was a tendancy for choosing middle category i.e. neutral answers.

e **Don't know answers:** comparing whether there was a tendancy for choosing don't know answers i.e. in our study, the 'no answer' option.

5. **Survey completion time:** can be measured by calculating the time taken to complete the entire survey.

Table 2.3 show a list of papers that obtained significant measures during their study. As we can see, few papers have reported positive outcomes such as reduction in the amount of straightlined-answers [24], reduction in middle-point responses [24], and an increase in survey completion rate (this is considered to be a positive outcome, as few authors [14, 29] suggest that an increase in time spent in answering the survey, is not necessarily a sign of heightened engagement, but the fact that participants found the survey to be more fun and voluntarily spent more time suggests that the increased duration is the result of an improved user experience). There are also a few papers that have reported positive outcomes such as an increase in the completion rate of the survey [15], lesser amount of item non-responses [23, 31], and an increase in the number of words used in the free-text fields [14, 15, 24, 29]. However, these studies were not able to find a significant difference between the different versions of their survey.

No.	DV	#Papers using DV	#Papers reporting SIG outcomes		#Papers reporting non-significant outcomes		#Papers which did not conduct significance tests
			Increase in DV	Decrease in DV	Increase in DV	Decrease in DV	
Behavioral Outcomes							
1	Survey completion rate	**9 |** 1, 3, 4, 5, 6, 8, 10, 12, 13	-	1, 6, 13	3, 12	5	4, 10
2	Item non-response rate	**5 |** 2, 4, 5, 6, 8	2, 5	-	4	6	8
3	Words in free-text answers	**7 |** 2, 3, 4, 5, 10, 13, 14	13, 14	2	3, 4, 5	-	10
4	Degree of careless responses						
a)	Straightlining	**5 |** 1, 3, 5, 6, 8	-	5	-	1, 3, 6	8
b)	Speeding	**2 |** 3, 8	-	-	3	-	8
c)	Middle-point responses	**1 |** 5	-	5	-	-	-
d)	Don't know answers	**1 |** 3	-	-	3	-	-
5	Survey completion time	**9 |** 1, 3, 4, 5, 6, 10, 12, 13, 14	4, 6, 13	-	1, 3, 5	12, 14	10
6	Number of choice sets completed	**1 |** 12				12	

Table 2.3: A list containing the statistical significance measurements of the different behavioral variables.

2.4.2 Psychological Variables

Psychological variables are used to measure the subjective experience of the participants after using the survey platform. Table 2.4 show a list of variables that were measured in previous studies. Similar to the grouping of the behavioral variables, we grouped together the psychological variables that had a similar meaning but differed in their names.

1. **User Experience:** In the study that we replicate, user experience is evaluated by measuring pragmatic and hedonic qualities. i.e. the pragmatic qualities refers to the perceived usefulness, efficiency, and ease-of-use of the survey platform, and hedonic qualities refers to the joy of use and particularly measures stimulation, attractiveness, and identification aspects that are generated during the use of the system. Therefore, in our study, we measure these qualities by using a standardized questionnaire named **AttrakDiff2**.

2. **Affect:** is used to understand how a users feels before and after the survey. i.e. in our case, to understand whether the moods of a user has any correlation to the perceived effectiveness of gamification. A standardized questionnaire namely **I-PANAS-SF** can be used to measured the positive and negative effects and understand the emotional states of the user.

3. **Perceived Fun:** can be additionally measured through open questions allowing us to gather more insights into the user's enjoyment.

4. **Perceived Duration:** is used to understand how the users perceived the time spent in filling out the survey. i.e. how long they felt the survey took.

5. **Preference:** can be measured by knowing the willingness of a user to use the gamified survey and/or recommend the survey to others.

6. **Perceived Usability:** measures how satisfied a user is after using the system. Therefore, this variable can be measured using the standarized questionnaire

Psychological variables measured in different studies	Grouping of psychological variables
Pragmatic qualities: Perceived usefulness, efficiency, ease-of-use Hedonic qualities: Joy of use, stimulation, identification, attractiveness	User Experience
Positive and negative affect, emotional state	Affect
Enjoyment	Perceived Fun
Duration	Perceived Duration
Preference	Preference
Satisfaction, confidence in using	Perceived Usability
Ease-of-use, easy to read, easy to answer	Perceived Effort
Attention	Attention

Table 2.4: A list of psychological variables measured in different studies. Left: Psychological variables. Right: Grouping of the psychological variables that holds similar meaning and interpretation. (A negation (-) denotes that the variable represent contrary meaning to that of the chosen variable.)

No.	DV	#Papers using DV	#Papers reporting SIG outcomes		#Papers reporting non-significant outcomes	#Papers which did not conduct significance tests
			Increase in DV	Decrease in DV		
Psychological Outcomes						
1	User experience	**1** | 3	3	-	-	-
2	Affect	**1** | 3	-	-	3	-
3	Perceived fun	**8** | 1, 3, 4, 5, 6, 9, 12, 13	1, 4, 5, 9, 13	-	3, 6, 12	-
4	Perceived duration	**1** | 1, 3	-	1	3	-
5	Preference	**5** | 3, 4, 9, 10, 12	3, 4, 9	12	-	10
6	Satisfaction	**3** | 1, 10, 13	1	-	13	10
7	Perceived effort	**3** | 1, 5, 6	1	5, 6	-	-
8	Perceived usability	**2** | 4, 13	-	4, 13	-	-
9	Attention	**3**| 9, 12, 13	9, 12	-	13	-

Table 2.5: A list containing the statistical significance measurements of the different psychological variables.

System Usability Scale (SUS) to understand the usability aspects and to understand how the user felt (eg: confident) while using the survey.

7. **Perceived User Effort:** this variables measures how much effort is needed from the user inorder to use/understand the system.
8. **Attention:** is used to understand how gamification has affected a users attention and cognitive state. In previous studies [38, 39], attention was measured using signaling theory and Stimulus-Organism-Response (S-O-R) framework.

Table 2.5 show a list of papers that obtained significant measures during their study. As we can see, most of the papers report positive outcomes such as increased user experience [15], an improvement in the perceived fun [12, 14, 24, 38, 39], a sense of feeling that allows users to think that they spent less time in the survey [12], and an improvement in satisfaction [12, 29], preference [14, 15, 29, 38], and overall attention towards the survey [38, 39].

2.4.3 Wrap-Up

Since our gamified survey platform is deployed online, we will not be able to observe the participants directly (in-person) to understand their overall experience of using the survey platform. Therefore, we measure behavioral variables based on the measurements collected such as number of participants took part in the survey, number of optional questions answered by the participants, number of characters used in free-text fields, the amount of time taken to complete the survey. In order to measue participants affective and psychological state, we include certain questions that evaluates the subjective experience such as perceived enjoyment, perceived duration, and overall experience during the survey session.

Chapter 3
Study Platform

In this chapter, we discuss the concept of the platform that was implemented to conduct both our user studies on (Experiment 1 and Experiment 2).

3.1 Gamified Survey Platform

The overall design of our experiment is similar to the study conducted by Harms et al. [15]. However, we modify a few aspects to keep the study design/outcomes consistent across the different conditions. In this section, we describe in detail the aspects we kept, the aspects we changed, and the how we intend to conduct the overall study.

We replicated the folowing aspects:

- The survey platform was available in German.
- The overall user interface design was kept consistent across the survey versions: non-gamified survey and gamified survey.
- The survey questions were reused for our experiment.
- The non-gamified survey comprised of a game element progress bar, which according to the authors were not particularly considered as a game element. Hence, we also kept progress bar in the non-gamified survey version.
- The gamified survey comprised of game element badges and progress bar. Again, the progress bar was not considered as a game element by the authors. Hence, we mimicked the design as such.
- The rules for awarding the badges in the gamified survey was kept. However, the semantics of one badge out of the ten badges was modified. Below, we discuss the aspects that led us to change the semantics of the badge.
- The goal of the study was kept undisclosed to the participants.

3.2 Additional Features

For the purpose of our study, we extended a few aspects from the above study:

- The survey platform was originally designed in German language. However, we designed the survey platform in English. The translations from German to English were made as accurately as possible, using the translator software DeepL[1].
- In the gamified survey, the authors kept gamification to be active in both the main survey and post survey. We keep gamification active only during the main survey and not during the post-survey, as the questions in the post-survey requires the user to reason about the experienced gamification and keeping gamification active during this part might be confusing for the users and might confound the results.
- The rules for awarding badges in the gamified survey is kept. However, the semantics of one badge out of the ten badges is modified. i.e. in the actual study, the badge 'Interview' can be unlocked by answering the post-survey questions. However, since we deactivate gamification in the post-survey session, we will be adding certain demographic questions at the end of the main survey, and award the badge 'Interview' for answering them. Similarly, we transferred the same rules and logic for the other gamified survey version, namely points and leaderboard.
- The ordering of the questions were changed. i.e. the pre-survey questions which requested users about their demographic data were adapted at the last page in order to match the semantics of Interview badge.
- A guided tour was added for both non-gamified and gamified survey with the notion of explaining the different game elements such as progress bar and badges that are present, and additionally allow them to experience and understand how these elements can be unlocked or used. The guided tour also includes steps for successfully completing the survey, and the possibility to quit the survey whenever needed.
- We added an additional condition (for experiment 1) where participants had a choice to either enable or disable gamification. Participants entering this condition will be redirected to a choice page after completing the pre-survey and they will experience a guided tour of the two survey versions. Later, they will be redirected to the choice page, where they can choose if they want to answer the survey with or without gamification through a single click of a button. A detailed explanation of the conditions of the two individual experiments are discussed in later chapters (see User Study).
- We use a round-robin method to allocate participants to the different study conditions. i.e. we distribute the participants in a 1:1:2 manner based on the completed surveys, in order to make the sample distribution equal across the conditions.

3.2.1 Choice

Since we wanted to provide users with a choice to select the survey platform they wanted, we took two phenomena into consideration while designing the customization options.

[1] https://www.deepl.com/en/translator

Firstly, we wanted to avoid choice overload, and secondly we wanted users to use their preferred gamified platform with minimum effort. In other words, we wanted to offer a choice which is not complex and does not need high user effort, in terms of setting up the gamified platform. This inspiration was took from Lessel et al. [22], where the authors allowed users to select their preferred image tagging platform (gamified platform or a non-gamified platform) using a single click of a button. That is, the users can select between a non-gamified platform versus a gamified platform, simply by clicking the enable or disable gamification option.

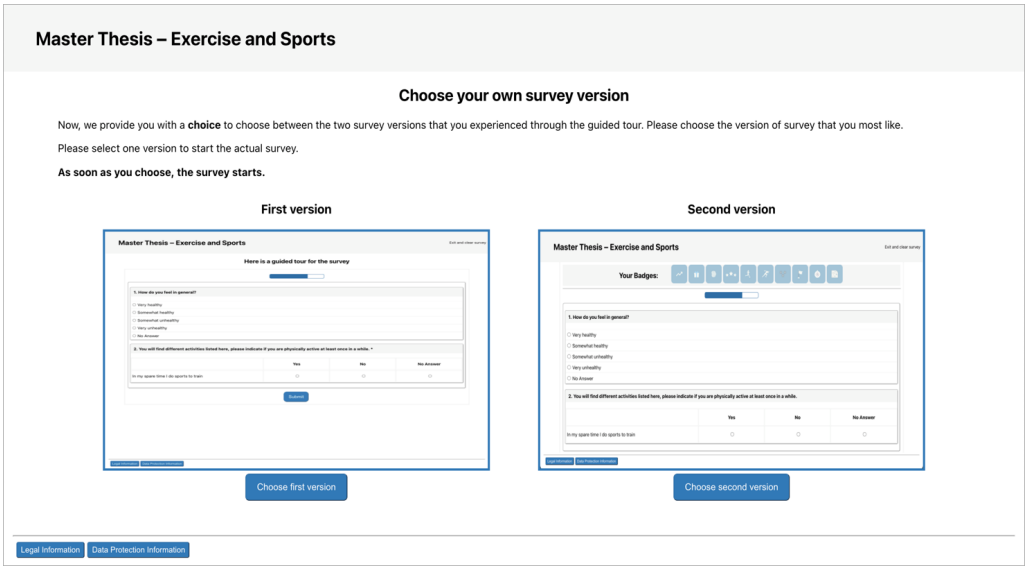


Figure 3.1: Screenshot of the page where users could make their choice of either choosing a survey platform that is non-gamified, or a survey platform that is gamified.

Figure 3.1 depicts the choice page where users choose the survey version which they want to use to complete the survey. i.e. they can choose between the plain survey version (the non-gamified version) or the gamified version of the survey with badges. This choice page is shown to the user after the user has seen a guided walkthrough (or tutorial) of both the survey versions. A more detailed explanation of the guided walkthrough is provided in the next section (see Tutorial). This is done in order to provide a visual understanding of the survey versions to the user. It is important to note that this choice is provided to the user only once, and the decision cannot be changed once the user starts the actual survey.

3.2.2 Tutorial

Originally, in the study that we replicated [15], there was no tutorial disaplyed to the user. Therefore, this guided tutorial (or walkthrough) is completely a new addition to our study platform. Users are provided a guided tutorial i.e. users will be walked through a serious of steps through pop-up windows providing a step-by-step explanation that a user has to perform within the survey platform. A few steps of the tutorial are discussed

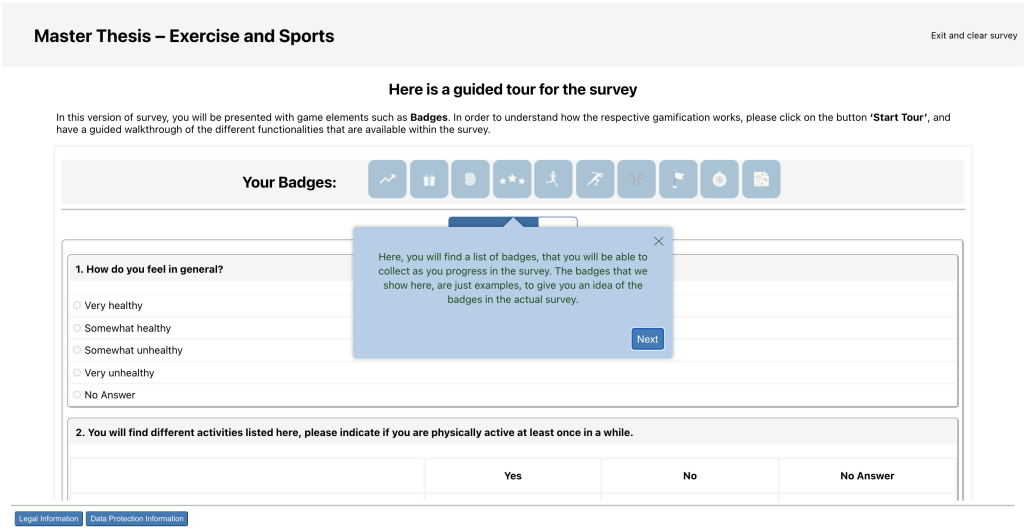


Figure 3.2: A screenshot of the guided tutorial for the survey version with the game element badges. The pop-up tooltip provides a step-by-step explanation of the survey platform.

here. Users will be shown where to exit the survey, users will shown where to view their progress within the survey, additionally users will be shown how to access or make use of the game element that is provided in the survey platform. This tutorial is mainly presented to allow users to experience both survey versions that is provided to them, and to allow them to make an informed choice. The tutorial of the gamified survey explains the users the different functionalities of the game element, namely badges, and how each badge could be unlocked by progressing over the survey.

3.3 Implementation

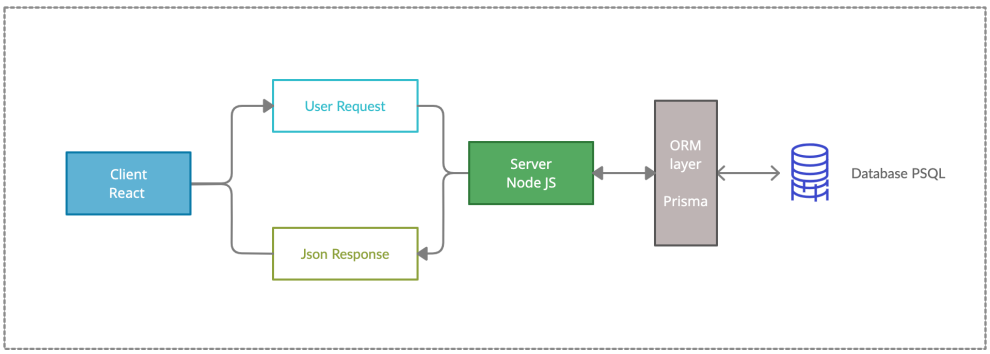


Figure 3.3: Architecture of the developed survey platform.

The survey web application was built based on the client server architecture. Nodejs[2] was used to build the server. The primary reasons to choose Nodejs express[3] as the server are:

- Easy to build scalable solutions.
- Consistency in the language in both Client and Server (Javascript or Typescript).
- Large and active community.

To build the user interface, React.js[4] was used as the frontend framework. React is one way data binding model where the data always flows from top to bottom and offers the environment to create reusable and interactive UI components. The major reasons to use React as the frontend framework are:

- With 165k stars on github it is one of the most widely used frontend frameworks with a large community support.
- Plethora of reusable libraries to speed up the development process.
- Possibility of building client side rendered and server side rendered applications.

The interaction between the client and server was achieved through the REST api.

[2]https://nodejs.org/en/
[3]https://expressjs.com/
[4]https://reactjs.org/

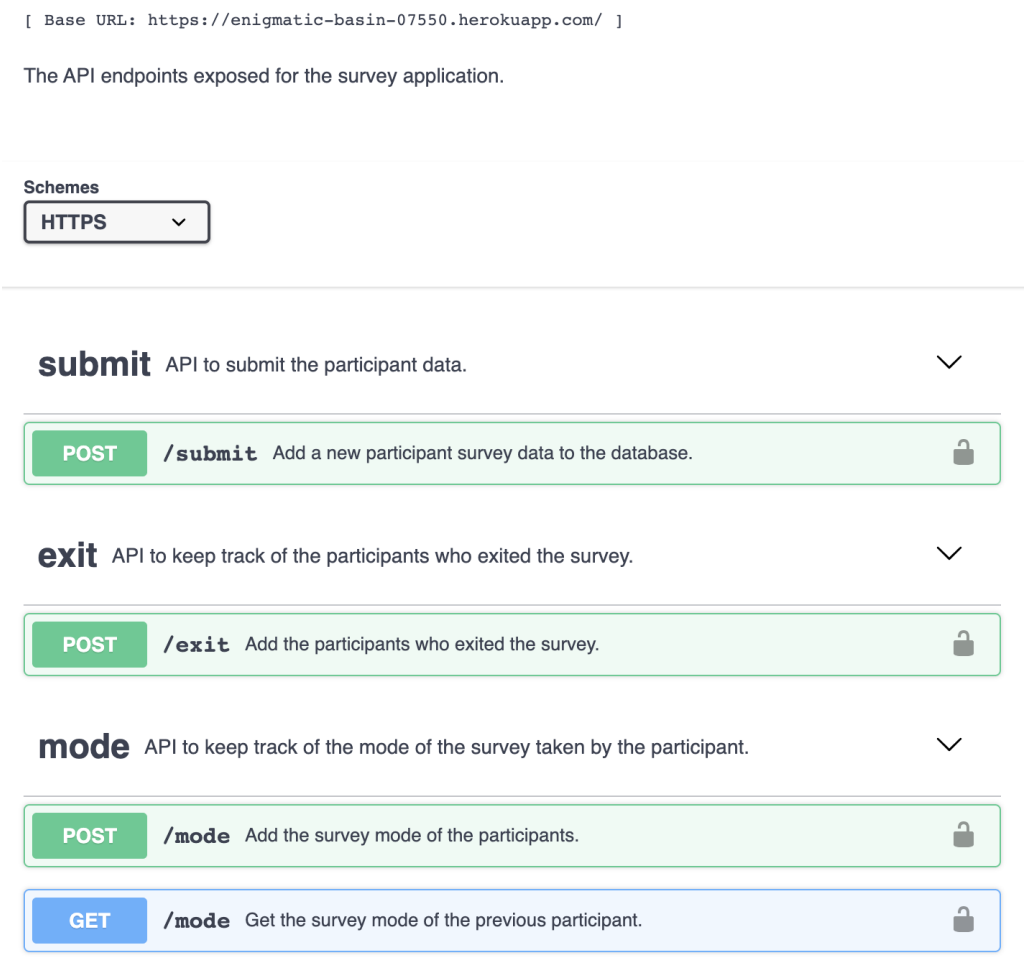


Figure 3.4: API Documentation

PostgreSQL was used as the database to store the participant information. The following tables were created for the application.

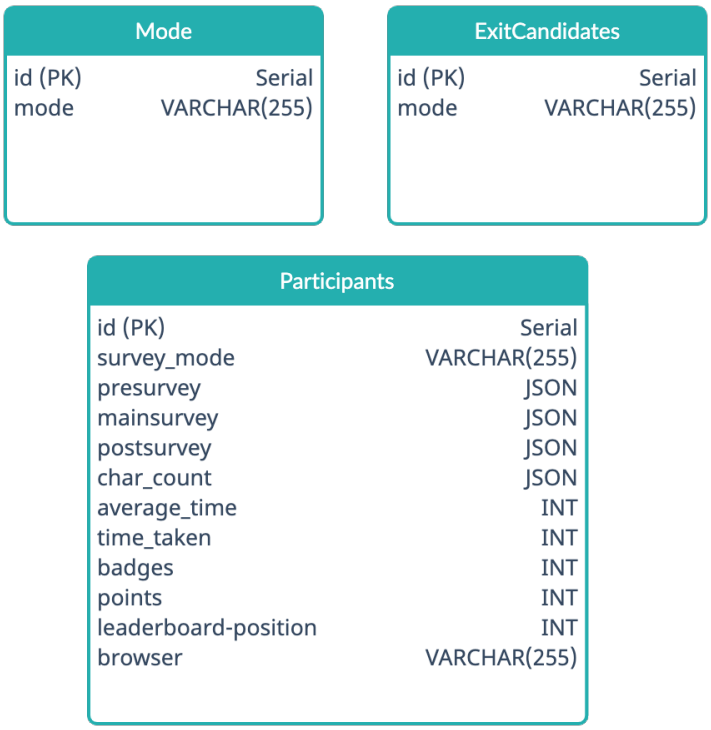


Figure 3.5: Structure of the database, containing the three main tables used for the survey platform. Mode: holds the information regarding which survey mode the user has entered; Exit Candidates: holds information about the users who has quit the survey; Participants: holds information about how the user performs within the survey.

The server and the database were made to communicate through an ORM library called Prisma[5]. The main objective behind using the ORM library was to declaratively define the application data models and build easily. The entire application is built using typescript[6] as the main language. Typescript was used to provide an error free environment to build the application. Also there is great advantage in using typescript with express server, react and prisma. The main advantages are:

- Errors caught at compile time.
- Shape of the APIs are well defined.
- Autocompletion to build the application at a faster pace.

The client, as mentioned, was envisioned using the React framework. In addition to using react there were few other libraries which were used for the building of the application. The core survey was built using the library SurveyJS[7]. However the UI theme for the survey was replicated from the Haams paper. The client uses the react-router library[8]

[5]https://www.prisma.io/
[6]https://www.typescriptlang.org/
[7]https://surveyjs.io/
[8]https://reactrouter.com/

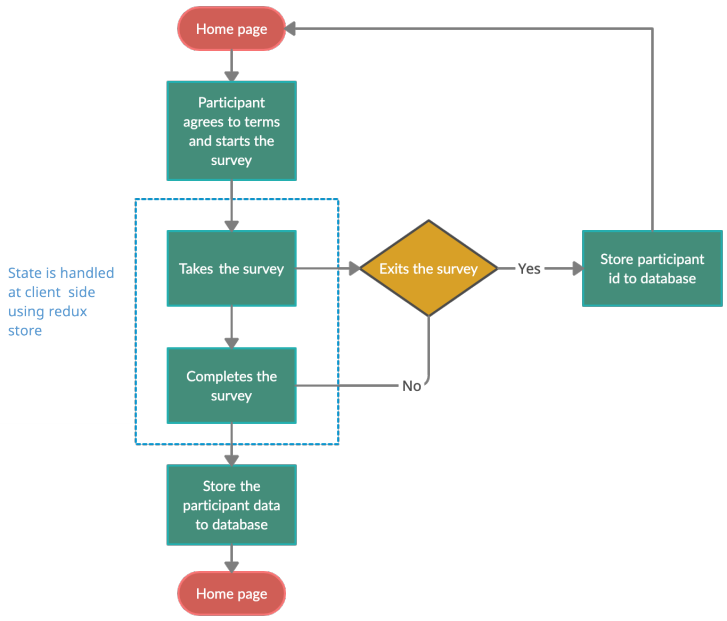


Figure 3.6: Workflow of the survey platform.

to build the navigation between the pages. The entire navigation of the application is handled at the client side using the react-router library. This means that the express server serves the fully built react application at startup and then react-router handles the routing.

The client uses the styled-components[9] library to handle the styling and theming for the user interface. While architecting the application it was observed that due to the gamification element involved there has to be a prominent state management maintained on the client side. For the state management on the client side, react-redux library[10] was used. On a superficial level, the workflow would be that the participants' entire data would be handled by the redux store and only after the completion of the survey the data related to the participant would be pushed to the database through the REST api call to the server.

The algorithm to direct the participant to the survey mode is also handled on the client side. This is done by tracking the survey mode of the previous participant and allotting the incoming participant to the specific survey mode. The algorithm is designed in such a way that the participants are split in the ratio of 1:1:2 for the first study which is basically Traditional Survey: Gamified Survey: Enable/Disable Gamification. The same ratio of 1:1:2 is maintained for the second study which is Badges Gamification: Leaderboard Gamification: Gamification Choice.

The deployment environment is Heroku[11] which is essentially a platform as a service

[9]https://styled-components.com/
[10]https://react-redux.js.org/
[11]https://www.heroku.com/

which helps the developers to build and deploy applications on a cloud environment. The entire application is developed with the use of version control system Github.

3.3.1 Technologies Used

- To implement the survey platform as a web application, we have chosen **React.js** as the frontend framework. React.js is implemented using Typescript as the primary language.
- Since our survey platform is intended to support two languages: English and German, we have used the internationalization library *react-i18next*.
- For implementing the survey questions, we have used the library *survey-react*.
- In order to provide a guided tour to the participants to decide which survey version they would like to use, we have used *react-joyride* library to visually understand and experience the different survey versions that are available.
- *React-redux* has been used as the frontend store. The styling of the components is achieved through *styled-components*.
- Unit tests are written using *JEST* framework with *@testing-library/react*.
- Navigation of the survey pages is achieved through *react-router* library.
- The backend is mainly based on **Node.js Express Server**. However, this is implemented with Typescript as the programming language.
- The survey is implemented using **REST API**.
- **PostgreSQL** has been used as the primary database. We also add an additional ORM layer between the database and the express server. For this purpose, *PRISMA* library has been used as the ORM layer.
- JEST is used as the unit testing framework for the backend as well.
- The primary package manager for both frontend and backend is Yarn.
- Finally, considering the size and complexity of the application, this is built as a monorepo.

Chapter 4
User Study

This chapter gives an overview of the gamified survey user study that was conducted in order to analyze the impact of choice in a gamified setting. The Overview section gives a brief explanation of the different aspects that we considered for both our experiments (Experiment 1 and Experiment 2) and the selection of the game configurations. Subsequently, the next chapters discusses the different study conditions and methods that were considered. Lastly, we discuss the results obtained from the study.

4.1 Overview

The existing body of **Related Work** showed the effectiveness of customization in a variety of contexts (crowd sourcing [21], image tagging [22], cooking themed game [34]). However, this has not yet been investigated in the context of gamified online surveys. This lack of research motivated the thesis to study the effect of choice on user behavior and perception in a gamified online survey. Inspired by the related work, a between-subject experimental design seemed to be the optimal way to investigate this, as each of the participants will only be exposed to a single survey version. This allows the participants to be not aware of other existing versions of the survey, thereby allowing them to exhibit their opinions towards the survey experienced.

- **Experiment 1: Enabling or Disabling Gamification**
- **Experiment 2: Selecting from Fixed Game Configurations**

The first experiment was conducted to investigate the effectiveness of giving the option to users to either enable the gamification provided in the survey, or to disable the gamification completely. For the purpose of investigating on this, we used the game element - badges, in the gamified survey. The second experiment was conducted to investigate whether providing a choice between multiple gamified surveys, without providing the option to disable gamification in the survey is effective or not. The second

study consisted of the game elements - badges, points and leaderboard (Refer Motivation for motivation behind the user studies).

In the next sections, we present both the experiments seperately.

4.2 Experiment 1: Enabling or Disabling Gamification

In this section, we firstly present the study conditions that were used. Secondly, we provide a detailed explanation of the hypotheses that were built for the study. Next, we provide participant details, and lastly present the obtained results.

4.2.1 Conditions

For the first experiment, we introduced three different study conditions namely No Gamification condition (NG), Gamification condition (G), and Choice condition (NG/G) (Refer Table 4.1).

- **No Gamification condition.** Participants in the No Gamification condition were given a non-gamified tutorial, in which a traditional survey without any game design elements is shown. Once the tutorial was done, participants were given with a survey that was about sports and leisure activities, which was replicated from Harms et al. [15]. In this condition, participants are directly given with a traditional survey without having a choice in selection of the survey, whether to have gamification or not.

- **Gamification condition.** Participants in the Gamification condition were given a guided walkthrough of the game element that were available in the survey, and corresponding functionalities of the game element. Once the participant completes the guided walkthrough, they were assigned to the gamified survey that contained similar questions to that of the non-gamified survey (i.e. about sports and leisure activities). Similar to the No Gamification condition, the participants were directly assigned to the gamified survey without having a choice to make a selection on whether to use game element or not.

- **Choice condition.** Participants in the choice condition completed a guided tour of both the survey versions - No Gamification and Gamification. The order in which the participants viewed both tutorials were randomized to avoid ordering bias. This meant that if one participant viewed the non-gamified tutorial first, then the next participant will likely be assigned to the gamified tutorial first. This is done to assure that the ordering effects are being counterbalanced (i.e. one half of the participants saw the gamified tutorial first, and the other half saw the non-gamified survey tutorial first). Once the participants completed both the tutorials, they will be directed to a choice page, where the participant views a snapshot of both the survey versions they experienced in the tutorial. Here, the participants were allowed to make a choice on whether they wanted to enable

gamification or disable the gamification in the survey. This decision splits the Choice condition into two sub-conditions: $\text{Choice}_{\text{Gamification}}$ and $\text{Choice}_{\text{No Gamification}}$. Participants who designed to enable gamification for the survey were assigned to $\text{Choice}_{\text{Gamification}}$, and participants who chose to disable gamification were assigned to $\text{Choice}_{\text{No Gamification}}$ condition. Once this assignment is made, participants remained in the same condition until they complete the entire study. Only the participants in the Choice condition were aware of both the survey versions that were available; therefore they were the only participants who were given a choice to choose their preferred survey version.

Study Conditions	Abbr.	Game Elements
No Gamification	NG	-
Gamification	G	Badges
Choice	NG/G	Badges

Table 4.1: Study Conditions of Experiment 1.

Participants were randomly assigned to either of the three conditions in a 1:1:2 fashion. This allowed us to investigate the effectiveness of having a simple choice to enable or disable gamification in a survey. This is done with the motive of identifying whether users are satisfied when they are allowed to make a personal choice and adapt the system to their desired survey version by having a closer look at the quality of the survey data.

4.2.2 Hypotheses

For the user study, we investigated the effectiveness of providing choice that allows users' to either enable gamification or disable gamification in the system. To study this effect, we formulated the following hypotheses:

- **H1.** Participants who completed the gamified survey exhibited positive psychological outcomes compared to the participants who completed the non-gamified survey, regardless of whether they are given a choice.
  - a Gamification increases user experience compared to non-gamified survey.
  - b Gamification does not have an effect on participants affect states.
- **H2.** Positive psychological outcomes does not improve participants behavior in answering the survey.
  - a Gamification decreases survey completion rate compared to non-gamified survey.
  - b Gamification increases item non-response rate compared to non-gamified survey.
- **H3.** Gamified survey increases the word count in open-ended items compared to non-gamified survey, regardless of whether there is a choice.
- **H4.** Gamified survey increases survey completion time compared to non-gamified survey.

H1 and H2 suggests that gamification has an impact on the psychological outcomes, and not on the behavioral outcomes. H3 and H4 suggests that gamification might lead to positive behavioral outcome in terms of word count, and time spent. In order to mitigate the problem of non-improvement in some of the behavioral attitudes (such as survey completion rate, item non-response rate), we investigate whether customization helps in achieving the desired behavioral outcomes. Therefore, we build the rest of the hypotheses (H5, H6) based on this.

- **H5.** Participants who have a choice to either enable or disable gamification experience positive psychological outcomes than the participants who did not have a choice.
    a Participants with choice encounter improved user experience compared to participants with no choice.
    b Participants with choice encounter improved positive affectivity compared to participants with no choice.
- **H6.** Participants who have a choice to either enable or disable gamification experience positive behavioral changes than the participants who did not have a choice.
    a $\text{Choice}_{\text{enable/disable gamification}}$ increases survey completion rate compared to no choice.
    b $\text{Choice}_{\text{enable/disable gamification}}$ decreases item non-response rate compared to no choice.
    c $\text{Choice}_{\text{enable/disable gamification}}$ increases the word count in open-ended answers compared to no choice.
    d $\text{Choice}_{\text{enable/disable gamification}}$ increases survey completion time compared to no choice.

The derived hypotheses allows us to verify if the customization option helps in improving users' behavioral and psychological outcomes in gamified online surveys. Since previous studies [10, 16] has shown that having a personal choice leads to higher satisfaction and task enjoyment, we intend to measure the same effect in the context of gamified online surveys.

H1a is derived based on the Table 2.5 where some of the studies found out that the usage of game design elements within surveys had a positive impact on users' psychological outcomes, resulting in higher user experience, and thereby improving aspects such as perceived fun, interest, preference, and overall usability of the survey. H1b is derived based on the study that we replicate [15]. This is done to measure whether the participants affectivity changes before and after the survey. This allows us to determine whether the participants have any positive affects or negative affects after filling out the survey. Our hypotheses suggest that gamification does not cause any changes to the participants moods and emotions, which is positive because the participant finishes the survey in the same mood as he started it with.

H2 is derived based on the Table 2.3 where most of the studies found out that positive user experience does not necessarily improve respondents' behavior in answering the survey i.e. studies show that a significant change and improvment in how users perceive

the survey does not necessarily affect the way in which they fill out the survey. Based on the results obtained by different authors in their studies, there seems to be little motivation for respondents in answering the surveys, resulting in higher amounts of survey dropouts (H2a), item non-responses (H2b).

H3 is derived based on the Table 2.3 where some of the studies (7 out of 14 studies) measured, and found out that a few studies (2 out of 7 studies) found an increase in the number of characters used in answering the open-ended questions. Therefore, in our study, we expect to receive more words in free-text fields.

H4 is derived based on the Table 2.3 where some of the studies (9 out of 14) measured, and found out that a few studies (3 out of 7) found an increase in the amount of time spent on answering the survey. The authors interpret this result as a positive outcome, since participants are willingly spending more time in answering the survey. Therefore, in our study, we expect that gamification would increase the overall survey completion time.

H5 and H6 follow from the previously presented related works, which not only show that having a choice is generally perceived well [10, 16], but also have an impact on their behavior. As a consequence, we expected participants in the Choice condition to have an overall positive effect in terms of enjoyment, as well as in the quality of responses given in the survey.

4.2.3 Participants

80 participants completed the study i.e. participants were allocated to the three study conditions, non-gamified, gamified, and choice condition, in a 1:1:2 manner. Therefore 20 participants completed the non-gamified survey, 20 participants completed the gamified survey, and 40 participants completed the survey with a choice. Out of the 40 participants who were assigned to the choice condition, 26 participants decided to enable gamification to take the survey, and 13 participants decided to disable gamification for the main survey, and one participant did not complete the survey. Most of the participants 37 participants (46.8%) were aged between 18 - 24 years, 31 participants (39.2%) were aged between 25 - 34 years, 7 participants (8%) were aged between 35 - 44 years, 3 participants were below 18 (4%), and only 1 participant was aged above 45 years. Out of 79 participants, 42 participants (52.5%) were female and 36 participants (46.3%) were male. One participant preferred not to answer.

4.2.4 Results

In this section, we discuss the data that was obtained in the user study. The results of the studies are derived based on descriptive statistics of the objective measurements as well as, based on inferential statistics, to verify and validate the statistical significance of the tests conducted. The section is structured as follows: Firstly, performance-related results, mainly item non-response rate, number of words in open questions, survey completion time, are presented. Subsequently, results on the psychological reactions obtained from the self-reported questionnaire are reported. Next, we give an insight into the HEXAD user type results, and participants of which user type performed better in the survey.

Performance-related Results

- **Item Response Rate.** We first analyzed the number of questions that were answered in the main survey. The mean responses were deviated across conditions only marginally (No Gamification: M = 25, Gamification: M = 25.6, $\text{Choice}_{\text{Gamification}}$: M = 24.66, $\text{Choice}_{\text{No Gamification}}$: M = 24.46). To compare the mean amount of responses across the conditions, we calculated a one-way ANOVA, for which we did not find a statistically significant difference between the groups (Welch's F(2, 76) = 1.950, p = 0.149). We report the Welch's F here, as Levene's test for homogenity of variance was significant for our datatset (p = 0.001) and Welch's F is more robust in this case. Table 4.2 show detailed overview of the amount of responses generated by participants in the main task across individual conditions.

Condition	n	Number of generated responses					
		M	SD	Mdn	Min	Max	
No Gamification	20	25	1.86	26	20	26	
Gamification	20	25.6	1.35	26	20	26	
Choice	39	24.58	2.07	26	20	26	

Table 4.2: Detailed overview of the amount of responses generated in the main survey (n = number of participants in the condition, M = mean amount of generated responses, SD = standard deviation, Mdn = median, Min = minimum, Max = maximum).

Condition	P_tukey's HSD	Significant
Gamified Vs Choice	0.127	No
Gamified Vs Baseline	0.568	No
Baseline Vs Choice	0.704	No

Table 4.3: The table shows pairwise comparisons across the conditions for the performance-related result - item response rate.

- **Number of characters in free-text fields.** We analyzed the number of characters generated by the participants in the free-text fields. To compare the means across the conditions, we calculated one-way ANOVA. We found no statistically significant difference in the average character count across the conditions (F(2, 76) = 0.159, p > 0.05). The means of character count across the three conditions are as follows: M(baseline) = 75.7, M(gamified) = 65.9, and M(choice) = 73.9. The Kruskal-Wallis test computes a p-value (=0.796) that is greater than the significance level alpha = 0.05, indicating that there are no significant differences across the conditions.

Condition	P_tukey's HSD	Significant
Gamified Vs Choice	0.878	No
Gamified Vs Baseline	0.865	No
Baseline Vs Choice	0.994	No

Table 4.4: The table shows pairwise comparisons across the conditions for the performance-related result - number of characters in free-text fields.

- **Survey Completion Time.** We analyzed the total time taken by the participants to

answer the items in the main survey. To compare the mean time taken across the conditions, we calculated one-way ANOVA. We found no statistically significant difference in the average time taken to answer the survey according to different conditions ($F(2, 76) = 1.343$, $p > 0.05$). The mean time taken across the three conditions are as follows: M(baseline) = 320.35, M(gamified) = 301.64, and M(choice) = 230.00. The Kruskal-Wallis test computes a p-value (=0.065) that is slightly greater than the significance level alpha = 0.05, indicating that we cannot prove that the survey completion time is higher in Choice condition.

Condition	P_tukey's HSD	Significant
Gamified Vs Choice	0.931	No
Gamified Vs Baseline	0.289	No
Baseline Vs Choice	0.355	No

Table 4.5: The table shows pairwise comparisons across the conditions for the performance-related result - survey completion time.

Psychological Results

We analyzed the responses to the IMI questionnaire, which was filled out by the participants after completing the main survey. The resulting scores depicted in Figure 4.1, where we cal clearly see that participants in the choice condition perceived the provided choice positively by giving higher ratings for the choice subscale.

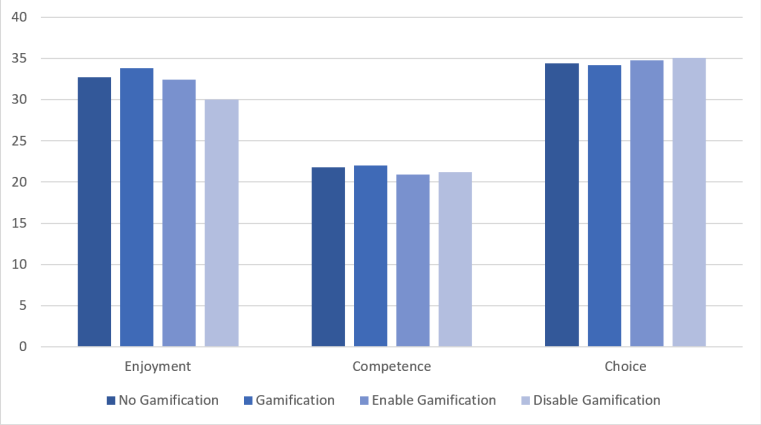


Figure 4.1: Mean ratings for the individual subscale of the IMI: Perceived Enjoyment; Perceived Competence; Perceived Choice.

- **Perceived Choice.** We calculated a one-way ANOVA for the subscale - perceived choice. The homogeneity of variance was not violated (Levene's test not being significant with p>.2), confirming that the variances between the groups are approximately equal. By comparing the means of perceived choice across the study conditions, we did not find any statistically significant difference (No Gamification: M = 34.4, Gamification: M = 34.2 Choice: M = 34.9, p = 0.870). Furthermore, we used the Fisher's LSD post-hoc procedure for pairwise comparisons of the mean

scores. The scores for the perceived choice subscale were not significantly different between any of the conditions.

Condition	Perceived Choice subscale		
	Difference	**P**	**Significant**
Choice vs Gamification	6.467	**<0.627**	No
Choice vs No Gamification	2.067	0.725	No
No Gamification vs Gamification	0.907	**0.003**	No

Table 4.6: The table shows pairwise comparisons across the conditions for perceived choice subscale.

- **Perceived Competence.** We calculated a one-way ANOVA for the subscale - perceived competence. The homogeneity of variance was not violated (Levene's test not being significant with p>.2), confirming that the variances between the groups are approximately equal. By comparing the means of perceived competence across the study conditions, we did not find a statistically significant difference (No Gamification: M = 29.8, Gamification: M = 29.9, Choice: M = 28.6, p = 0.600). Furthermore, we used the Fisher's LSD post-hoc procedure for pairwise comparisons of the mean scores. The scores for the perceived choice subscale were not significantly different between Choice and Gamification (M = 28.6 vs M = 29.9, p = 0.383), and between the conditions Gamification and No Gamification (M = 29.9 vs M = 29.8, p = 0.929), and between the Choice and No Gamification condition was not significant (M = 28.6 vs M = 29.8, p = 0.441).
- **Perceived Enjoyment.** Similarly, for the enjoyment subscale, we calculated a one-way ANOVA. The homogeneity of variance was not violated (Levene's test not being significant with p>.2), confirming that the variances between the groups are approximately equal. By comparing the means of perceived competence across the study conditions, we did not find a statistically significant difference (No Gamification: M = 32.7, Gamification: M = 33.8, Choice: M = 31.5, p = 0.514). Furthermore, we used the Fisher's LSD post-hoc procedure for pairwise comparisons of the mean scores, and did not find any significant differences between the groups.

Self-reported User Experience and User Types Results

Lastly, we analyzed the *Positive and Negative Affect Schedule* scale, and a significant difference across the conditions was found in Pre-Survey *Positive Affect* subscale (p = 0.036). The analysis revealed a difference between $\text{Gamification}_{\text{Points and Leaderboard}}$ and $\text{Gamification}_{\text{Badges}}$ (M = 14.06 vs M = 11.20, p = 0.027), and between Choice and $\text{Gamification}_{\text{Badges}}$ (M = 13.86 vs M = 11.20, p = 0.018). By taking a closer look at the mean values, the analysis suggests that participants in the $\text{Gamification}_{\text{Points and Leaderboard}}$ and Choice conditions experienced positive affect significantly more while starting with the survey.

We analyzed self-reported measures of 2 questionnaires (I-PANAS-SF, and HEXAD User Types). The participants were asked to fill out the positive and negative affect scale and HEXAD user types questionnaire before starting with the main survey. We use this data to analyze whether the participants user type had an effect on their overall performance

and interest in the survey. Having a closer look at the participants from the *Choice* condition, 10 participants out of the 24 who enabled gamification (42%), were of the user type *Player* and *Achiever*. 8 participants out of the 15 who disabled gamification (53%) were also of the same player type *Player* and *Achiever*. We mainly report the findings our the user type PLA and ACH since the chosen game element - badges is an indication of certain achievements within a task.

The highest number of badges achieved by the participants were 7 badges out of the available 10 badges. Out of the 8 participants who scored the highest number of badges, 5 participants (62%) belong to the user type player and achiever.

International-Positive and Negative Affect Schedule-Short Form (I-PANAS-SF)

Next, we report the scores obtained from I-PANAS-SF questionnaire which was collected before the participant entered the main survey, and after completing the main survey, in order to analyze if the survey had an impact on the positive and negative affect states of each participant. There was no significant differences across the conditions found in both *Positive Affect* subscale and *Negative Affect* subscale collected during the pre- and post survey questionnaires.

4.3 Experiment 2: Selecting from Fixed Game Configurations

4.3.1 Conditions

For the second experiment, we introduced three study conditions namely $\text{Gamification}_{\text{Badges}}$ condition (G1), $\text{Gamification}_{\text{Points and Leaderboard}}$ condition (G2), and Choice condition (G1/G2) (Refer Table 4.7).

Study Conditions	Abbr.	Game Elements
$\text{Gamification}_{\text{Badges}}$ condition	G1	Badges
$\text{Gamification}_{\text{Points and Leaderboard}}$ condition	G2	Points, Leaderboard
Choice	G1/G2	Badges, Points, Leaderboard

Table 4.7: Study Conditions of Experiment 2.

- **$\text{Gamification}_{\text{Badges}}$ condition.** Participants in the $\text{Gamification}_{\text{Badges}}$ condition were given a guided walkthrough, in which a gamified survey with the game design element - badges, is shown. In the tutorial, participants are exposed to the functioning of the game element, and provides instructions to participants' on how to unlock badges as they progress through the survey. Once the tutorial was done, participants were given with a survey on sports and leisure activities. In this condition, participants are provided with a forced gamification survey without having a choice in selecting which survey version they want to use.

- **$\text{Gamification}_{\text{Points and Leaderboard}}$ condition.** Participants in this condition were given with a guided walkthrough of the game elements - points and leaderboard that

were available in the survey, and it's corresponding functionalities. Instructions were provided as to how a participant could achieve points as they progress through the survey, and how they could compare themselves with participants' on the Leaderboard. Once the participant completes the guided walkthrough, they were assigned to the gamified survey that contained similar questions to that of the G1 condition (i.e. about sports and leisure activities). Similar to the previous study condition, the participants were forcefully assigned to the gamified survey containing points and leaderboard, without having a choice to make a selection on which gamified survey version they wanted to use.

- **Choice condition.** Participants in the choice condition completed a guided tour of both the survey versions - $\text{Gamification}_{\text{Badges}}$ and $\text{Gamification}_{\text{Points and Leaderboard}}$. The order in which the participants viewed both tutorials were randomized so as to avoid ordering bias. This meant that if one participant viewed the badges tutorial first, then the next participant will likely be assigned to the points and leaderboard tutorial first. This is done to assure that the ordering effects are being counterbalanced (i.e. one half of the participants saw the gamified tutorial first, and the other half saw the non-gamified survey tutorial first). Once the participants completed both the tutorials, they will be directed to the Choice page, where the participant views a snapshot of both the survey versions they experienced in the tutorial. Here, the participants were allowed to make a choice on whether they wanted to use badges in the survey, or to use points and leaderboard in the survey. This decision splits the Choice condition into two sub-conditions: $\text{Choice}_{\text{Badges}}$ and $\text{Choice}_{\text{Points and Leaderboard}}$.

  Participants who designed to use badges in the survey were assigned to $\text{Choice}_{\text{Badges}}$, and participants who chose to use points in the survey were assigned to $\text{Choice}_{\text{Points and Leaderboard}}$ condition. Once this assignment is made, participants remained in the same condition until they complete the entire study. Only the participants in the Choice condition were aware of both the survey versions that were available; therefore they were the only participants who were given a choice to choose their preferred survey version.

Participants were randomly assigned to either of the three conditions in a 1:1:2 fashion (Participants in the Choice condition will receive double the number of participants than the forced Gamification conditions). This experiment aimed at identifying whether users are satisfied when they are allowed to make a personal choice and adapt the system to their desired survey version, by having a closer look at the quality of the survey data.

4.3.2 Hypotheses

The second experiment focused on the second research goal of this thesis: to investigate the effectiveness of choice that allows users' to choose between different gamified interventions consisting of different game configurations (or game elements). To study this effect, we formulated the following hypotheses:

- **H7:** Participants who have a choice to choose between game configurations experience positive psychological changes than the participants who did not have a choice.

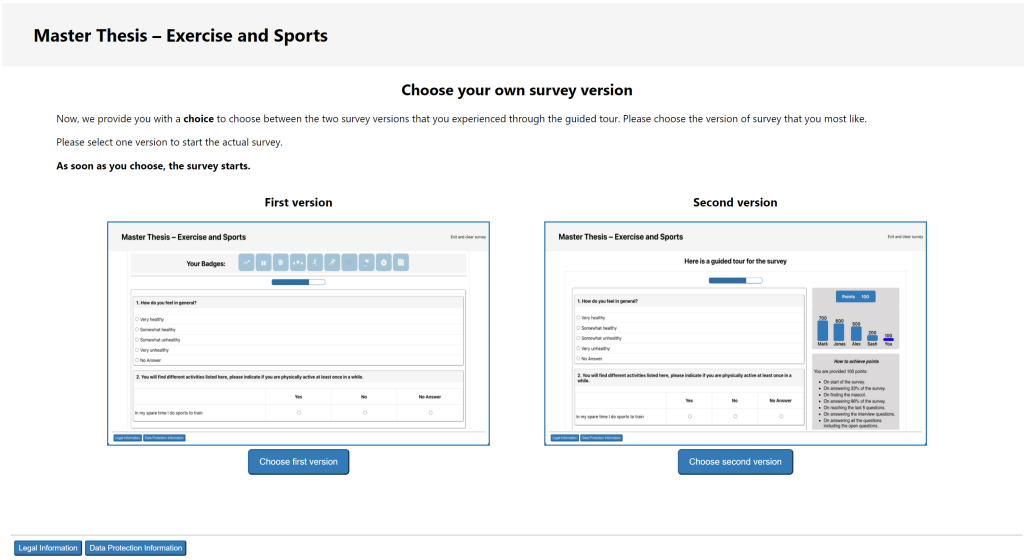


Figure 4.2: Screenshot of the choice page that was displayed to the participants in the Choice condition (Experiment 2).

a Participants with choice encounter higher user experience compared to the participants who had no choice.

b Participants with choice encounter higher positive affectivity compared to the participants who had no choice.

- **H8:** Participants who have a choice to choose between game configurations experience positive behavioral changes than the participants who did not have a choice.

  a $\text{Choice}_{G1/G2}$ increases survey completion rate compared to forced gamification.

  b $\text{Choice}_{G1/G2}$ decreases item non-response rate compared to forced gamification.

  c $\text{Choice}_{G1/G2}$ increases the word count in open-ended answers compared to forced gamification.

  d $\text{Choice}_{G1/G2}$ has positive effect on the degree of careless responses.

  – $\text{Choice}_{G1/G2}$ reduces the number of straightlined responses compared to forced gamification.

  – $\text{Choice}_{G1/G2}$ reduces speeding compared to forced gamification.

  – $\text{Choice}_{G1/G2}$ reduces middle-point responses compared to forced gamification.

  – $\text{Choice}_{G1/G2}$ reduces the number of no-answer responses compared to forced gamification.

  e $\text{Choice}_{G1/G2}$ increases survey completion time compared to forced gamification.

  – Hypotheses H7 and H8 are derived based on the previously presented related works, which show that having a personal choice is perceived well, and ultimately leads to positive affective and motivational consequences [10, 16]. Therefore, we expect that participants in the choice condition would exert positive psychological and behavioral outcomes compared to participants who did not have the choice to choose between different fixed game configurations.

4.3.3 Participants

60 participants completed the study i.e. participants were allocated to the three study conditions, G1, G2, G1/G2, in a 1:1:2 manner. Therefore 15 participants completed the gamified survey with badges, 15 participants completed the gamified survey with points and leaderboard, and 30 participants completed the survey with a choice between the two gamified versions. Out of the 30 participants who were assigned to the choice condition, 14 participants decided to apply badges enabled gamification to take the survey, and 16 participants decided to apply points and leaderboard enabled gamification for the main survey.

4.3.4 Results

In this section, we discuss the data that was acquired in the study. Firstly, performance related results, mainly item non-response rate, number of words in open questions, survey completion time, are presented. Subsequently, results on the psychological reactions obtained from the self-reported questionnaire are reported. Next, we give an insight into the HEXAD user type results, and participants of which user type performed better in the survey.

Performance-related Results

- **Item Response Rate.** We first analyzed the number of questions that were answered in the main survey. The mean responses were deviated across conditions only marginally ($\text{Gamification}_{\text{Badges}}$: M = 23.4, $\text{Gamification}_{\text{Points and Leaderboard}}$: M = 22.9, Choice: M = 22.9). To compare the mean amount of responses across the conditions, we calculated a one-way ANOVA, for which we did not find a statistically significant difference between the groups (Welch's F(2, 57) = 0.072, p = 0.931). Table 4.8 show detailed overview of the amount of responses generated by participants in the main task across individual conditions.

Condition	n	Number of generated responses					
		M	SD	Mdn	Min	Max	Sum
Badges	15	23.4	6.65	26	0	26	352
Points and Leaderboard	15	22.9	2.86	24	17	26	344
Choice	30	22.9	4.29	24.5	9	26	688

Table 4.8: Detailed overview of the amount of responses generated in the main survey (n = number of participants in the condition, M = mean amount of generated responses, SD = standard deviation, Mdn = median, Min = minimum, Max = maximum).

Condition	P_tukey's HSD	Significant
Badges Vs Choice	0.932	No
Badges Vs Points and Leaderboard	0.949	No
Points and Leaderboard Vs Choice	1.000	No

Table 4.9: Results of Tukey (HSD) - Analysis of the differences between the categories with a confidence interval of 95%.

- **Number of characters in free-text fields.** We analyzed the number of characters generated by the participants in the free-text fields in the main survey. To compare the means across the conditions, we calculated one-way ANOVA. We found no statistically significant difference in the average character count across the conditions (F(,) =). The means of character count across the three conditions are as follows: M(baseline) = 75.7, M(gamified) = 65.9, and M(choice) = 73.9. The Kruskal-Wallis test computes a p-value (=0.796) that is greater than the significance level alpha = 0.05, indicating that there are no significant differences across the conditions.

- **Survey Completion Time.** We analyzed the total time taken by the participants to answer the items in the main survey. To compare the mean time taken across the conditions, we calculated one-way ANOVA. We found no statistically significant difference in the average time taken to answer the survey according to different conditions ($F(2, 76) = 1.343$, $p > 0.05$). The mean time taken across the three conditions are as follows: M(baseline) = 320.35, M(gamified) = 301.64, and M(choice) = 230.00. The Kruskal-Wallis test computes a p-value (=0.065) that is slightly greater than the significance level alpha = 0.05, indicating that we cannot prove that the survey completion time is higher in Choice condition.

Psychological Results

We analyzed the responses to the IMI questionnaire, which was filled out by the participants after completing the main survey. The resulting scores depicted in Figure 4.3, where we cal clearly see that participants in the choice condition perceived the provided choice positively by giving higher ratings for the choice subscale.

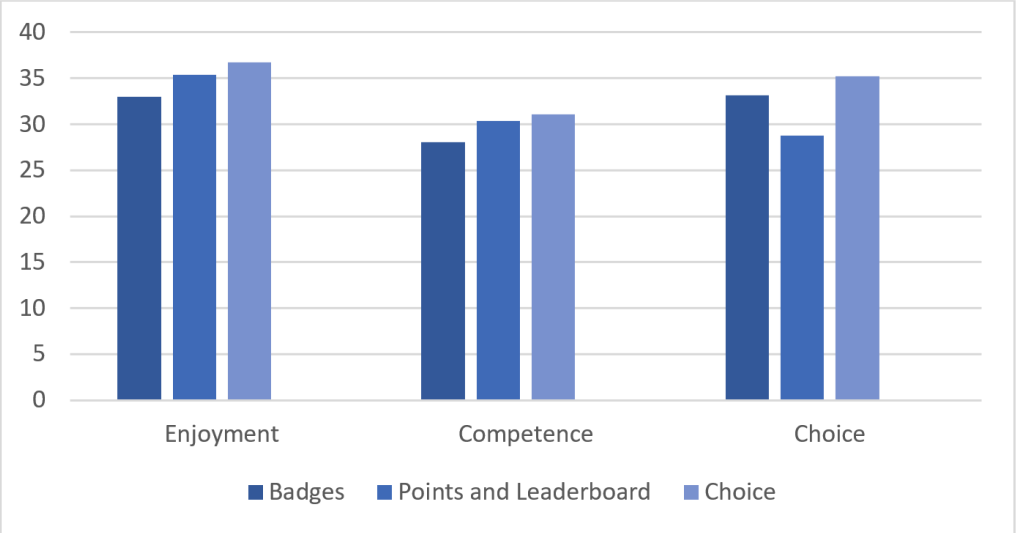


Figure 4.3: Mean ratings for the individual subscale of the IMI: Perceived Enjoyment; Perceived Competence; Perceived Choice.

- **Perceived Choice.** We calculated a one-way ANOVA for the subscale - perceived choice. The homogeneity of variance was not violated (Levene's test not being significant with p>.2), confirming that the variances between the groups are approximately equal. By comparing the means of perceived choice across the study conditions, we found a statistically significant difference ($\text{Gamification}_{\text{Badges}}$: M = 33.1, $\text{Gamification}_{\text{Points\& Leaderboard}}$: M = 28.7, Choice: M = 35.2, p < 0.0001). Furthermore, we used the Fisher's LSD post-hoc procedure for pairwise comparisons of the mean scores. The scores for the perceived choice subscale were significantly different between Choice and $\text{Gamification}_{\text{Points and Leaderboard}}$ (M = 35.2 vs M = 28.7, p < 0.0001), and between the conditions $\text{Gamification}_{\text{Badges}}$ and $\text{Gamification}_{\text{Points and Leaderboard}}$ (M = 33.1 vs M = 28.7, p = 0.003). Even though the difference between the Choice and $\text{Gamification}_{\text{Badges}}$ condition was not significant (M = 35.2 vs M = 33.1, p = 0.097), it seems as if the Choice condition overall increased participants' feeling of having a choice.

Condition	Perceived Choice subscale		
	Difference	**P**	**Significant**
Choice vs Points and Leaderboard	6.467	**<0.0001**	**Yes**
Choice vs Badges	2.067	0.097	No
Badges vs Points and Leaderboard	4.400	**0.003**	**Yes**

Table 4.10: The table shows pairwise comparisons across the conditions for perceived choice subscale.

- **Perceived Competence.** We calculated a one-way ANOVA for the subscale - perceived competence. The homogeneity of variance was not violated (Levene's test not being significant with p>.2), confirming that the variances between the groups are approximately equal. By comparing the means of perceived competence across the study conditions, we did not find a statistically significant difference ($\text{Gamification}_{\text{Badges}}$: M = 28.0, $\text{Gamification}_{\text{Points\& Leaderboard}}$: M = 30.3, Choice: M = 31.1, p = 379). Furthermore, we used the Fisher's LSD post-hoc procedure for pairwise comparisons of the mean scores. The scores for the perceived choice subscale were not significantly different between Choice and $\text{Gamification}_{\text{Points and Leaderboard}}$ (M = 31.1 vs M = 30.3, p = 0.725), and between the conditions $\text{Gamification}_{\text{Badges}}$ and $\text{Gamification}_{\text{Points and Leaderboard}}$ (M = 28.0 vs M = 30.3, p = 0.368), and between the Choice and $\text{Gamification}_{\text{Badges}}$ condition was not significant (M = 31.1 vs M = 28.0, p = 0.167).

- **Perceived Enjoyment.** Similarly, for the enjoyment subscale, we calculated a one-way ANOVA. The homogeneity of variance was not violated (Levene's test not being significant with p>.2), confirming that the variances between the groups are approximately equal. By comparing the means of perceived competence across the study conditions, we did not find a statistically significant difference ($\text{Gamification}_{\text{Badges}}$: M = 33.0, $\text{Gamification}_{\text{Points and Leaderboard}}$: M = 35.4, Choice: M = 36.7, p = 0.111). Furthermore, we used the Fisher's LSD post-hoc procedure for pairwise comparisons of the mean scores. The scores were significantly different between Choice and $\text{Gamification}_{\text{Badges}}$ (M = 36.7 vs M = 33.0, p = 0.037). However, the scores were not significant between the conditions $\text{Gamification}_{\text{Badges}}$ and $\text{Gamification}_{\text{Points and Leaderboard}}$ (M = 33.0 vs M = 35.4, p = 0.239), and between the Choice and $\text{Gamification}_{\text{Points and Leaderboard}}$ condition was not significant (M = 36.7 vs M = 35.4, p = 0.449).

Condition	IMI					
	Enjoyment		Competence		Choice	
	M	S.D.	M	S.D.	M	S.D.
Badges	33.0	6.30	28.0	5.49	33.1	3.62
Points and Leaderboard	35.4	5.94	30.3	7.70	28.7	3.10
Choice	36.7	4.87	31.1	6.99	35.2	4.31

Table 4.11: Self-reported Results - IMI Scale

Self-reported User Experience and User Types Results

Lastly, we analyzed the *Positive and Negative Affect Schedule* scale, and a significant difference across the conditions was found in Pre-Survey *Positive Affect* subscale ($p = 0.036$). The analysis revealed a difference between $\text{Gamification}_{\text{Points and Leaderboard}}$ and $\text{Gamification}_{\text{Badges}}$ ($M = 14.06$ vs $M = 11.20$, $p = 0.027$), and between Choice and $\text{Gamification}_{\text{Badges}}$ ($M = 13.86$ vs $M = 11.20$, $p = 0.018$). By taking a closer look at the mean values, the analysis suggests that participants in the $\text{Gamification}_{\text{Points and Leaderboard}}$ and Choice conditions experienced positive affect significantly more while starting with the survey.

4.4 Method

This section is common to both Experiment 1 and Experiment 2.

A link to the online survey platform was shared on survey software (*SurveySwap*), social media and other networking platforms such as LinkedIn, Twitter, (web) WhatsApp. We explicitly mentioned to the participants to take the survey in laptops and computers, rather than on handheld devices like mobile phones. Furthermore, participants who clicked on the link via a smaller screen were not allowed to participate until the screen size was maximized. This was done to ensure that the features of the survey platform such as game elements were displayed properly. Participants who clicked on the link were directed to an introductory webpage where they were informed about the context and approximate duration to complete the survey. Here, no additional information were revealed about the study such as the true purpose of the study. This was done with the intent to avoid persuasion of participants [3], since this study is solely based on one's motivation and requires participants to truly exhibit their experiences during the survey. Hence, we decided that revealing the research purpose would diminish the effects. Once the participant has read and understood the data protection regulations, they were requested to provide consent to participate in the survey.

In the next page, participants were provided with a pre-survey questionnaire containing 10 basic questions about how one feels psychologically before entering the survey. For this purpose, the I-PANAS-SF standardized questionnaire [36] was used where participant chooses on a 5-point Likert scale (ranging from extremely to not at all) on how positive or negative they feel (example: 'Active', 'Nervous', 'Attentive'). Followed by this, 24-item HEXAD user types questionnaire [37] were used to collect information about a participants' user type (example: 'Player', 'Achiever'). Once this data is collected, we assign the participant to one of the six conditions. After entering the condition that the participant has been allocated to, they receive a guided tour of how to proceed with

the survey version that they received, and get to know what are the different game elements that are available, and the overall functioning of the survey platform. After experiencing the guided tour, participants enter the actual survey session where a survey related to sports and leisure activities is shown. For participants in the Choice condition, an additional page displaying information regarding the offered choice is shown (see Figure 4.2). Here, for experiment 1, they are allowed to choose which version of the survey platform - non-gamified or gamified - they wanted to keep for the main survey. And for participants assigned to experiment 2, they are allowed to choose which version of the gamified survey platform - badges or points and leaderboard - they wanted to keep for the main survey. On the page, we provided a short informative text, in which we explained the choice to the participants. We did not reveal the participants any terms such as 'game', 'game elements', or 'gamification', but instead mentioned the choices as 'survey versions' [37]. Besides the information, snapshots of the two versions were displayed for the participants, to have a clear visual idea of each of the survey version. The main survey (on sports and leisure activities) consists of 26 voluntary questions including open questions, multiple choice questions, and single choice questions (example: open question - *There is a variety of physical activities, from playing football to things like rafting, hiking and dancing. What do you do?*).

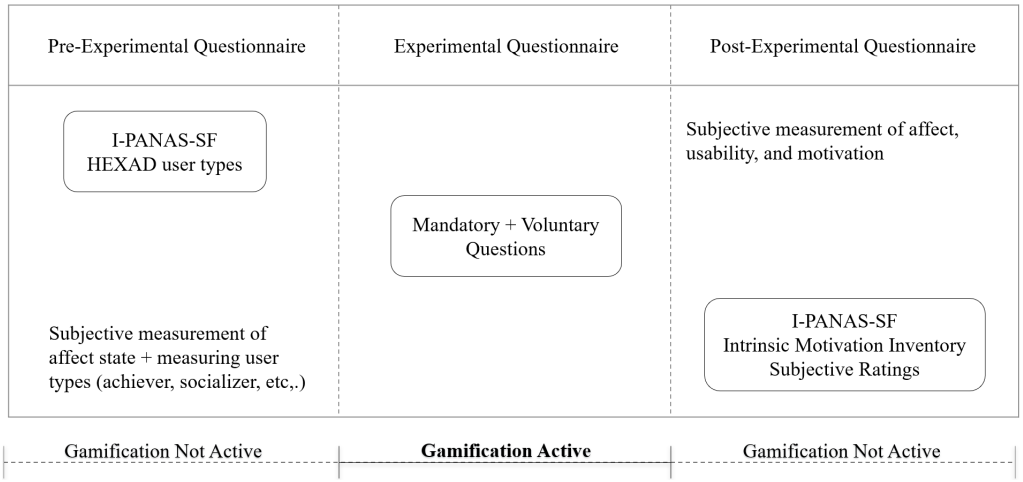


Figure 4.4: Experimental Design (If a participant belongs to a condition where the survey version is gamified, then gamification is kept active only while answering the sports and leisure activities survey.

After completing the survey, the participants receive a post-survey questionnaire containing the same ten questions from I-PANAS-SF [36] that was previously asked during the pre-survey - in order to determine a participants feeling after completing the survey. 20 items from the Intrinsic Motivation Inventory (IMI) were included in order to measure how users perceived the survey with respect to choice, competence, and overall enjoyment by using a 7-point Likert scale (ranging from 'Strongly Disagree' to 'Strongly Agree'). The following subscales were used for this purpose: Perceived Choice, Perceived Competence, Enjoyment/Interest subscale. At the end of the post-survey questionnaire, we include certain items for self-reported measures on how participants view having a game element in their survey (Refer Appendix A). Figure 4.4 depicts the structure of the experiment.

4.5 Summary of Results

R1: Participants player type had no impact for their choice

By using the data obtained in the Choice condition, we calculated a binary logistic regression to analyze whether participants' player type had an impact on the participants' choice of the survey version - in other words, whether certain player types were more likely to enable gamification or more likely to disable gamification than others, or whether players of certain types were more likely to activate badges in a survey than points and leaderboard. The regression analysis did not reach statistical significance for any of the player types ($p>0.40$), meaning that the player types did not have an affect on the participants' decision.

R2: Participants in the Choice condition appreciated the given choice to choose their survey version

We evaluated the answers to the self-reported measures, where participants gave their opinion on of the game element. We had a closer look at the results obtained in the self-reported questionnaire that was asked at the end of the post-survey questionnaire. In particular, we dderived participants' answers to the six items that measured their perception of the game elements (see Figure 4.5).

As already described, the six statements that measure the perception of game elements were rated on a 5-point Likert scale. We asked unique statements in each of the three conditions (No Gamification, Gamification, and Choice condition), therefore we did not individually compare the results across the conditions. Participants in the Gamification condition reported with the highest mean rating that they liked that a game element was available in the survey (No Gamification: M = 4.15, SD = 1.18). Similarly, particpants in the Choice condition reported with a high mean rating that they liked that a choice was available to enable or disable the game element in the survey (Choice: M = 3.74, SD = 1.18). For the fifth statement (Figure 4.5) asked in the Choice condition, we calculated an independent t-test to compare the respective mean score between the $\text{Choice}_{\text{Gamification}}$ and $\text{Choice}_{\text{No Gamification}}$ condition and found that the difference was significant ($p<0.05$). This suggests that participants in $\text{Choice}_{\text{Gamification}}$ enjoyed the choice to enable or disable the game element than the participants in the $\text{Choice}_{\text{No Gamification}}$.

Condition	**Enjoyment subscale**		
	M	**SD**	**Mdn**
No Gamification	32.75	7.23	33.5
Gamification	33.8	6.48	33
$\text{Choice}_{\text{Gamification}}$	32.45	8.94	35
$\text{Choice}_{\text{No Gamification}}$	30	5.31	29

Table 4.12: The table shows the ratings of the enjoyment subscale of the IMI recorded during the post-survey questionnaire.

To further find evidence for the assessment, we had a closer look at the enjoyment subscale measures collected at the end of the post-survey questionnaire. The ratings did not show significant differences across the conditions ($F(2,76) = 0.672$, $p = 0.513$).

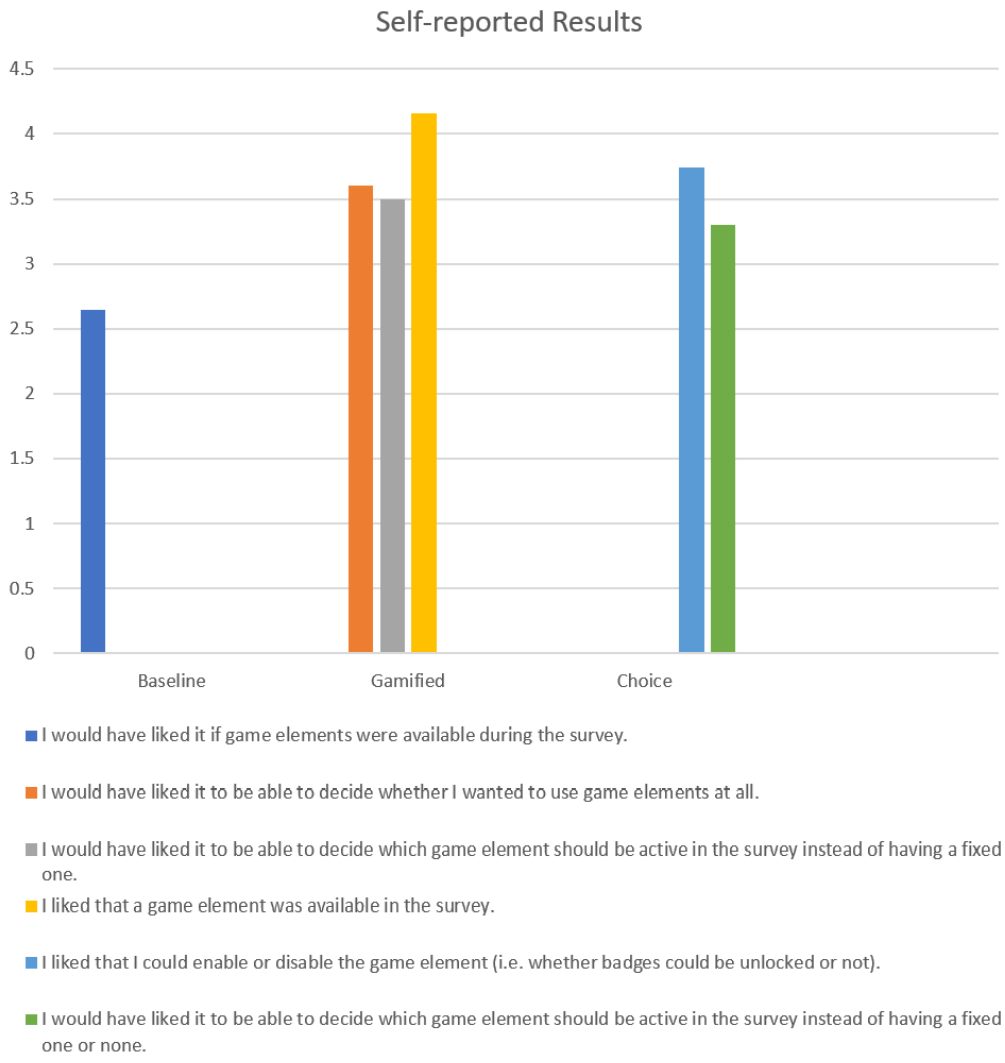


Figure 4.5: Self-reported results.

R3: Participants in the Choice condition who enabled gamification enjoyed the game elements more than participants who disabled gamification

In the survey version where a choice was given to either choose a survey where the gamification is enabled or choose a survey version where the gamification was disabled, we asked the participants to give their perception of the choice, and what made them choose the version they chose. We now take a closer look at the results obtained from answering the free-text question *"what made you to enable or disable the game element provided to you?"*. 14 out of 39 participants (35.8%) felt positive about the choice and provided positive statements about the gamification they experienced. The results show that participants felt *curious* to try out a survey with game elements, and found it *interesting*, *exciting*, and *enjoyable* to have badges in a survey. 7 out of 39 participants provided statements that are not entirely negative but some useful statements that can be used to improvise the survey versions. There were mixed opinions from participants about the gamified survey. Two participants felt that having a gamified survey with badges was good, but the questions asked were boring and not compelling enough to motivate them to answer. Four participants felt that having badges in a survey were not useful and that it will not change the way users feel while answering surveys. One participant felt that they liked to know the status of their progress (33% achieved, 66% achieved) in the survey, but did not have the motivation to achieve any badges. 5 participants out of 39 participants were in favor of the non-gamified version of the survey. The participants used the words "*simple design, simplicity, simpler* to describe why they liked the non-gamified survey version. 10 out of 39 participants did not respond to the open-question and 3 participants did not provide meaningful statements (example: bjh).

Subscales	p	p (BA - GA)	p (BA - CH)	p (GA - CH)
Interest/Enjoyment	0.168	0.200	0.941	0.372
Perceived Competence	0.148	0.996	0.719	0.656
Perceived Choice	0.870	0.992	0.934	0.877
Subscales	p	M (BA)	M (GA)	M (CH)
Interest/Enjoyment	-	33.2	34.0	33.2
Perceived Competence	-	29.80	29.95	28.66
Perceived Choice	-	34.4	34.2	34.9

Table 4.13: Self-reported Results - IMI Scale

Additionally, we also measured participants' perception of enjoyment, competence, and choice within the survey. The resulting scores clearly show that participants did not perceive the different subscales differently. Participants in all conditions have provided similar ratings indicating that the there were no change in the perception of the different survey versions.

4.6 Discussion

First of all, we were able to derive at the same findings as Harms et al. [15]. The gamification intervention of badges did not negatively affect participants' motivation and performance, in the sense that they provided approximately same amount of responses as with the participants' who responded without the game elements. Furthermore, gamification nor choice motivated participants to generate more number of characters in

free-text fields or spend more time on the survey responses.

We found that the offered choice was appreciated more by the participants who were in the Choice condition and used the choice to enable gamification (SR2). Participants have reported that this choice increased their curiosity to try out the new version consisting of game elements, rather than going on for a survey with no game elements. Thus, having a choice to enable or disable gamification had a positive effect on those who decided to use gamification. This choice has allowed them to view both versions of the survey, and make a decision because of their interest and curiosity.

Similarly, having a choice to choose between fixed gamified interventions has also proven to have a positive effect. While we believe that the concept of choice to be useful and profitable when it aligns with the survey context, there are a few aspects that has to be considered: Gamification did not seem to provide positive effects such as no effect on the amount of optional questions answered, no change in the amount of characters generated in free-text fields, no change in the time taken to complete the survey. Eventhough, the performance of the participants' were not negatively impacted, there is a huge concern regarding the absence of positive effects.

Overall, we arrive at the conclusion that providing choice between fixed gamified interventions, and providing choice to enable or disable gamification did not have any particular benefits on the quality of the generated responses, but providing a choice is shown to enhance the psychological reactions of the participants in the survey.

Chapter 5
Conclusion

In this chapter, we give a conclusive overview of the two experiments and discuss the significance of our results. Furthermore, we discuss the limitations and provide directions for future work.

5.1 Overview

In this thesis we explored the simplest form of customization for gamified online surveys, namely enabling or disabling gamification in the survey (experiment 1), and providing a choice between fixed gamified surveys (experiment 2). Since in the previous research, having a choice was generally shown to be beneficial for user motivation in different contexts and situations, we hypothesized that offering users the choice in a gamified survey context should be beneficial as well. To investigate this further, we implemented an online survey based on the existing literature, where the bare minimum of game elements generated similar positive effects as that of a non-gamified survey. Similar to Harms et al. [15], we used the game element badges for the gamified version of the survey, and reused the same logic for awarding the badges based on certain criteria. This developed survey platform served as the basis for our user study. The participants were distributed across the six conditions: non-gamified survey, gamified survey, and the survey with a choice to enable or disable gamification versions, gamified survey with badges, gamified survey with points and leaderboard, and a survey providing choice between the two gamified surveys. Based on the results obtained, we compared the participants' performance across the conditions, using the obtained behavioral and psychological variable measurements.

5.2 Contributions

First of all, we replicated the survey platform that was developed by Harms et al. [15], and recreated both non-gamified and gamified versions of the survey platform from scratch. Additionally, we designed the gamified survey with points and leaderboard

aplying the same logic as that of badges. We were able to replicate the core findings of the previous studies, showing that gamification is overall not considered effective in online surveys, which in itself is an important contribution.

Additionally, we identified that providing users with a choice to choose their preferred gamified survey version improves participants mood and survey experience. Furthermore, based on the qualitative responses of participants, we identify that participants motivation were dependent on the context or the topic of the survey (in our case, survey on sports and leisure activities) rather than purely based on game elements that were available in the survey. Participants specifically mentioned that they liked both badges and points and leaderboard, and that the gamification drove them to be curious and made the survey experience quite interesting. Consequently, it seems that having a choice positively affects all users as they have the ability to choose the version they like, without having to do much customization.

The concept of customizing the survey platform within a single step is in itself another contribution. The first experiment serves as a customization contribution to enable or disable gamification in the survey, whereas the second experiment conducted served as a customization contribution where a choice between two fully gamified survey versions were made available. In such a case, this thesis contributed towards providing minimal custmization options to the user, yet satisfying users' need for autonomy and freedom in choosing the desired survey version.

All in all, our research indicates that offering a choice to either enable or disable gamification, and also a choice between fixed gamified surveys, poses a feasible and easy-to-realize customization option. This customization option would be beneficial when the survey questions are in itself compelling enough. Though, we did not have enough evidence to prove that choice is beneficial in the context of surveys, we provide recommendations for future work in the next sections, to successfully carry out the study.

5.3 Limitations

One of the main limitations that we identify in our study is that, the participants in the conditions: No Gamification and Gamification had no chance to know or understand what it means to have a choice to enable or disable gamification in the survey. A visual tour of the other versions could have helped participants decide whether they would have liked such an approach. This effect might have had an impact on the post-survey questions that were asked to the participants about how one perceives game element. The results of the self-reported measures on the perception of choice to enable or disable gamification shows that participants would not have appreciated such a choice, and we believe that they responded this way as it was hard for them to visualize how it would have been to have a choice.

Similarly, participants in the second experiment, where two different gamified surveys were used, one with the game element badges, and another with the game element points and leaderboard, had no chance to know how the other gamified survey version looked like, or what it is to have a choice between the game elements. This might have had an impact on how participants perceived the availability of such a choice. As we

already stated, this effect might be due to the fact that it was difficult for participants to fully understand their options. Although these assumptions were made based on the received responses, and we were unable to gain any conclusive information. However, the self-reported measures on the perception of choice were positive and results from the Choice condition show that participants liked that there was a choice available in the survey.

Another limitation is that, we chose the game elements that are known for motivating players and achievers to perform better in the survey, by answering more number of questions and with improved quality. However, we feel that the gamified setup was somewhat not that motivating for a survey context. This conclusion is made based on the responses we received from the open-questions asked in the post-survey questionnaire. Some participants did not feel the urge to collect badges, as they felt that it is truly not worth the effort, given that it a survey. This might also be due to the fact that the questionnaire used on sports and leisure activities did not arouse interest in most of the participants as they felt the questionnaire to be boring. It would have been interesting to have other game elements that perfectly aligned with the survey context. From the results, we found that participants' of player types - philanthropists and free spirts were motivated to generate more badges than participants' of other player types. As a consequence, we believe that it is possible for the participants' performance and motivation to have diminished from the discussed limitations.

5.4 Future Work

One possible direction for future work would be to investigate the effect of survey questionnaire used. Since we found from the above discussions that participants felt that the questions were not compelling enough, it is highly relevant to use a different questionnaire that is particularly enjoyable, and then analyze whether providing a choice on top of the gamification intervention has any effect on the participants motivation who already enjoy the game elements.

Furthermore, it would be interesting to investigate whether offering more than one choice in a gamified survey would affect participants positively. This allows us to understand whether participants are purely driven by their interest towards the game element offered, or whether they are not affected by any additional choices offered. To investigate this, we propose to consider different game elements for each of the survey versions, and additionally provide choices to users to choose between the many gamified survey options. This approach allows us to investigate if providing more number of options, yet with minimal customization effort from the user (a single click of a button to select which gamified survey a user wants), is beneficial or not.

Another possibility for future work, is to apply more customization options (a higher user effort), and see whether higher levels of customization options has an impact on people's perception of the gamified survey platform. However, based on the received user reactions, we believe that users' might not be willing to invest more time and effort into customizing the survey platform, as it might not be provide any benefit to the user at the end. One idea is that, if the survey platform returns the user with some type of reward, then this might enhance their mood, and might have an impact on how they

want their survey setup to be. In this case, users might be willing to customize their gamified setup.

Another direction for further investigation is that, we suggest to add one or more conditions to the study, namely one condition where participants see all available survey versions during the tutorial but are forced to complete either one of the gamified (or non-gamified) survey versions encountered. By including such conditions, it is possible to examine whether participants in the selective top-down conditions would be more interested in having choice if they were fully aware of their options and had seen the task both with and without game elements or with all types of game elements that are available.

Lastly, we propose to investigate the experiments on a larger sample number and identify which specific aspects apart from people's perception and enjoyment of game elements, would have had a moderating effect on their decision about the survey version.

Appendix A
Questionnaires

In this thesis, we use three types of questionnaires: pre-experimental questionnaire, experimental questionaire, and post-experimental questionnaire.

A.1 Pre-Experimental Questionnaire

For the pre-experimental questionnaire, we use two standarized questionnaires: one from the I-PANAS-SF to investigate the positive and negative affect states of the users before starting the survey. Another questionnaire including HEXAD user type questions, to understand which user type each of the participant belongs to.

A.1.1 International Positive and Negative Affect Schedule Short Form (I-PANAS-SF)

Each item was rated on a 5-point scale, labeled with 1=Not at all, 2=Sligtly, 3=Moderately, 4=Very, 5=Extremely.

The following words describe different feelings and sensations. Read each word and then enter the intensity of the feeling you are experiencing in the scale next to each word. You have the possibility to choose between five different levels.

- Active
- Upset
- Hostile
- Inspired
- Ashamed
- Alert

- Nervous
- Determined
- Attentive
- Afraid

A.1.2 HEXAD: A Player Type Framework for Gamification Design

Each item was rated on a 7-point scale, labeled with 1=Strongly disagree, 2=Disagree, 3=Somewhat disagree, 4=Neither agree nor disagree, 5=Somewhat agree , 6=Agree, 7=Strongly agree.

To what extend do you agree with the following statements?

- It makes me happy if I am able to help others.
- I like helping others to orient themselves in new situations.
- I like sharing my knowledge.
- The wellbeing of others is important to me.
- Interacting with others is important to me.
- I like being part of a team.
- It is important to me to feel like I am part of a community.
- I enjoy group activities.
- It is important to me to follow my own path.
- I often let my curiosity guide me.
- Opportunities for self expression are important to me.
- Being independent is important to me.
- I like overcoming obstacles.
- It is important to me to continuously improve my skills.
- I enjoy emerging victorious out of difficult circumstances.
- I like mastering difficult tasks.
- I like to provoke.
- I like to question the status quo.
- I see myself as a rebel.
- I dislike following rules.
- I like competitions where a prize can be won.
- Rewards are a great way to motivate me.
- Return of investment is important to me.
- If the reward is sufficient I will put in the effort.

A.2 Experimental Questionnaire

A.2.1 Voluntary Questionnaire on Sports and Leisure Activities

1. How do you feel in general?
   Please choose one of the following answers:
   - Very healthy
   - Somewhat healthy
   - Somewhat unhealthy
   - Very unhealthy
   - No Answer
2. And how would you rate yourself in all things, are you.
   Please choose one of the following answers:
   - Very athletic
   - rather sporty
   - rather not sporty
   - not athletic at all
   - No Answer
3. And how would you describe yourself? Are you...
   Please choose one of the following answers:
   - more sociable and likes to be around people
   - rather lonely and likes to be alone
   - No Answer
4. Today's survey is about exercise and sport. This refers to ALL the physical activities you do. This can be either in a sports club or training in your free time. But it also includes daily activities in between (cycling to school/work, walking instead of driving,...) and if you do a lot of exercise during your education/job. In this sense, are you at least physically active from time to time?
   - Yes
   - No
   - No answer
5. You will find different activities listed here, please indicate if you are physically active at least once in a while. (Options: Yes, No, No answer)
   - In my spare time I do sports to train
   - I am active in my free time, but mostly spontaneously and without a fixed training goal
   - I am physically active within the scope of school, my training or my job
   - I am active in one or more sports clubs
6. How often are you physically active?
   Please choose one of the following answers:

- several times a week
- several times per month
- approximately once a month
- less than once a month
- No Answer

7. And how would you rate your sporting and physical activities? Do you...
   Please choose one of the following answers
   - enough
   - rather too much
   - rather too little
   - No response
8. If you are physically active, where is that?
   Please select one or more items from the list.
   - at school or at my training place
   - in public spaces (streets, parks, green spaces, etc.)
   - within the framework of services provided by youth organizations or youth centers
   - elsewhere in nature
   - in a sports club
   - in public sports facilities (running tracks, swimming pools, climbing facilities, etc.)
   - in fitness centers
   - at home
   - at my workplace
9. And when you are physically active, do you prefer to be active together with other people, or do you prefer to be alone?
   Please choose one of the following answers:
   - I prefer to be active with other people
   - I prefer to travel alone
   - likes both
   - No Answer
10. There is a variety of physical activities, from playing football to things like rafting, hiking and dancing. What do you do?
11. Why are you physically active?
12. Here are a few more reasons to be physically active, please indicate whether they apply to you very much, rather already, rather not or not at all.
    - It gives me a chance to think
    - My friends push me to

- I like doing things with other people
- It is a good way for me to relax
- I like to measure myself with others
- It is healthy for me
- I want to be physically attractive
- I can gain strength for everyday life
- I have time for myself
- It is fun
- I get to know my limits
- My parents push me to
- I can show what I can do
- It helps me relax and relieve stress

13. What's your circle of friends like? Are the people in your circle of friends, all things considered...
    Please choose one of the following answers:

    - physically more active than you
    - about as physically active as you
    - physically less active than you
    - No Answer

14. Why are you not physically active?

15. Here are a few more reasons to be physically active, please indicate whether they apply to you very much, rather already, rather not or not at all.

    - Sports does more harm than benefits
    - Sports in unnecessary
    - I cannot do sport, due to health or physical reasons
    - I do not have the opportunity to do so, there is none or too little offers
    - I do not like the talk of sports and fitness
    - I cannot motivate myself
    - I do not have time for sports
    - I do not have anybody to workout or do sport with
    - I cannot afford it
    - I feel uncomfortable when others watch me
    - I am not fit enough and that frustrates me
    - I have bad experience

16. Here you will find again some statements, please indicate in each case, if you want to give them very much agree, rather agree, rather disagree or disagree at all.

    - Exercise must first and foremost be fun
    - Non-sporty people are bullied more often

- A daily gym or exercise lesson at school is important
- Offers for trendy sports are hard to find
- Opening hours are too short at most sports facilities
- Sport is primarily about performance and competition
- Sport is taken far too seriously in our society
- Exercise is first and foremost for health
- Exercise helps me to get a more beautiful body

17. Below are a few measures to motivate young people to be more active. Please indicate whether you believe that these young people can be motivated to take more exercise:

    - More physical activity in school/at work
    - TV commercials with short fitness exercises
    - Fitness stations in public areas
    - More sports activities in clubs

18. When you think of your circle of friends now, what would you say? Do most people in your circle of friends wear their outfits when they exercise...

    - highly relevant
    - rather important
    - rather not important
    - not relevant

19. Another question: How do you feel with your looks...

    - very happy
    - rather content
    - rather dissatisfied
    - very unhappy

20. Finally, we ask you to provide some statistics. What state do you live in?
21. What is your highest completed education?
22. And you are in which line of work?
23. Are you married, single, in a relationship?
24. Do you have children of your own?
25. You live:

    - at your parents' house
    - a separate apartment/house
    - in a shared apartment

26. How big is the place where you live?

    - up to 5,000 inhabitants
    - up to 50,000 inhabitants
    - over 50.000 inhabitants

Mandatory Questionnaire on Demographics

1. How old are you?
    - Below 18
    - 18 – 24
    - 24 – 34
    - 34 – 44
    - 44 and above
2. Your Gender:
    - Female
    - Male
    - Non-binary/Genderqueer
    - Gender Variant/Non-Conforming
    - Not Listed
    - Prefer Not to Answer
3. Do you play computer games? This also includes games on smartphones, tablets, consoles and handhelds (Nintendo 3Ds, Sony Vita,.).
    - Yes
    - No
    - No answer
4. Are you familiar with badges (or points and leaderboard) from computer games?
    - Yes
    - No
    - No answer

A.3 Post-Experimental Questionnaire

A.3.1 International Positive and Negative Affect Schedule Short Form (I-PANAS-SF)

Each item was rated on a 5-point scale, labeled with 1=Not at all, 2=Sligtly, 3=Moderately, 4=Very, 5=Extremely.

The following words describe different feelings and sensations. Read each word and then enter the intensity of the feeling you are experiencing in the scale next to each word. You have the possibility to choose between five different levels.

- Active
- Upset
- Hostile

- Inspired
- Ashamed
- Alert
- Nervous
- Determined
- Attentive
- Afraid

A.3.2 Intrinsic Motivation Inventory (IMI)

Each item was rated on a 7-point scale, labeled with 1=Strongly disagree, 2=Disagree, 3=Somewhat disagree, 4=Neither agree nor disagree, 5=Somewhat agree , 6=Agree, 7=Strongly agree.

1. I felt like I had to do this survey.
2. I felt like it was not my own choice to do this survey.
3. I did this survey because I wanted to.
4. I believe I had some choice about doing this survey.
5. I did not really have a choice about doing this survey.
6. I did this survey because I had no choice.
7. I did this survey because I had to.
8. I thought this survey was quite enjoyable.
9. I would describe this survey as very interesting.
10. This survey was fun to do.
11. While I was doing this survey, I was thinking about how much I enjoyed it.
12. I thought this was a boring survey.
13. This survey did not hold my attention at all.
14. I enjoyed doing this survey very much.
15. After working at this survey for awhile, I felt pretty competent.
16. I think I did pretty well at this survey, compared to others participants.
17. This was a survey that I could not do very well.
18. I was pretty skilled at doing this survey.
19. I think I am pretty good at taking this survey.
20. I am satisfied with my performance at this task.

A.3.3 Self-Report Questionnaire

Each item was rated on a 5-point scale, labeled with 1=Strongly disagree, 2=Somewhat disagree, 3=Neither agree nor disagree, 4=Somewhat agree, 5=Strongly agree.

1. Completing the questionnaire took a lot of time.
2. The questionnaire was designed differently from other online questionnaires.
3. I liked the questionnaire better than other online questionnaires.
4. I feel that the questions asked in the questionnaire were reasonable.
5. I feel that the answer options provided in the questionnaire were reasonable.

These five questions were common to all conditions (No gamification condition, Gamification conditions (both badges and points and leaderboard), and Choice conditions). However, for the participants in different individual conditions, we used an additional questionnaire to evaluate how they perceived choice.

Questions used in, Experiment 1 - Condition 1: No Gamification condition.

1. I would have liked it if game elements were available during the survey (For example, scoring points for filled out questions, comparing scores with other participants or unlocking badges for filled out questions, etc,.).

Questions used in, Experiment 1: Condition 2: Gamification condition (with badges) and Experiment 2 - Condition 1: $\text{Gamification}_{\text{Badges}}$ and $\text{Gamification}_{\text{Points\&Leaderboard}}$.

1. I would have liked it to be able to decide whether I wanted to use game elements at all (i.e. I would have liked the option to disable game elements completely).
2. I would have liked it to be able to decide which game element should be active in the survey instead of having a fixed one.
3. I liked that a game element was available in the survey.

Questions used in Experiment 1: Condition 3: $\text{Choice}_{\text{enabledisable gamification}}$.

1. I liked that I could enable or disable the game element (i.e. whether badges could be unlocked or not).
2. I would have liked it to be able to decide which game element should be active in the survey instead of having a fixed one or none.

Questions used in Experiment 2: Condition 3: $\text{Choice}_{\text{BadgesPoints\&Leaderboard}}$.

1. I liked that I could select which game element I want to use i.e. badges or points and leaderboard.
2. I would have liked it if I had other game elements other than "Badges, Points and Leaderboard" to select from.

3. I would have liked it to be able to activate both game elements (i.e. badges and points and leaderboard).

Open-ended question asked in Choice condition: enable or disable gamification.
What made you to enable or disable the game element provided to you?

Open-ended question asked in Choice condition: Badges or Points and Leaderboard.
What made you choose the particular game element (points and leaderboard OR badges)?

Appendix B
Derivation of Dependent Variables

The dependent variables for this study were derived from the following eleven papers. Refer the table below for the exact reference of the papers that are mentioned.

[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**
[12]	[6]	[15]	[14]	[24]	[23]	[17]	[31]	[38]	[29]	[39]	[4]	[18]	[2]

1. Survey Completion rate		
[1]		**Significant**

	Treatment cell				
	Text only N = 251	Decoratively visual N = 251	Functionally visual N = 252	Gamified N = 253	Total
Summary					
Completion rate (%)	94	93	94	58*	80
Average length (minutes)	14.1	15.2	15.5	15.3	15.0
Demographics					
Male (%)	44	48	50	49	48
Younger than 35 (%)	22	23	29	24	24
College graduate (%)	59	51	56	61	57
Income <$25K (%)	19	22	15***	18	18
Play games daily/weekly (%)	69	54**	61	64	62
Play games seldom/never (%)	21	27	22	24	24
Average hours online per week	20.0	20.8	36.9**	20.9	24.3

*$p < 0.001$; **$p < 0.01$; ***$p < 0.05$

	Significant

d) Respondent Behavior									
	Completion of the survey	66	0.86	0.346	60	0.83	0.376	$\chi^2(1)=0.229$	0.632
	Time spent in the survey	57	08:04	03:12	50	08:19	04:50	U=1366.0	0.713
	Words in free-text answers	57	19.96	14.874	50	20.24	18.083	U=1268.5	0.724
	Speeding	64	0.58	1.307	59	0.59	0.949	U=1714.5	0.292
	Straightlining	61	0.38	0.553	57	0.26	0.552	U=1535.0	0.182
	Empty Answers	57	2.28	8.474	50	3.16	7.614	U=1134.0	0.057

Respondent Behavior: The gamified survey's completion rate of 86% (N=57 out of 66) was only insignificantly higher than the conventional survey's completion rate of 83% (N=50 out of 60), see Table 2d.

[3]	**Not Significant**

[4]	**Significant**

7.1 Respondent Behavior and Engagement

Respondent behavior was automatically logged during use. The gamified survey had a lower *response rate* of 70% (21 out of 30 persons), as opposed to the conventional survey with a response rate of 86% (26 out of 30 persons). We also measured the amount of time spent in the survey and the question where participants cancelled the

Results indicate that gamification successfully increased the users' perceived fun, the average time spent, as well as their willingness to use and recommend the survey, without introducing a strong bias in the survey results, albeit with a lower overall response rate. This improvement in user experience is in line with related studies on

	Text-Only Survey	Visual Survey	Gamified Survey	$\chi^2(df)$
Number of completed interviews	372	324	354	
Number of screened out	100	63	116	
Break off rate	14.0% (77)	18.0% (85)	17.5% (100)	3.675(2) (n.s.)

[5]	**Not Significant**

	Text-only survey	Visual survey	Gamified survey	$\chi^2(df)$
First wave				
Number of completed interviews	372	324	354	
Breakoff rate	14.0% (77)	18.0% (85)	17.5% (100)	3.68(2) (n.s.)
Second wave				
Number of invitations	368	323	351	
Participation rate	70.7% (260)	72.1% (233)	69.5% (244)	0.56 (2) (n.s.)
Breakoff rate	4.8% (13)	5.7% (14)	10.3% (28)	7.36(2)*

*p<0.05

[6]	**Significant**

<table>
<tr>
<td></td>
<td>
In the first wave of the study no significant differences in breakoff rates were found: 14.0% in the text-only, 18.0% in the visual, and 17.5% in the gamified survey ($\chi^2(2)= 3.68$, p=0.159, see Table 1). Contrary to expectations, the breakoff rate in the second wave was the highest in the gamified survey among the three conditions: 4.8% in the text-only, 5.7% in the visual, and 10.3% in the gamified survey ($\chi^2(2)=7.36$, $p<0.05$). The finding is consistent with the results of the experiment among adults conducted by Downes-Le Guin and his colleagues (2012). Though

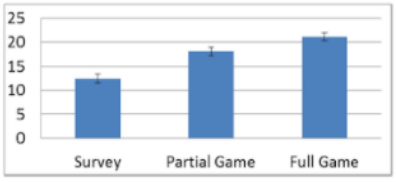
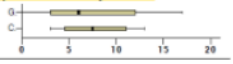
In total, 737 respondents completed the second wave with no difference between the conditions: 70.7% in the text-only, 72.1% in the visual, and 69.5% in the gamified survey ($\chi^2(2)=0.56$, p=0.756, see Table 1). A multivariate logistic regression predicting nonresponse in the second wave shows a significant effect of the score in the cognitive test. The higher the score in the first
</td>
<td></td>
</tr>
<tr>
<td>[10]</td>
<td>
<table>
<tr><th>Questionnaire type</th><th>Return rate broken off and completed feedbacks</th><th>Return rate with completed feedbacks</th><th>Return rate (1x invoice and 1x feedback from the same user</th></tr>
<tr><td>Classical</td><td>21,01%</td><td>19,49%</td><td>27,46%</td></tr>
<tr><td>Gamified</td><td>14,89%</td><td>14,22%</td><td>28,09%</td></tr>
</table>

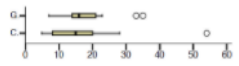
Table 5. Distribution of dropouts and completions in the classical questionnaire
<table>
<tr><th></th><th>Frequency</th><th>Percent</th></tr>
<tr><td>Valid conclusion</td><td>77</td><td>93%</td></tr>
<tr><td>Dropout</td><td>6</td><td>7%</td></tr>
<tr><td>Total</td><td>83</td><td>100%</td></tr>
</table>

Table 6. Distribution of dropouts and completions in the gamified questionnaire
<table>
<tr><th></th><th>Frequency</th><th>Percent</th></tr>
<tr><td>Valid conclusion</td><td>64</td><td>96%</td></tr>
<tr><td>Dropout</td><td>3</td><td>4%</td></tr>
<tr><td>Total</td><td>67</td><td>100%</td></tr>
</table>
</td>
<td>Significance not tested</td>
</tr>
<tr>
<td>[12]</td>
<td>
29.4.2 Drop-Off Rate

Approximately 40 % of those who were qualified to complete the survey dropped off before completion in each of the gamified versions. Because of the greater complexity of the conjoint experiment compared to a typical survey, it was expected that the drop-off rate would be higher than for a typical survey. This was indeed the case: other researchers have found a fairly consistent drop-off rate of 25 % in online surveys (Cape, 2009). Table 29.6 details the drop-off rates of the four different games.
Gamification did not significantly reduce the survey drop-off rate at the 95 % confidence level, although the sample size may not have been powerful enough to show a difference. Respondents who participated in the regular survey-style conjoint experiment were just as likely to discontinue participating in the survey as those who played the gamified versions.

Table 29.6 Drop-off rates
<table>
<tr><th></th><th>Version 1—simple choice sets</th><th>Version 2.1—salesperson—no incorrect answers</th><th>Version 2.2—salesperson—with incorrect answers</th><th>Version 3—monsters—with incorrect answers</th></tr>
<tr><td>Total completes</td><td>104</td><td>90</td><td>102</td><td>101</td></tr>
<tr><td>Total drop-offs</td><td>128</td><td>58</td><td>68</td><td>63</td></tr>
<tr><td>Total qualified to complete</td><td>232</td><td>148</td><td>170</td><td>164</td></tr>
<tr><td>Percentage of drop-offs over the total who started</td><td>55 %</td><td>39 %</td><td>40 %</td><td>38 %</td></tr>
</table>
</td>
<td>Not significant</td>
</tr>
</table>

2. Survey Participation		
[2]	*Did Increasing Gamification Increase Participation?* A univariate ANOVA test showed a main effect of *gamification level* on total questions completed ($F_{2,641}$=23.9, p<0.001). Tukey HSD post-hoc pairwise comparisons of the three levels revealed a significant difference between Full Game and Survey (p<0.001), Survey and Partial Game (p<0.001), and Full Game and Partial Game (p=0.031). The average number of questions completed increased with gamification level (Figure 2).  **Figure 2. Mean questions completed by level of gamification.**	**Significant**
[4]	b.) Amongst respondents who cancelled the survey: After how many questions did they cancel? Gamified 9 7.56 6.00 6.54 Conventional 4 7.75 7.50 4.27 U=110 p=0.956 Total 13 7.62 6.00 5.75 	**Not Significant**
[5]	Contrary to expectations, the highest item nonresponse rates were found in the gamified (11.73%) and visual surveys (11.12%) than in the text-only survey (1.73%; see Table 2). A linear multivariate regression showed significantly lower item nonresponse rate in the text-only ($\beta = -8.924$, $p < .001$) compared to the gamified survey. In addition, those who filled out the questionnaire via PC rather **Table 2.** Overall Item Nonresponse Rate. Text-Only Survey | Visual Survey | Gamified Survey | Statistics *N* 372 324 354 Overall item nonresponse rate 1.73% (4.43%) 11.12% (16.33%) 11.73% (16.32%) $F(2, 1047) = 63.421^{***}$ Excluding Flash-based questions Overall item nonresponse rate 1.89% (5.72%) 2.06% (7.01%) 2.11% (5.64%) $F(2, 1047) = 0.123$; (n.s.) $^{***}p < .001$, standard deviation in parentheses.	**Significant**
[6]	*Item nonresponse rate* Item nonresponse rate – *first wave* 1.73% (4.43%) 11.12% (16.33%) 11.73% (16.32%) F(2,1047)=63.42*** Item nonresponse rate – *second wave* 1.29% (3.35%) 1.29% (5.07%) 1.23% (3.59%) F(2,734)=0.02 (n.s.)	**Not Significant**
[13]	It has been argued that the use of gamification in research helps increase completion rate and respondents' engagement (e.g. Leedy & Ruyle, 2011; Puleston, 2011b). However, in this study, H3 is rejected and H4 is only partly confirmed, casting doubt on the validity of previous research. Gamification has not contributed to increases in completion rate ($p \leq .950$); both questionnaires had similar completion rates. Gamification's effects on	**Significant**

3. Words in free-text answers		
[2]	Responses in this game are given as free-form text, making direct comparisons more difficult. For our initial analysis we compared the average response length in characters, between the three gamification levels. Univariate ANOVA found a significant difference in average response length for gamification level ($F_{2,410}$=8.723, p<0.001). A Tukey HSD post-hoc pairwise comparison found a significant difference between Partial Game and Full Game conditions (p<0.001, with responses 1.79 characters shorter in the Full Game) but no significant difference between Survey and Partial Game (p=0.069) or Survey and Full Game (p=0.221).	**Significant**
[3]		**Not Significant**
[4]	d.) Amongst respondents who completed the survey: Did gamification increase the word counts of the plain text answers? Gamified 21 17.76 16.00 6.71 Conventional 26 15.46 15.00 10.40 U=115 Total 47 16.49 15.00 8.92 p=0.120 	**Not Significant**
[5]	Contrary to expectations, no difference in the length of the responses, the number of reasons children reported for liking school, and the distribution of these reasons was found between the conditions. On average, respondents wrote 49.9 characters in the text only (*SD* = 58.1), 43.6 characters in the visual (*SD* = 57.6), and 47.7 characters (*SD* = 64.0) in the gamified survey, $F(2, 987) = 0.901$, $p = .407$.	**Not Significant**
[10]	**Table 3.** Evaluation of the number of characters in the classical and gamified questionnaire **Questionnaire type** / **Number of feedbacks** / **Number of text answers** / **Text reply quote** / **Average of used characters** / **Median on used characters** / **Max** / **Min** Classical 77 39 57,53% 15,23 9 89 1 Gamified 64 35 54,69% 44,53 25 626 2	**Significance not tested**
[13]	Many authors (e.g. Puleston & Sleep, 2011; Stevens, 2011; Strube & Zdanowicz, 2008) have suggested that gamification helps increase the volume of data collected in online surveys and the time respondents spend answering questions. In this study, H1 and H2 are both supported as there was a significant difference in volume of feedback ($p \leq .016$) and time spent on completion activity ($p \leq .000$) between the conventional questionnaire and the gamified questionnaire. Scores of both variables were higher in the gamified survey (16% increase in word count and 30% increase in time spent).	**Significant**
[14]	**Findings** The results of the study in relation to our first objective clearly demonstrated that the application of soft gamification did indeed provide richer data (i.e. significantly greater number of words for each of the three questions) via the use of a framework, applying a rule (supporting the previous findings of Jon Puleston and Deborah Sleep (2011)) and also subtly changing the response layout for open-ended questions (see Table 1).	**Significance not tested**

<table>
<tr><th colspan="3">4. Degree of careless responses</th></tr>
<tr><th colspan="3">a) Straight-lining</th></tr>
<tr><td>[1]</td><td>

	Text only N = 251	Decoratively visual N = 251	Functionally visual N = 252	Gamified N = 253	Total
Debrief questions					
How interesting?[a]	5.2	5.4	5.7***	6.0***	5.6
How easy to read?[a]	6.1	6.3	6.4***	6.2***	6.2
How easy to answer?[a]	5.9	6.0	6.3***	6.3***	6.1
How fast?[a]	5.3	5.3	5.4***	5.0***	5.3
How much enjoyed?[a]	5.0	5.3***	5.4***	5.7***	5.4
How many minutes?	12.7	13.0	13.8***	15.1**	13.7
Engagement trap failures					
Inconsistent response (%)	18	22	22	20	20
Failed to select 'strongly agree' (%)	13	12	10	13	12
Straightlined in two or more grids (%)	17	20	17	20	18

</td><td>Not Significant</td></tr>
<tr><td>[3]</td><td>

d) Respondent Behavior								
	Completion of the survey	66	0.86	0.346	60	0.83	0.376	$\chi^2(1)=0.229$	0.632
	Time spent in the survey	57	08:04	03:12	50	08:19	04:50	U=1366.0	0.713
	Words in free-text answers	57	19.96	14.874	50	20.24	18.083	U=1268.5	0.724
	Speeding	64	0.58	1.307	59	0.59	0.949	U=1714.5	0.292
	Straightlining	61	0.38	0.553	57	0.26	0.552	U=1535.0	0.182
	Empty Answers	57	2.28	8.474	50	3.16	7.614	U=1134.0	0.057

We investigated the influence of gamification on the answers given by respondents using separate tests for each individual survey question (Mann-Whitney U-tests for ordinal questions, Chi-Square tests for boolean questions, Dunn-Sidak adjusted p-Values). None of the questions revealed a significant influence, the smallest adjusted p-Value being p=0.861.

</td><td>Not Significant</td></tr>
<tr><td>[5]</td><td>

Table 3. Straight-Lining, Extreme Response Style, and Middle-Point Response Style (Odds Ratio).

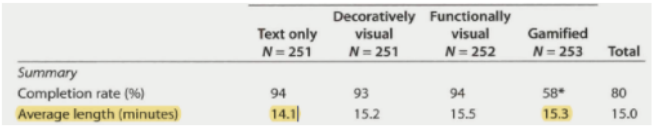
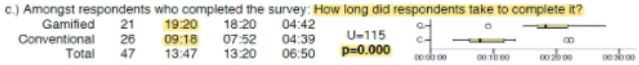
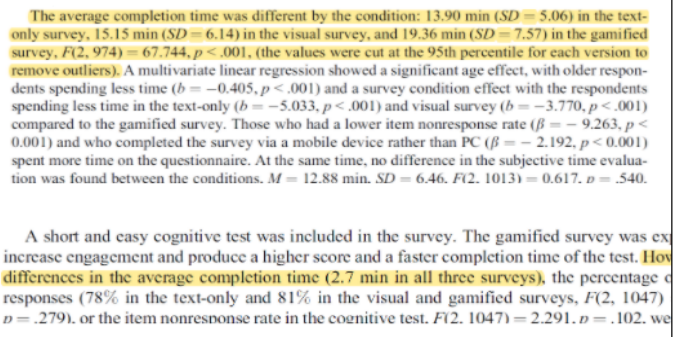
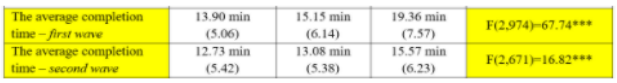
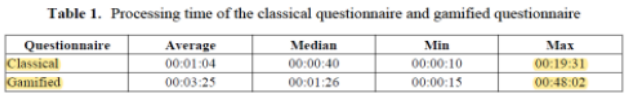
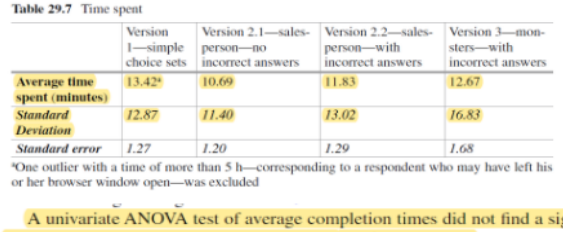
	Straight-Lining (A Logistic Regression)	Extreme Response Style (A Negative Binomial Model)	Middle-Point Response Style (A Negative Binomial Model)
Intercept	0.303 [0.041, 2.232]	13.075 [7.805, 21.906]***	1.541 [0.878, 2.703]
Text-only survey	3.840 [1.762, 8.370]***	1.094 [0.919, 1.303]	1.223 [1.014, 1.475]*
Visual survey	0.840 [0.300, 2.347]	0.980 [0.812, 1.182]	0.971 [0.792, 1.191]
Age	0.876 [0.760, 1.010]	0.960 [0.927, 0.993]*	1.079 [1.040, 1.120]***
Girls	1.060 [0.580, 1.934]	1.011 [0.873, 1.170]	0.987 [0.842, 1.156]
Good academic performance	2.049 [1.087, 3.860]*	1.287 [1.110, 1.494]***	0.711 [0.606, 0.834]***
Cognitive test (percentage of correct responses)	0.118 [0.048, 0.295]***	0.885 [0.656, 1.194]	1.278 [0.937, 1.741]
Using the Internet every day or almost every day	1.672 [0.771, 3.622]	1.071 [0.885, 1.297]	0.810 [0.659, 0.996]*
Using the Internet 7 or more hours per day	1.146 [0.306, 4.293]	1.077 [0.810, 1.433]	0.978 [0.722, 1.327]
Mobile web survey mode	0.211 [0.027, 1.669]	1.065 [0.741, 1.530]	0.910 [0.618, 1.340]
Presence of bystanders	2.121 [1.057, 4.256]*	0.932 [0.779, 1.116]	1.095 [0.903, 1.327]
Getting help while completing the survey	0.500 [0.209, 1.192]	0.966 [0.768, 1.215]	1.023 [0.805, 1.302]
Partially labeled scale +	0.690 [0.379, 1.256]	0.957 [0.829, 1.105]	0.983 [0.842, 1.147]
Model	$\chi^2(11) = 57.119^{***}$, R^2 (Cox & Snell) = .065, R^2 (Nagelkerke) = .175	Likelihood ratio $\chi^2(12) = 20.743$, $p = .054$	Likelihood ratio $\chi^2(12) = 45.946^{***}$

Note. Reference category = gamified survey; + = the experiment with partially vs. fully labeled responses was implemented in Q15.
*p < .05. **p < .01. ***p < .001, 95% CI in brackets.

As expected, the text-only condition produced a higher level of straight-lining: 11.4% in the text-only survey, 2.8% in the visual survey, and 3.2% in the gamified survey. No difference between the visual and gamified surveys was found. This is rather the difference between the types of questions—grids and drag and drop as well as grids and slider questions—than the difference between the survey conditions (see Drolet, Butler, & Davis, 2009). However, there was also a higher item

</td><td>Significant</td></tr>
</table>

<table>
<tr><td>[6]</td><td>
<table>
<tr><td>Straight-lining</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Straight-lining – first wave</td><td>11.4%</td><td>2.8%</td><td>3.2%</td><td>$\chi^2(2)$ 25.24***</td></tr>
<tr><td>Straight-lining – second wave</td><td>15.4%</td><td>9.2%</td><td>14.5%</td><td>$\chi^2(2)$ 4.76 (n.s.)</td></tr>
</table>
***p <0 .001, *p<0.05, standard deviation in parentheses.
</td><td>Not Significant</td></tr>
<tr><td>[8]</td><td>
Schonlau and Toepoel, 2015) to assess survey data quality. In order to assess data quality within our lab experiment, we, operationalize data quality by considering the indicators listed in Table 1.
<table>
<tr><th colspan="2">Indicators for data quality</th></tr>
<tr><td>1.</td><td>Voluntary continuing with the web survey
• Number of additional items answered in voluntary part of the questionnaire
• Additional time spent on voluntary part of the questionnaire</td></tr>
<tr><td>2.</td><td>Item nonresponse (Mavletova 2015)</td></tr>
<tr><td>3.</td><td>Length of answers to open questions (Mavletova 2015)</td></tr>
<tr><td>4.</td><td>Rounded answers to numeric questions (Schober et al. 2015)</td></tr>
<tr><td>5.</td><td>Straight-lining (Schonlau and Toepoel 2015)</td></tr>
<tr><td>6.</td><td>Speeding (Zhang and Conrad 2013)</td></tr>
<tr><td>7.</td><td>Bogus item flags (Meade and Craig 2012)</td></tr>
<tr><td>8.</td><td>Consistency items (Meade and Craig 2012)</td></tr>
</table>
</td><td>Not measured</td></tr>
<tr><td colspan="3">b) Speeding</td></tr>
<tr><td>[3]</td><td>
<table>
<tr><td>d) Respondent Behavior</td><td>Completion of the survey</td><td>66</td><td>0.86</td><td>0.346</td><td>60</td><td>0.83</td><td>0.376</td><td>$\chi^2(1)=0.229$</td><td>0.632</td></tr>
<tr><td></td><td>Time spent in the survey</td><td>57</td><td>08:04</td><td>03:12</td><td>50</td><td>08:19</td><td>04:50</td><td>U=1366.0</td><td>0.713</td></tr>
<tr><td></td><td>Words in free-text answers</td><td>57</td><td>19.96</td><td>14.874</td><td>50</td><td>20.24</td><td>18.083</td><td>U=1268.5</td><td>0.724</td></tr>
<tr><td></td><td>Speeding</td><td>64</td><td>0.58</td><td>1.307</td><td>59</td><td>0.59</td><td>0.949</td><td>U=1714.5</td><td>0.292</td></tr>
<tr><td></td><td>Straightlining</td><td>61</td><td>0.38</td><td>0.553</td><td>57</td><td>0.26</td><td>0.552</td><td>U=1535.0</td><td>0.182</td></tr>
<tr><td></td><td>Empty Answers</td><td>57</td><td>2.28</td><td>8.474</td><td>50</td><td>3.16</td><td>7.614</td><td>U=1134.0</td><td>0.057</td></tr>
</table>
</td><td>Not Significant</td></tr>
<tr><td>[8]</td><td>
Schonlau and Toepoel, 2015) to assess survey data quality. In order to assess data quality within our lab experiment, we, operationalize data quality by considering the indicators listed in Table 1.
<table>
<tr><th colspan="2">Indicators for data quality</th></tr>
<tr><td>1.</td><td>Voluntary continuing with the web survey
• Number of additional items answered in voluntary part of the questionnaire
• Additional time spent on voluntary part of the questionnaire</td></tr>
<tr><td>2.</td><td>Item nonresponse (Mavletova 2015)</td></tr>
<tr><td>3.</td><td>Length of answers to open questions (Mavletova 2015)</td></tr>
<tr><td>4.</td><td>Rounded answers to numeric questions (Schober et al. 2015)</td></tr>
<tr><td>5.</td><td>Straight-lining (Schonlau and Toepoel 2015)</td></tr>
<tr><td>6.</td><td>Speeding (Zhang and Conrad 2013)</td></tr>
<tr><td>7.</td><td>Bogus item flags (Meade and Craig 2012)</td></tr>
<tr><td>8.</td><td>Consistency items (Meade and Craig 2012)</td></tr>
</table>
</td><td>Not measured</td></tr>
<tr><td colspan="3">c) Middle-point responses</td></tr>
<tr><td>[5]</td><td>
About 14.0% of the responses were given using the middle-point category. A negative binomial model showed that the text-only survey produced a higher rate of middle responses compared to the gamified survey (OR = 1.223, $p < .05$; see Table 3). No difference between the gamified and visual surveys was found. Younger respondents (OR = 1.079, $p < .001$), those with higher academic
</td><td>Significant</td></tr>
<tr><td colspan="3">d) Don’t know answers</td></tr>
<tr><td>[3]</td><td>
<table>
<tr><td>d) Respondent Behavior</td><td>Completion of the survey</td><td>66</td><td>0.86</td><td>0.346</td><td>60</td><td>0.83</td><td>0.376</td><td>$\chi^2(1)=0.229$</td><td>0.632</td></tr>
<tr><td></td><td>Time spent in the survey</td><td>57</td><td>08:04</td><td>03:12</td><td>50</td><td>08:19</td><td>04:50</td><td>U=1366.0</td><td>0.713</td></tr>
<tr><td></td><td>Words in free-text answers</td><td>57</td><td>19.96</td><td>14.874</td><td>50</td><td>20.24</td><td>18.083</td><td>U=1268.5</td><td>0.724</td></tr>
<tr><td></td><td>Speeding</td><td>64</td><td>0.58</td><td>1.307</td><td>59</td><td>0.59</td><td>0.949</td><td>U=1714.5</td><td>0.292</td></tr>
<tr><td></td><td>Straightlining</td><td>61</td><td>0.38</td><td>0.553</td><td>57</td><td>0.26</td><td>0.552</td><td>U=1535.0</td><td>0.182</td></tr>
<tr><td></td><td>Empty Answers</td><td>57</td><td>2.28</td><td>8.474</td><td>50</td><td>3.16</td><td>7.614</td><td>U=1134.0</td><td>0.057</td></tr>
</table>
</td><td>Not Significant</td></tr>
</table>

5. Survey Completion time		
[1]		**Not Significant**
[3]		**Not Significant**
[4]		**Significant**
[5]		**Not Significant**
[6]		**Significant**
[10]		**Significance not tested**
[12]		**Not significant**

[1]

	Text only N = 251	Decoratively visual N = 251	Functionally visual N = 252	Gamified N = 253	Total
Summary					
Completion rate (%)	94	93	94	58*	80
Average length (minutes)	14.1	15.2	15.5	15.3	15.0

[3]

d) Respondent Behavior								
	Completion of the survey	66	0.86	0.346	60	0.83	0.376	$\chi^2(1)=0.229$	0.632
	Time spent in the survey	57	08:04	03:12	50	08:19	04:50	U=1366.0	0.713
	Words in free-text answers	57	19.96	14.874	50	20.24	18.083	U=1268.5	0.724
	Speeding	64	0.58	1.307	59	0.59	0.949	U=1714.5	0.292
	Straightlining	61	0.38	0.553	57	0.26	0.552	U=1535.0	0.182
	Empty Answers	57	2.28	8.474	50	3.16	7.614	U=1134.0	0.057

[4]

c.) Amongst respondents who completed the survey: How long did respondents take to complete it?

Gamified	21	19:20	18:20	04:42	U=115	
Conventional	26	09:18	07:52	04:39	p=0.000	
Total	47	13:47	13:20	06:50		

[5]

The average completion time was different by the condition: 13.90 min (SD = 5.06) in the text-only survey, 15.15 min (SD = 6.14) in the visual survey, and 19.36 min (SD = 7.57) in the gamified survey, $F(2, 974) = 67.744$, $p < .001$, (the values were cut at the 95th percentile for each version to remove outliers). A multivariate linear regression showed a significant age effect, with older respondents spending less time ($b = -0.405$, $p < .001$) and a survey condition effect with the respondents spending less time in the text-only ($b = -5.033$, $p < .001$) and visual survey ($b = -3.770$, $p < .001$) compared to the gamified survey. Those who had a lower item nonresponse rate ($\beta = -9.263$, $p < 0.001$) and who completed the survey via a mobile device rather than PC ($\beta = -2.192$, $p < 0.001$) spent more time on the questionnaire. At the same time, no difference in the subjective time evaluation was found between the conditions. $M = 12.88$ min, $SD = 6.46$, $F(2, 1013) = 0.617$, $p = .540$.

A short and easy cognitive test was included in the survey. The gamified survey was ex
increase engagement and produce a higher score and a faster completion time of the test. Hov
differences in the average completion time (2.7 min in all three surveys), the percentage c
responses (78% in the text-only and 81% in the visual and gamified surveys, $F(2, 1047)$
$p = .279$), or the item nonresponse rate in the cognitive test. $F(2, 1047) = 2.291$, $p = .102$, we

[6]

The average completion time – *first wave*	13.90 min (5.06)	15.15 min (6.14)	19.36 min (7.57)	F(2,974)=67.74***
The average completion time – *second wave*	12.73 min (5.42)	13.08 min (5.38)	15.57 min (6.23)	F(2,671)=16.82***

[10]

Table 1. Processing time of the classical questionnaire and gamified questionnaire

Questionnaire	Average	Median	Min	Max
Classical	00:01:04	00:00:40	00:00:10	00:19:31
Gamified	00:03:25	00:01:26	00:00:15	00:48:02

[12]

Table 29.7 Time spent

	Version 1—simple choice sets	Version 2.1—salesperson—no incorrect answers	Version 2.2—salesperson—with incorrect answers	Version 3—monsters—with incorrect answers
Average time spent (minutes)	13.42[a]	10.69	11.83	12.67
Standard Deviation	*12.87*	*11.40*	*13.02*	*16.83*
Standard error	*1.27*	*1.20*	*1.29*	*1.68*

[a]One outlier with a time of more than 5 h—corresponding to a respondent who may have left his or her browser window open—was excluded

A univariate ANOVA test of average completion times did not find a significant difference between the four versions ($F_{3,390}=0.517$, $p=0.67$). The realism of the

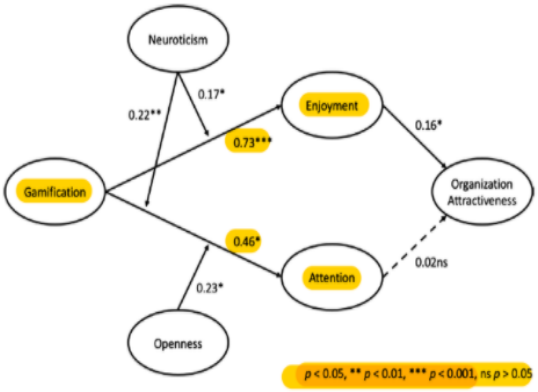
<table>
<tr><td>[13]</td><td>Many authors (e.g. Puleston & Sleep, 2011; Stevens, 2011; Strube & Zdanowicz, 2008) have suggested that gamification helps increase the volume of data collected in online surveys and the time respondents spend answering questions. In this study, H1 and H2 are both supported as there was a significant difference in volume of feedback ($p \leq .016$) and time spent on completion activity ($p \leq .000$) between the conventional questionnaire and the gamified questionnaire. Scores of both variables were higher in the gamified survey (16% increase in word count and 30% increase in time spent).</td><td>Significant</td></tr>
<tr><td>[14]</td><td>Comparison of completion times
Although contrary to the previous research of Puleston and Sleep (2011), the increased detail of responses in the gamified survey cell did not significantly impact the time taken to complete the survey. Earlier research had observed participants spending far longer completing gamified surveys, which was interpreted as a sign that participants were more engaged in the survey as they were willing to spend extra time providing the detail. However, in the case of this research the questionnaire was relatively short (compared to many online surveys) and took an average of only ten minutes to complete in both test cells.</td><td>Not Significant</td></tr>
<tr><td colspan="3">6. Number of choice sets completed</td></tr>
<tr><td>[12]</td><td>To our surprise, the game elements did not have a positive effect on the number of choice sets that respondents completed. Although a univariate ANOVA test did not show a significant difference ($F_{3,390}=1.17$, $p=0.322$), respondents who received the survey-like version completed an average of 10 more choice sets than in any of the game versions, completely contrary to expectations.
<table>
<tr><th></th><th>Version 1—simple choice sets</th><th>Version 2.1—— salesperson—no incorrect answers</th><th>Version 2.2— salesperson—with incorrect answers</th><th>Version 3—monsters— with incorrect answers</th></tr>
<tr><td>Average number of choice sets completed</td><td>47</td><td>37</td><td>36</td><td>37</td></tr>
<tr><td>Standard deviation</td><td>52.9</td><td>48.2</td><td>39.6</td><td>45.8</td></tr>
<tr><td>Standard error</td><td>5.08</td><td>4.73</td><td>4.32</td><td>4.21</td></tr>
</table>
<table>
<tr><th></th><th>Version 1—simple choice sets</th><th>Version 2.1— salesperson—no incorrect answers</th><th>Version 2.2— salesperson—with incorrect answers</th><th>Version 3—monsters— with incorrect answers</th></tr>
<tr><td>Percentage who did at least 36 choice tasks</td><td>41 %</td><td>24 %</td><td>33 %</td><td>30 %</td></tr>
<tr><td>Percentage who completed all 192 choice tasks</td><td>7 %</td><td>3 %</td><td>0 %</td><td>3 %</td></tr>
<tr><td>Percentage who did only the minimum (6 choice tasks)</td><td>21 %</td><td>29 %</td><td>29 %</td><td>29 %</td></tr>
</table></td><td>Not significant</td></tr>
</table>

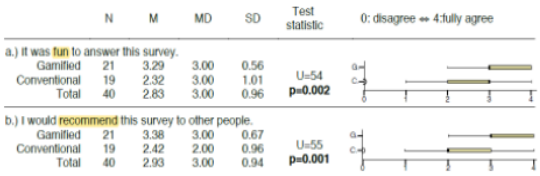
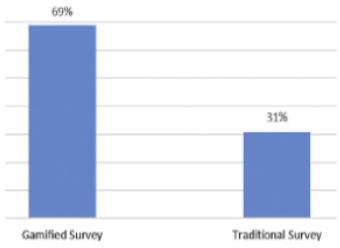
1. User Experience		
[3]	b) User Experience / AttrakDiff2 [11]: Pragmatic Quality (48, 1.38, 0.634, 42, 1.31, 0.735, U=948.5, 0.851); Hedonic Quality - Identity (43, 0.74, 0.721, 38, 0.72, 0.749, U=787.0, 0.779); Hedonic Quality - Stimulation (48, 0.90, 1.000, 41, 0.23, 0.885, U=536.5, <0.001); Attractiveness (48, 1.43, 1.429, 41, 1.09, 0.852, U=746.0, 0.049) scores for "hedonic quality - stimulation" and "attractiveness" were significantly higher (i.e., better) in the gamified survey.	**Significant**
2. Affect		
[3]	**Affect:** The participant's affect before and after filling the survey was measured using I-PANAS-SF [18] questions in the pre- and post-test questionnaires. Affect scores ranged from 0 to 50, the higher the stronger the emotion. Differential affect scores were calculated as post-test minus pre-test scores. None of the scores differed significantly depending on survey design, see Table 2a. a) Affect / I-PANAS-SF [18]: Pre-Test Positive Affect (58, 14.66, 3.354, 56, 14.70, 3.264, U=1581.0, 0.808); Pre-Test Negative Affect (60, 6.15, 1.858, 58, 6.29, 2.392, U=1728.0, 0.945); Post-Test Positive Affect (55, 14.42, 3.521, 47, 14.49, 3.406, U=1257.0, 0.813); Post-Test Negative Affect (55, 5.67, 1.001, 48, 5.73, 1.976, U=1163.5, 0.220); Differential Positive Affect (53, -0.45, 2.081, 44, -0.50, 2.029, U=1127.5, 0.779); Differential Negative Affect (55, -0.29, 1.133, 46, -0.22, 0.696, U=1238.5, 0.827)	**Not Significant**
3. Perceived fun/Enjoyment		
[1]	Debrief questions: How interesting?[a] 5.2 / 5.4 / 5.7*** / 6.0*** / 5.6; How easy to read?[a] 6.1 / 6.3 / 6.4*** / 6.2*** / 6.2; How easy to answer?[a] 5.9 / 6.0 / 6.3*** / 6.3*** / 6.1; How fast?[a] 5.3 / 5.3 / 5.4*** / 5.0*** / 5.3; How much enjoyed?[a] 5.0 / 5.3*** / 5.4*** / 5.7*** / 5.4; How many minutes? 12.7 / 13.0 / 13.8*** / 15.1** / 13.7	**Significant**
[3]	c) Subjective Ratings / 4-item Likert-type questions: Fun (55, 2.98, 0.828, 48, 2.83, 0.975, U=1233.5, 0.545); Time consuming (56, 1.66, 0.837, 48, 1.69, 0.689, U=1261.5, 0.557); Preferred over other surveys (53, 3.09, 0.714, 42, 2.64, 0.958, U=821.5, 0.019)	**Not Significant**
[4]	a.) It was fun to answer this survey. Gamified (21, 3.29, 3.00, 0.56); Conventional (19, 2.32, 3.00, 1.01); Total (40, 2.83, 3.00, 0.96); U=54, p=0.002 b.) I would recommend this survey to other people. Gamified (21, 3.38, 3.00, 0.67); Conventional (19, 2.42, 2.00, 0.96); Total (40, 2.93, 3.00, 0.94); U=55, p=0.001 (N, M, MD, SD, Test statistic; 0: disagree ↔ 4:fully agree)	**Significant**

<table>
<tr>
<td>[5]</td>
<td>More children found it more enjoyable to fill out the gamified questionnaire, $\chi^2(8) = 42.412$, $p < .001$: 28.6% selected the top response on the 5-point scale (“I liked it very much”) in the text-only survey, 33.8% in the visual survey, and 48.1% in the gamified survey. A multivariate logistic regression predicting the top response showed that it was less enjoyable to complete the text-only (OR = 0.434, $p < .001$) and visual survey (OR = 0.569, $p < .01$). Younger respondents (OR = 0.923, $p < .05$), girls (OR = 1.369, $p < .01$), those who performed well at school (OR = 1.389, $p < .05$) and who spent on average 7 or more hours on the Internet per day (OR = 2.028, $p < .01$) provided more positive evaluation. Predicting the top two responses (“liked it very much” and “rather liked it”) showed a significant difference between the text-only (OR = 0.539, $p < .01$) and gamified survey and no significant difference between the visual (OR = 0.852, $p = .502$) and gamified survey.</td>
<td>Significant</td>
</tr>
<tr>
<td>[6]</td>
<td>
<table>
<tr><td>Enjoyment of completing the survey – first wave
Scale: 1-5(I liked it very much-I did not like it at all)</td><td>2.05
(0.95)</td><td>1.86
(0.79)</td><td>1.70
(0.82)</td><td>F(2,1036)=15.58***</td></tr>
<tr><td>Enjoyment of completing the survey – second wave</td><td>1.91
(0.95)</td><td>1.89
(0.82)</td><td>1.76
(0.79)</td><td>F(2,728)=2.45
(n.s.)</td></tr>
</table>
</td>
<td>Not Significant</td>
</tr>
<tr>
<td>[9]</td>
<td>To test the effect of gamification on cognitive reactions (H1), we performed a paired samples t-test on the rescaled factor scores of cognitive reactions towards the game ($M = 3.349$, $SD = 0.954$) and cognitive reactions towards the survey ($M = 3.291$, $SD = 0.969$). The mean difference is not significant ($M_{diff} = 0.058$, $p > 0.05$), leading us to reject H1. Thus, gamification seems not to affect cognitive reactions. To test the effect of gamification on affective reactions (H2), we performed a paired samples t-test on the rescaled factor scores of affective reactions towards the game ($M = 3.409$, $SD = 0.974$) and affective reactions towards the survey ($M = 3.251$, $SD = 0.938$). The mean difference is significant ($M_{diff} = 0.159$, $p < 0.001$), providing support for H2. In other words, gamification appears to increase affective reactions. To test hypotheses H3a/b through H6a/b, we evaluated the corresponding path coefficients in the path model as shown in Fig. 5.

As can be seen in Fig. 5, the path from attitude towards the extrinsic mechanism to cognitive reactions towards the game is not significant ($\beta = 0.055$, $SE = 0.057$, $p > 0.05$). Contrary to expectations, it appears that attitude towards the extrinsic mechanism does not impact cognitive reactions towards the game. Hence, H3a is not supported. On the other hand, the path from attitude towards the extrinsic mechanism to affective reactions towards the game is positive and significant ($\beta = 0.118$, $SE = 0.056$, $p < 0.05$), which suggests that attitude towards the extrinsic mechanism increases affective reactions. Thus, H3b is supported. In contrast, the paths from attitude towards the intrinsic mechanism to cognitive reactions ($\beta = 0.214$, $SE = 0.055$, $p < 0.001$), and affective reactions ($\beta = 0.258$, $SE = 0.054$, $p < 0.001$), towards the game are both positive and significant. Thus, attitude towards the intrinsic mechanism increases cognitive and affective reactions. Likewise, the paths from attitude towards the self-presentation mechanism to cognitive reactions ($\beta = 0.134$, $SE = 0.045$, $p < 0.01$) and affective reactions ($\beta = 0.116$, $SE = 0.044$, $p < 0.01$) are both positive and significant. This suggest that attitude towards the self-pre-

Table 6
The results of hypotheses testing.
<table>
<tr><th>Hypothesis</th><th>Statistical Test</th><th>Supported</th></tr>
<tr><td>H1: Gamified survey > cognitive reactions</td><td>$M_{diff} = 0.058$, $p > 0.05$</td><td>No</td></tr>
<tr><td>H2: Gamified survey > affective reactions</td><td>$M_{diff} = 0.159$, $p < 0.001$</td><td>Yes</td></tr>
<tr><td>H3a: Extrinsic design mechanism > cognitive reactions</td><td>$\beta = 0.055$, $SE = 0.057$, $p > 0.05$</td><td>No</td></tr>
<tr><td>H3b: Extrinsic design mechanism > affective reactions</td><td>$\beta = 0.118$, $SE = 0.056$, $p < 0.05$</td><td>Yes</td></tr>
<tr><td>H4a: Intrinsic design mechanism > cognitive reactions</td><td>$\beta = 0.214$, $SE = 0.055$, $p < 0.001$</td><td>Yes</td></tr>
<tr><td>H4b: Intrinsic design mechanism > affective reactions</td><td>$\beta = 0.258$, $SE = 0.054$, $p < 0.001$</td><td>Yes</td></tr>
<tr><td>H5a: Self-presentation mechanism > cognitive reactions</td><td>$\beta = 0.134$, $SE = 0.045$, $p < 0.01$</td><td>Yes</td></tr>
<tr><td>H5b: Self-presentation mechanism > affective reactions</td><td>$\beta = 0.116$, $SE = 0.044$, $p < 0.01$</td><td>Yes</td></tr>
<tr><td>H6a: Cognitive reactions > signaler</td><td>$\Delta\beta = -0.065$, $p > 0.05$</td><td>No</td></tr>
<tr><td>H6b: Affective reactions > signaler</td><td>$\Delta\beta = -0.506$, $p < 0.001$</td><td>Yes</td></tr>
<tr><td>H7: Gamification > positive signal</td><td>$t = 87.74$, $p < .001$</td><td>Yes</td></tr>
</table>
</td>
<td>Significant</td>
</tr>
<tr>
<td>[11]</td>
<td>As can be seen in Fig. 3, the path from gamification to enjoyment is positive and significant ($\beta = 0.729$, $SE = 0.213$, $p < 0.001$). As hypothesized, it appears that gamification increases survey enjoyment. Thus, H1 is supported. Like-</td>
<td>Significant</td>
</tr>
</table>

<table>
<tr><td></td><td> 
</td><td></td></tr>
<tr><td>[12]</td><td>Univariate ANOVA tests showed that gamification had no significant effect on respondent ratings of how enjoyable the game was ($F_{3,390}$=1.12, p=0.34) or how
<table>
<tr><th></th><th>Version 1: simple choice sets</th><th>Version 2.1: salesperson—no incorrect answers</th><th>Version 2.2: salesperson—with incorrect answers</th><th>Version 3: monsters—with incorrect answers</th></tr>
<tr><td>Enjoyment (How enjoyable was this survey game?)</td><td>2.9</td><td>3.0</td><td>3.4</td><td>3.1</td></tr>
<tr><td>Standard deviation</td><td>1.79</td><td>1.69</td><td>1.90</td><td>1.83</td></tr>
<tr><td>Standard error</td><td>0.18</td><td>0.19</td><td>0.18</td><td>0.18</td></tr>
</table></td><td>Not Significant</td></tr>
<tr><td>[13]</td><td>previous research. Gamification has not contributed to increases in completion rate ($p \leq .950$); both questionnaires had similar completion rates. Gamification's effects on engagement are unclear. Gamification had little impact on enjoyment ($p \leq .026$; 6% increase in enjoyment scores), a slight negative impact on perceived ease of use ($p \leq .025$; 4% decrease in</td><td>Significant</td></tr>
<tr><td colspan="3">4. Perceived Duration</td></tr>
<tr><td>[1]</td><td><table>
<tr><th></th><th>Text only N = 251</th><th>Decoratively visual N = 251</th><th>Functionally visual N = 252</th><th>Gamified N = 253</th><th>Total</th></tr>
<tr><td>Debrief questions</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>How interesting?[a]</td><td>5.2</td><td>5.4</td><td>5.7***</td><td>6.0***</td><td>5.6</td></tr>
<tr><td>How easy to read?[a]</td><td>6.1</td><td>6.3</td><td>6.4***</td><td>6.2***</td><td>6.2</td></tr>
<tr><td>How easy to answer?[a]</td><td>5.9</td><td>6.0</td><td>6.3***</td><td>6.3***</td><td>6.1</td></tr>
<tr><td>How fast?[a]</td><td>5.3</td><td>5.3</td><td>5.4***</td><td>5.0***</td><td>5.3</td></tr>
</table></td><td>Significant</td></tr>
<tr><td>[3]</td><td><table>
<tr><td>c) Subjective Ratings
4-item Likert-type questions</td><td>Fun
Time consuming
Preferred over other surveys</td><td>55
56
53</td><td>2.98
1.66
3.09</td><td>0.828
0.837
0.714</td><td>48
48
42</td><td>2.83
1.69
2.64</td><td>0.975
0.689
0.958</td><td>U=1233.5
U=1261.5
U=821.5</td><td>0.545
0.557
0.019</td></tr>
</table></td><td>Not Significant</td></tr>
<tr><td colspan="3">5. Preference</td></tr>
<tr><td>[3]</td><td><table>
<tr><td>c) Subjective Ratings
4-item Likert-type questions</td><td>Fun
Time consuming
Preferred over other surveys</td><td>55
56
53</td><td>2.98
1.66
3.09</td><td>0.828
0.837
0.714</td><td>48
48
42</td><td>2.83
1.69
2.64</td><td>0.975
0.689
0.958</td><td>U=1233.5
U=1261.5
U=821.5</td><td>0.545
0.557
0.019</td></tr>
</table></td><td>Significant</td></tr>
</table>

	Subjective Ratings: The post-test questionnaire included three Likert-type questions where participants rated fun ("The survey was fun"), perceived duration ("The survey took a lot of time"), and subjective preference ("I liked the survey better than other surveys"). The available answers were coded as "strongly disagree" (1), "disagree" (2), "agree" (3) and "strongly agree" (4). Evaluation results (Table 2c) show no significant differences regarding the first two questions, but preference was higher (better) in the gamified survey.	
[4]		**Significant**
[9]	 **Fig. 6.** Preference for Signaler. The difference between the percentage of respondents preferring a signaler with a gamified survey vis-a-vis one with a traditional survey is significant ($t = 87.74$, $p < .001$), which lends support for H7. Thus, we conclude that organizations employing gamified surveys will be preferred over those employing traditional surveys. The summary of all results is reported in Table 6. **Table 6** The results of hypotheses testing. Hypothesis / Statistical Test / Supported H1: Gamified survey > cognitive reactions — $M_{diff} = 0.058$, $p > 0.05$ — No H2: Gamified survey > affective reactions — $M_{diff} = 0.159$, $p < 0.001$ — Yes H3a: Extrinsic design mechanism > cognitive reactions — $\beta = 0.055$, $SE = 0.057$, $p > 0.05$ — No H3b: Extrinsic design mechanism > affective reactions — $\beta = 0.118$, $SE = 0.056$, $p < 0.05$ — Yes H4a: Intrinsic design mechanism > cognitive reactions — $\beta = 0.214$, $SE = 0.055$, $p < 0.001$ — Yes H4b: Intrinsic design mechanism > affective reactions — $\beta = 0.258$, $SE = 0.054$, $p < 0.001$ — Yes H5a: Self-presentation mechanism > cognitive reactions — $\beta = 0.134$, $SE = 0.045$, $p < 0.01$ — Yes H5b: Self-presentation mechanism > affective reactions — $\beta = 0.116$, $SE = 0.044$, $p < 0.01$ — Yes H6a: Cognitive reactions > signaler — $\Delta\beta = -0.065$, $p > 0.05$ — No H6b: Affective reactions > signaler — $\Delta\beta = -0.506$, $p < 0.001$ — Yes H7: Gamification > positive signal — $t = 87.74$, $p < .001$ — Yes gamification or by a traditional survey. Hence, H6a is not supported. On the other hand, the paths from affective reactions towards the game ($\beta = -0.361$, $SE = 0.063$, $p < 0.001$) and affective reactions towards the survey ($\beta = 0.145$, $SE = 0.058$, $p < 0.05$) to preference towards signaler are both positive and significant. Thus, the data suggest that affective reactions increase preference towards the signaler. More importantly, the difference between the two path coefficients is also significant ($\Delta = -0.506$, $p < 0.001$). This suggests that affective reactions caused by gamification have a stronger impact on preference towards the signaler than affective reactions caused by a traditional survey. Thus, H6b is supported.	**Significant**

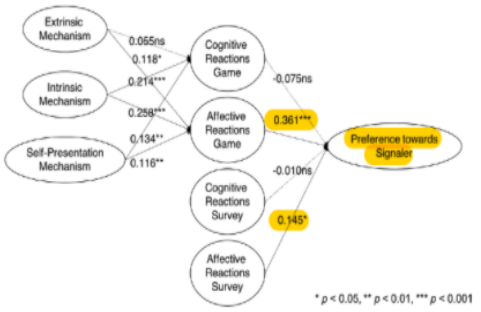


Fig. 5. Results of path analysis.

[10]

According to the Hypothesis 1, the commitment of the users has not decreased. More users have tended to answer open questions and tend to use more characters. According to Hypothesis 2, the use of game elements also has no influence on customer satisfaction. There are slight tendencies that users prefer to fill out the gamified questionnaire. The processing time is slightly longer for the gamified questionnaire than for the classic questionnaire. In the comparison between the classical and the gamified questionnaire the customer satisfaction remained at least the same. That means that the evaluation of

In summary, it can be stated that due to the small number of participants with 150 persons, only slight tendencies can be derived. The use of game elements in the questionnaire has shown that there has been no deterioration in the involvement of users. The game elements have also had no negative impact on customer satisfaction. A further result is that the guests prefer the design of the gamified questionnaire. It is im-

Significance not tested

[12]

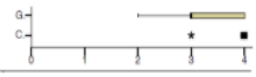
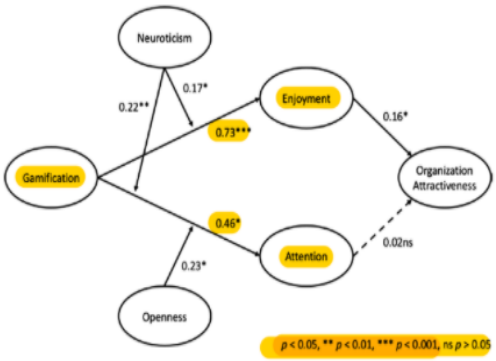
	Version 1: simple choice sets	Version 2.1: salesperson—no incorrect answers	Version 2.2: salesperson—with incorrect answers	Version 3: monsters—with incorrect answers
Enjoyment (How enjoyable was this survey game?)	2.9	3.0	3.4	3.1
Standard deviation	*1.79*	*1.69*	*1.90*	*1.83*
Standard error	*0.18*	*0.19*	*0.18*	*0.18*
Motivation (How motivated were you to complete the survey game?)	3.0	3.3	3.5	3.1
Standard deviation	*1.70*	*1.78*	*1.88*	*1.82*
Standard error	*0.18*	*0.19*	*0.18*	*0.18*
Interest (How interested were you in the subject matter of the questions?)	3.5	3.7	3.6	3.0
Standard deviation	*1.76*	*1.73*	*1.86*	*1.72*
Standard error	*0.18*	*0.19*	*0.18*	*0.18*

Univariate ANOVA tests showed that gamification had no significant effect on respondent ratings of how enjoyable the game was ($F_{3,390}=1.12$, $p=0.34$) or how motivated they were to complete the survey ($F_{3,390}=1.95$, $p=0.12$). We did find a significant difference in respondent interest in the subject matter of the questions ($F_{3,390}=2.84$, $p=0.038$).

As there were no significant differences found between versions 2.1 and 2.2, the addition of correct and incorrect answers in the salesperson scenario does not seem to have had an effect on enjoyment, motivation or interest. However, it seems likely that some element of the design of version 3 had an effect on respondents' interest in the subject matter: the fantastical setting or the different tone and writing style of the in-game customer statements *decreased* interest in the subject, opposite to our expectations

Significant

<table>
<tr><th colspan="3">6. Satisfaction</th></tr>
<tr><td>[1]</td><td>Other research in this area has relied heavily on debrief questions at the end of the survey that ask respondents about their satisfaction with the survey experience, often in comparison to other surveys they have taken. We asked five such questions, all on 7-point scales in which only the endpoints were labelled. The mean responses for those questions are shown in Table 2. After reviewing the results we concluded that, in general and despite one or two exceptions, respondents in the Functionally Visual and Gamified cells were more satisfied with their experience than those in the other two cells.</td><td>Significant</td></tr>
<tr><td>[10]</td><td>Table 7. Customer satisfaction and rating to the questionnaire tool
<table><tr><th></th><th>Customer satisfaction</th><th>Rating questionnaire</th></tr><tr><td>Classical questionnaire</td><td>4.4</td><td>4.1</td></tr><tr><td>Gamified questionnaire</td><td>4.4</td><td>4.1</td></tr></table></td><td>Significance not tested</td></tr>
<tr><td>[13]</td><td>($p \leq .950$); both questionnaires had similar completion rates. Gamification's effects on engagement are unclear. Gamification had little impact on enjoyment ($p \leq .026$; 6% increase in enjoyment scores), a slight negative impact on perceived ease of use ($p \leq .025$; 4% decrease in perceived ease of use scores) and no influence on perceived control ($p \leq .185$) and</td><td>Not Significant</td></tr>
<tr><th colspan="3">7. Perceived effort</th></tr>
<tr><td>[1]</td><td><table><tr><th></th><th>Text only
N = 251</th><th>Decoratively visual
N = 251</th><th>Functionally visual
N = 252</th><th>Gamified
N = 253</th><th>Total</th></tr><tr><td>Debrief questions</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>How interesting?[a]</td><td>5.2</td><td>5.4</td><td>5.7***</td><td>6.0***</td><td>5.6</td></tr><tr><td>How easy to read?[a]</td><td>6.1</td><td>6.3</td><td>6.4***</td><td>6.2***</td><td>6.2</td></tr><tr><td>How easy to answer?[a]</td><td>5.9</td><td>6.0</td><td>6.3***</td><td>6.3***</td><td>6.1</td></tr></table></td><td>Significant</td></tr>
<tr><td>[5]</td><td>Significantly more children found it easier to fill out the gamified survey, $\chi^2(8) = 38.003$, $p < .001$. For instance, the top response on the 5-point scale "very easy" was selected among 31.0% of the respondents in the text-only, 30.4% in the visual, and 44.4% in the gamified survey. A multivariate logistic regression predicting the top response showed that it was easier for older respondents (OR = 1.130, $p < .001$), for those who had good academic performance (OR = 1.688, $p < .001$), and who used the Internet almost every day (OR = 1.501, $p < .05$). It was more difficult to complete the text-only (OR = 0.522, $p < .001$) and visual survey</td><td>Significant</td></tr>
<tr><td>[6]</td><td><table><tr><th></th><th>Text-only survey</th><th>Visual survey</th><th>Gamified survey</th><th>Statistics</th></tr><tr><td>Survey burden</td><td></td><td></td><td></td><td></td></tr><tr><td>Ease of completing the survey – first wave
Scale: 1-5 (very easy-very difficult)</td><td>2.09
(0.94)</td><td>1.96
(0.83)</td><td>1.80
(0.86)</td><td>F(2,1034)=10.22***</td></tr><tr><td>Ease of completing the survey – second wave</td><td>1.88
(0.82)</td><td>1.80
(0.81)</td><td>1.62
(0.80)</td><td>F(2,730)=7.04***</td></tr></table></td><td>Significant</td></tr>
</table>

8. Perceived Usability		
[4]	**7.3 Self-Rated User Experience** Perceived usability and user experience were assessed upon completion of the sports survey through a post-test questionnaire that included System Usability Scale (SUS) [4] questions. It was filled out by an overall number of 40 respondents (21 gamified, 19 conventional). Pair-wise comparison of *individual SUS questions* (see Table 5) revealed that respondents were significantly more inclined to frequently use the gamified survey (2.81 ± 0.75), compared to the conventional version (1.16 ± 1.02). However, respondents felt significantly less confident using the gamified survey (3.1 ± 0.7), compared to the conventional version (3.79 ± 0.42). There was no significant effect of survey design on any other SUS question. *Overall SUS scores* for both survey versions were comparable as well, with the gamified survey scoring 77.98 points and the conventional survey scoring 79.08. Answers to further questions in the post-test questionnaire (Table 6) showed that respondents found the gamified survey (3.29 ± 0.56) significantly more *fun to use* than the conventional survey (2.32 ± 1.01). They were also significantly more inclined to *recommend* the gamified survey (3.38 ± 0.67), compared to the conventional survey (2.42 ± 0.96). SUS 1: I think that I would like to use this system frequently. Gamified 21 2.81 3.00 0.75 Conventional 19 1.16 1.00 1.02 U=140 p=0.000 Total 40 2.03 2.00 1.21  SUS 9: I felt very confident using the system. Gamified 21 3.10 3.00 0.70 Conventional 19 3.79 4.00 0.42 U=145 p=0.003 Total 40 3.43 4.00 0.68 	**Significant**
[13]	engagement are unclear. Gamification had little impact on enjoyment ($p \leq .026$; 6% increase in enjoyment scores), a slight negative impact on perceived ease of use ($p \leq .025$; 4% decrease in perceived ease of use scores) and no influence on perceived control ($p \leq .185$) and	**Significant**
9. Attention		
[9]	towards the extrinsic mechanism increases affective reactions. Thus, H3b is supported. In contrast, the paths from attitude towards the intrinsic mechanism to cognitive reactions ($\beta = 0.214$, $SE = 0.055$, $p < 0.001$), and affective reactions ($\beta = 0.258$, $SE = 0.054$, $p < 0.001$), towards the game are both positive and significant. Thus, attitude towards the intrinsic mechanism increases cognitive and affective reac- Likewise, the paths from attitude towards the self-presentation mechanism to cognitive reactions ($\beta = 0.134$, $SE = 0.045$, $p < 0.01$) and affective reactions ($\beta = 0.116$, $SE = 0.044$, $p < 0.01$) are both positive and significant. This suggest that attitude towards the self-pre-	**Significant**
[11]	increases survey enjoyment. Thus, H1 is supported. Likewise, the path from gamification to attention is significant ($\beta = 0.459$, $SE = 0.218$, $p < 0.05$). As expected, it appears that gamification does increase attention while filling out surveys. Hence, H2 is supported. The path from enjoyment  	**Significant**

<table>
<tr>
<td>[12]</td>
<td>
<table>
<tr><th></th><th>Version 1: simple choice sets</th><th>Version 2.1: salesperson—no incorrect answers</th><th>Version 2.2: salesperson—with incorrect answers</th><th>Version 3: monsters—with incorrect answers</th></tr>
<tr><td>Enjoyment (How enjoyable was this survey game?)</td><td>2.9</td><td>3.0</td><td>3.4</td><td>3.1</td></tr>
<tr><td>Standard deviation</td><td>1.79</td><td>1.69</td><td>1.90</td><td>1.83</td></tr>
<tr><td>Standard error</td><td>0.18</td><td>0.19</td><td>0.18</td><td>0.18</td></tr>
<tr><td>Motivation (How motivated were you to complete the survey game?)</td><td>3.0</td><td>3.3</td><td>3.5</td><td>3.1</td></tr>
<tr><td>Standard deviation</td><td>1.70</td><td>1.78</td><td>1.88</td><td>1.82</td></tr>
<tr><td>Standard error</td><td>0.18</td><td>0.19</td><td>0.18</td><td>0.18</td></tr>
</table>
Univariate ANOVA tests showed that gamification had no significant effect on respondent ratings of how enjoyable the game was ($F_{3,390}=1.12$, $p=0.34$) or how motivated they were to complete the survey ($F_{3,390}=1.95$, $p=0.12$). We did find a
</td>
<td>Not significant</td>
</tr>
<tr>
<td>[13]</td>
<td>($p \leq .950$); both questionnaires had similar completion rates. Gamification's effects on engagement are unclear. Gamification had little impact on enjoyment ($p \leq .026$; 6% increase in enjoyment scores), a slight negative impact on perceived ease of use ($p \leq .025$; 4% decrease in perceived ease of use scores) and no influence on perceived control ($p \leq .185$) and concentration/focus ($p \leq .101$). The gamified questionnaire was even perceived as being</td>
<td>Not Significant</td>
</tr>
</table>

Bibliography

[1] Judd Antin and Elizabeth F. Churchill. 2011. Badges in Social Media: A Social Psychological Perspective. *Gamification Workshop Proceedings* (01 2011), 4.

[2] Pippa Bailey, Gareth Pritchard, and Hollie Kernohan. 2015. Gamification in Market Research: Increasing Enjoyment, Participant Engagement and Richness of Data, but what of Data Validity? *International Journal of Market Research* 57 (01 2015), 17–28. DOI:http://dx.doi.org/10.2501/IJMR-2015-003

[3] Florian Brühlmann. 2015. The Effects of Framing in Gamification: A Study of Failure. (10 2015). DOI:http://dx.doi.org/10.13140/RG.2.1.4641.2884

[4] Briana Brownell, Jared Cechanowicz, and Carl Gutwin. 2015. *Gamification of Survey Research: Empirical Results from Gamifying a Conjoint Experiment*. 569–591. DOI: http://dx.doi.org/10.1007/978-3-319-10208-5_29

[5] Marc Busch, Elke Mattheiss, Wolfgang Hochleitner, Christina Hochleitner, Michael Lankes, Peter Fröhlich, Rita Orji, and Manfred Tscheligi. 2016. Using player type models for personalized game design - An empirical investigation. 28 (01 2016), 145–163.

[6] Jared Cechanowicz, Carl Gutwin, Briana Brownell, and Larry Goodfellow. 2013. Effects of Gamification on Participation and Data Quality in a Real-World Market Research Domain. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications (Gamification '13)*. Association for Computing Machinery, 58–65. DOI:http://dx.doi.org/10.1145/2583008.2583016

[7] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011a. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. Association for Computing Machinery, 9–15. DOI: http://dx.doi.org/10.1145/2181037.2181040

[8] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. 2011b. Gamification. Using Game-Design Elements in Non-Gaming Contexts. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. Association for Computing Machinery, 2425–2428. DOI:http://dx.doi.org/10.1145/1979742.1979575

[9] Elisabeth Deutskens, Ko de Ruyter, Martin Wetzels, and Paul Oostervald. 2004. Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters* (2004), 21 – 36. DOI:http://dx.doi.org/10.1023/B:MARK.0000021968.86465.00

[10] Hynes A. Graeme. 2009. Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Journal of psychology and marketing* 26 (02 2009), 204–212. DOI: http://dx.doi.org/10.1002/mar.20269

[11] DH Granello and Joe Wheaton. 2004. Online data collection: Strategies for research. *Journal of counseling and development: JCD* 82 (09 2004), 387–393.

[12] Theo Downes-Le Guin, R. Baker, Joanne Mechling, and Erica Ruyle. 2012. Myths and Realities of Respondent Engagement in Online Surveys. *International Journal of Market Research* 54 (2012), 613 – 633.

[13] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? — A Literature Review of Empirical Studies on Gamification. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE, 3025 – 3034. DOI: http://dx.doi.org/10.1109/HICSS.2014.377

[14] Johannes Harms, Stefan Biegler, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. 2015a. Gamification of online surveys: Design process, case study, and evaluation. In *In Proc. 15th IFIP TC 13 Int. Conf. Human-Computer Interaction, INTERACT 2015, J. Abascal, S. D. J. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, Eds., vol. 9296 of Lecture Notes in Computer Science. Springer (2015) (INTERACT 2015)*. Springer Link, 214 – 231. https://doi.org/10.1007/978331922701616

[15] Johannes Harms, Dominik Seitz, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. 2015b. Low-Cost Gamification of Online Surveys: Improving the User Experience through Achievement Badges. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. Association for Computing Machinery, 109–113. https://doi.org/10.1145/2793107.2793146

[16] Sheena Iyengar and Mark Lepper. 2001. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of personality and social psychology* 79 (01 2001), 995–1006. DOI:http://dx.doi.org/10.1037/0022-3514.79.6.995

[17] Florian Keusch and Chan Zhang. 2017. A Review of Issues in Gamified Surveys. *Social Science Computer Review* 35, 2 (2017), 147–166. DOI:http://dx.doi.org/10.1177/0894439315608451

[18] Nicole Koenig-Lewis. 2013. The effects of gamification on market research engagement and response. Academy of Marketing Conference.

[19] Pascal Lessel, Maximilian Altmeyer, and Antonio Krüger. 2015. Analysis of Recycling Capabilities of Individuals and Crowds to Encourage and Educate People to Separate Their Garbage Playfully. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, 1095–1104. DOI:http://dx.doi.org/10.1145/2702123.2702309

[20] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2016. "Don't Whip Me With Your Games": Investigating "Bottom-Up" Gamification. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, 2026–2037. DOI:http://dx.doi.org/10.1145/2858036.2858463

[21] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2017. Measuring the Effect of "Bottom-up" Gamification in a Microtask Setting. In *Proceedings of the 21st International Academic Mindtrek Conference (AcademicMindtrek '17)*. Association for Computing Machinery, New York, NY, USA, 63–72. DOI:http://dx.doi.org/10.1145/3131085.3131086

[22] Pascal Lessel, Maximilian Altmeyer, Lea Verena Schmeer, and Antonio Krüger. 2019. "Enable or Disable Gamification?": Analyzing the Impact of Choice in a Gamified Image Tagging Task. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, Article Paper 150, 12 pages. DOI:http://dx.doi.org/10.1145/3290605.3300380

[23] Aigul Mavletova. 2015a. A Gamification Effect in Longitudinal Web Surveys among Children and Adolescents. *International Journal of Market Research* 57, 3 (2015), 413–438. DOI:http://dx.doi.org/10.2501/IJMR-2015-035

[24] Aigul Mavletova. 2015b. Web Surveys Among Children and Adolescents: Is There a Gamification Effect? *Social Science Computer Review* 33, 3 (2015), 372–398. DOI: http://dx.doi.org/10.1177/0894439314545316

[25] Jane McGonigal. 2011. *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin Group , The. 400 pages.

[26] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. Association for Computing Machinery, 1137–1142. DOI: http://dx.doi.org/10.1145/2468356.2468559

[27] Hendrik Müller and Aaron Sedley. 2015. Designing Surveys for HCI Research. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. Association for Computing Machinery, 2485–2486. DOI:http://dx.doi.org/10.1145/2702613.2706683

[28] Rita Orji, Kiemute Oyibo, and Gustavo F. Tondello. 2017. A Comparison of System-Controlled and User-Controlled Personalization Approaches. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17)*. Association for Computing Machinery, New York, NY, USA, 413–418. DOI:http://dx.doi.org/10.1145/3099023.3099116

[29] David Prott and Martin Ebner. 2020. The Use of Gamification in Gastronomic Questionnaires. 2 (2020), 101–118. DOI:http://dx.doi.org/10.3991/ijim.v14i02.11695

[30] Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. 2017. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior* 69 (2017), 371 – 380. DOI:http://dx.doi.org/https://doi.org/10.1016/j.chb.2016.12.033

[31] Silvia Schacht, Florian Keusch, Nils Bergmann, and Stefan Morana. 2017. Web Survey Gamification - Increasing Data Quality in Web Surveys by Using Game Design Elements. In *Twenty-Fifth European Conference on Information Systems (ECIS), Guimarães, P, June 5-10, 2017*. Universidade do Minho, Guimarães, P.

[32] Amy Schade. 2016. Customization vs. personalization in the user experience. *Nielsen Norman Group World Leaders in Research-Based User Experience* (2016).

[33] Sofia Schöbel, Matthias Söllner, and Jan Marco Leimeister. 2016. The Agony of Choice - Analyzing User Preferences Regarding Gamification Elements in Learning Management Systems. In *International Conference on Information Systems (ICIS)*. Social Science Research Network, 22. DOI:http://dx.doi.org/10.2139/ssrn.3159163

[34] Kristin Siu and Mark O. Riedl. 2016. Reward Systems in Human Computation Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 266–275. DOI:http://dx.doi.org/10.1145/2967934.2968083

[35] Ferran Teixes. 2017. Yu-Kai Chou (2016). Actionable Gamification: beyond points, badges and leaderboards. Octalysis Media: Fremont. CA. *Revista Internacional de Organizaciones* (06 2017), 137. DOI:http://dx.doi.org/10.17345/rio18.137-144

[36] Edmund R. Thompson. 2007. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology* 38, 2 (2007), 227–242. DOI:http://dx.doi.org/10.1177/0022022106297301

[37] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 229–243. DOI:http://dx.doi.org/10.1145/2967934.2968082

[38] Tamilla Triantoro, Ram Gopal, Raquel Benbunan-Fich, and Guido Lang. 2019. Would you like to play? A comparison of a gamified survey with a traditional online survey method. *International Journal of Information Management* 49 (12 2019), 242–252. DOI:http://dx.doi.org/10.1016/j.ijinfomgt.2019.06.001

[39] Tamilla Triantoro, Ram Gopal, Raquel Benbunan-Fich, and Guido Lang. 2020. Personality and games: enhancing online surveys through gamification. *Information Technology and Management* (04 2020). DOI:http://dx.doi.org/10.1007/s10799-020-00314-4