Exploring the inner workings of the leaderboard for use in educational contexts

Sjoerd Alexander André de la Porte
under supervision of Ana Ciocarlan and Hanna Hauptmann

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Master of Science
of
Utrecht University.

Universiteit Utrecht

Department of Information and Computing Sciences

August 24, 2022

Abstract

How can we make tedious, repetitive, and difficult tasks on the computer more engaging and fun? Is it possible to make education or work more exciting, by making the experience feel more like a game without compromising the effectiveness? Gamification, being the addition of game elements into non-game environments, has become a popular software trend. It has been criticized for its widespread implementation due to a lack of scientific evidence and thorough understanding of its inner workings. Studies on gamification have most often reported an overall increases in motivation and engagement, while other studies have found that it can have a demotivating effect on part of the user group. The game-element that is most often linked to negative effects such as loss of performance is the leaderboard. To advance the discipline, we have set our focus on providing a deeper understanding of what aspects cause, influence and effect the gamification element leaderboard and provide requirements and recommendations for implementing the leaderboard into an education context. Two studies were conducted for this purpose. The first study was a quantitative study (N=34) in which students competed in a knowledge quiz. Consequently each student received a score, and with that a rank in their group's leaderboard. This part of the study ended with a reflection on their experience. The second part of the study consisted of two qualitative 1-hour focus group sessions (N=8), in which a more thorough understanding of the experience and the results was discussed. Seven factors that influence the leaderboard effect were identified, as well as five requirements when implementing the leaderboard in educational context. Our study shows an increase in motivation for some people, while simultaneously creating a less enjoyable learning environment for others. This research presents an overview of what causes this motivation or demotivation to occur, as well as recommendations on how to design, implement and evaluate the leaderboard in educational contexts.

Contents

List of Tables

List of Figures

Preface

During my educational carrier, laptops, smartboards and other digital tools have largely replaced the traditional books, chalk board, pen and paper on schools and universities. In primary school, I remember having to save my word-files to a floppy-disk, and having to 'dial-in' on my fathers expensive dial-up connection to access the internet. As I am now ending my educational carrier, I am almost entirely working on a permanent internet connection in the cloud, accompanied by the soft buzz of a robot vacuuming my floor. Technology has moved at a rapid rate, often transforming traditional tools towards digital versions of the same tool. As we gotten used to using these digital product, we have sometimes forgotten the possibilities of software to go much beyond the capabilities of digital tools. Educational tools are still largely digitized books (pdf's, websites), combined with a digitized pen (keyboard, or stylus). One could argue that this is, to say the least, not very creative.

My motivation is derived from friends who failed in the regular education system. Smart people, who were unable to concentrate and sit still. They were punished, given medicine to concentrate, and demotivated. These same people were, to the discontent of parents and teachers, able to concentrate for hours on end playing video games. I've wondered what it was about videogames that made them forget they had issues concentrating. I have always believed that if we could understand what creates this concentration and motivation, and implement it in the educational systems, people like them might become the best performing students in class. This master thesis is dedicated towards that goal, aiming to create educational tools that go much beyond the capabilities of the traditional, especially in motivation.

I would like to express my heartfelt gratitude to my first supervisor, Ana Ciocarlan, who gave me the freedom and guidance to truly do the research that I was inspired to complete. I want to thank her for all the online meetings, positive reinforcement, suggestions and effort she has spend to allowed this thesis to be completed. Furthermore, I would want to thank Hanna Hauptmann, my second supervisor, for her feedback, support during the quizzes, feedback on multiple aspects of this thesis and for helping me in finding the thesis topic.

Chapter 1

Introduction

How can one make tedious, repetitive, and difficult tasks on the computer more engaging and fun? Is it possible to make education or work more exciting, by making the experience feel more like a game, without compromising the effectiveness? The idea that game design holds valuable principles to make even the most tedious activity more enjoyable has a long history in Human-Computer Interaction, design and education, re-emerging under names such as ludic design, serious games, game-based learning or playful interaction (Deterding, 2014a). Gamification can be seen as one of the newest iterations of this idea, though it has its flaws (Deterding, 2014b).

> "Mowing the lawn or waiting in a dentist's office can become enjoyable provided one restructures the activity by providing goals, rules, and the other elements of enjoyment" - Csikszentmihalyi and Csikzentmihaly (1990)

Corporations and educational institutions are implementing gamified experiences across the board in order to encourage productivity, boost morale, and improve engagement (Zhonggen, 2019). Gamification has become a widespread technological trend in recent years (Cagiltay et al., 2015). It is defined as the process of adding game-elements towards otherwise non-game contexts (Deterding et al., 2011). The most commonly implemented game-elements are levels, points, badges, leaderboards and avatars (Barata et al., 2017). One example is the popular language learning application Duolingo, which currently holds 300 million users worldwide. It is generally seen as a strong representation of gamification with game-elements such as systematic levels, reward incentives and the ranking of users according to their achievements (Shortt et al., 2021).

Though promising, the popularity and widespread implementation of gamification has been criticized for the lack of scientific evidence and thorough understanding of its inner workings (Dichev and Dicheva, 2017). Research often focuses on metrics such as increased motivation and engagement, disregarding why these effects occur. More critically, Toda et al. (2017) showed that implementation may cause various problems if not well designed and supported by theory. One of the problems that could arise is a *decrease* in user performance or motivation. There is a gap in the literature as to which aspects influence the positive or negative response of individual game-elements. Studying these elements individually was proposed by Toda et al. (2017), which will be the main focus of this paper. This research focuses on uncovering the inner workings of one the most implemented game-element in gamification: the leaderboard (Barata et al., 2017). The leaderboard is perhaps the most challenging game-elements, which is both one of the elements that can create the greatest increase in motivation, as well the being the most likely to demotivate

users (Toda et al., 2017).

Therefore, the first research goal is to explore if social comparison theory can be used to understand the psychology of the leaderboard. The second research goal is to investigate what aspects of the leaderboard create an increase or decrease in motivation. The third and final research goal is to discover requirements that could reduce the demotivation and increase the motivation that occurs due to the implementation of the leaderboard.

1.1 Scope, contributions and research question

The primary goal of this research is to follow a user-centered approach to explore the inner workings of the leaderboard and proposes a different psychological background. The research is divided into four parts:

- RQ1 How is the leaderboard used?
- RQ2 How are people effected by the leaderboard?
- RQ3 Can social comparison theory be used to understand the leaderboard?
- RQ4 What requirements make the leaderboard applicable to an educational context?

Having defined the goals and questions of this research, chapter 2 discusses the relevant definitions, concepts and theories, as well as related gamification literature. How these questions are answered are explained in the methodology chapter 3. Finally, the results of the study are shown in chapter 4 and are summarized, interpreted and discussed in chapter 5 and is combined with the limitations of this research and options for future work. The main contributions to the research are concluded in chapter 6.

Chapter 2

Literature Review

In this chapter, relevant background information and related work is presented by investigating key components in the form of a theoretical review. We will firstly discuss what gamification is and its relation to persuasive technology and education. Secondly, we investigate current gamification research, how it is implemented, and what gaps in literature still exist. Here, we also discuss the psychological background that underlies gamification research. Thirdly, we elaborate upon what may influence the effectiveness of gamification according to prior work.

2.1 Gamification and its effect

2.1.1 What is gamification?

The term *Gamification* became popular in 2010 and is in most studies used to describe the implementation of game-inspired elements into non-game environments (Deterding et al., 2011). Specific game-mechanics are referred to as game elements, the most common are points, badges, avatars and leaderboards (Barata et al., 2017). They are meant to engage and motivate the user, and to enhance the user experience (Zainuddin et al., 2020). A popular gamification example can be found in the language learning application Duolingo, which currently holds 300 million users worldwide. Duolingo implemented (amongs others) the elements points, leaderboards and badges in order to create a more interesting learning experience (Shortt et al., 2021). Gamification has shown to be able to stimulate learners to achieve greater goal orientation by increasing their persistence, motivation and engagement (Zainuddin et al., 2020). Gamification is sometimes confused with "Serious Games" or "Game-based learning", as all three topics of research are inspired by games. However, game-based learning makes use of existing games and serious games is the creation of custom-made games for the purpose of learning, whereas gamification adds an extra layer of game elements to existing (non-game) environments (Plass et al., 2015).

2.1.2 Gamification in educational context

'Gamification in Education' refers to the introduction of game design elements into the design of the learning process (Dichev and Dicheva, 2017). Within education, gamification often aims to increase motivation and engagement for learning objectives as they are considered to be important predictors for student academic achievement (Linehan et al., 2011). By incorporating game elements in the design of a learning process, it is possible to engage learners in a more productive learning experience (Caponetto et al., 2014). Researchers have given increased attention to gamification in educational context in the last decade due to its effectiveness and the emergence of technology assisted education (Zhonggen, 2019). This interest was further increased due to the

COVID-19 pandemic, as it forced the sector to switch primarily towards digital education (Zainuddin et al., 2020). It was implemented to support learning in various ways, such as improving self-guided study, increasing completion of assignments and promoting collaboration (Caponetto et al., 2014). Gamification proved to play a crucial role in improving upon the low completion rates of Massive Open Online Courses (MOOC's), such as Coursera and Skillshare, where the improved completion rate often results in an improvement of the user's learning performance (Khalil et al., 2018)

2.1.3 Gamification as a Persuasive Technology

Research in the area of Human-Computer Interaction (HCI) has been exploring the role of Persuasive Technologies in order to encourage and motivate people towards more sustainable lifestyle (DiSalvo et al., 2010). Games and gamification are being used increasingly to persuade players to change their behavior, or to adopt a particular perspective on complex (social) issues, often combining persuasive strategies with game-elements (Kors et al., 2015). Persuasive Technology is an interdisciplinary research field that focuses on the design and development of technologies that are aimed at changing attitudes and behaviour through persuasion and social influence, but not through coercion or deception (Hamari et al., 2014a). For a technology to be called a persuasive technology, the persuasion must be intentional (Fogg, 1998). The technology must be designed to guide the user towards an attitude or behavioral change which guides the design process.

However, while Persuasive Technology may employ some strategies used in gamification, persuasive strategies differ from the goal of gamification. For example, a user might be rewarded a badge after a specific set of actions has been completed, like finishing the first level of a game within a certain time. The badge is in gamification described as a mechanism to show the user his or her progress and achievements within the system (Thiel, 2016). In Persuasive Technology on the other hand, the badge would be seen as a reward given to drive behavioral change. For example, rewarding a badge when a user completes at least one learning-exercise, every day, for one week. Badges are rewarded in both examples, but only in the latter is the reward given to specifically shape behaviour. In the study of (Shih and Jheng, 2017), persuasive strategies were linked to 29 game-elements in order to create an effective intervention to encourage energy-saving behaviour, which is only one example that shows how game-elements and persuasive technology can work together.

2.2 Existing research

2.2.1 Existing psychological background for leaderboard's

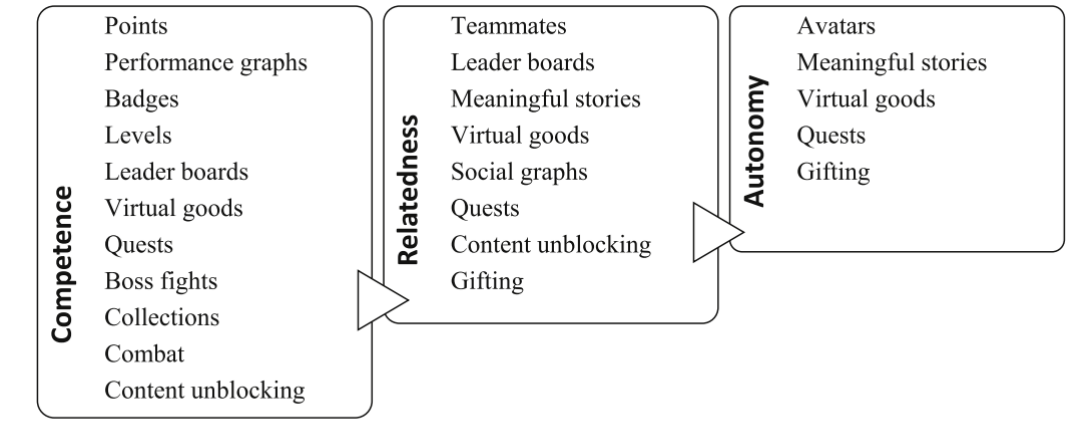
In order to understand the leaderboard, this study makes the assumption that the effects of a leaderboard can be explained by first understanding the psychological background related to the leaderboard, including social comparisons theory. As such, this section dives into the leaderboards psychological background.

Prior research on gamification and gaming has often aimed to find specific effects, such as increased motivation (Przybylski et al., 2010). The techniques and theories they use assess the degree to which gamification can have a favorable, negative, or no impact on particular outcomes under a specific set of conditions. In gamification research, motivation often refers to intrinsic

motivation based upon the SDT theory (Seaborn and Fels, 2015). Other measures such as engagement most often refer to immersion, presence and absorption (Abbasi et al., 2017). It is less widely studied what underlay these conditions.

The underlying theoretical framework in gamified learning is in most studies the Self-Determination Theory (SDT) (Seaborn and Fels, 2015). According to the theory, human motivation can be understood as a drive to satisfy three basic psychological needs: Competence, Relatedness and Autonomy (Deci and Ryan, 2012). Competence describes the need to be effective in dealing with the environment, and being able to build competence and develop mastery over important tasks. Autonomy described the need to control the course of one's life, and the need to feel control over behavior. Relatedness (also called connection) described the need of a sense of belonging and connection with others. To motivate students and satisfy these needs, (Buil et al., 2020) provided empirical evidence that shows how various game design elements can effectively be used for this purpose. Figure 2.1 shows which game-elements were linked to the psychological needs from the SDT theory.

Figure 2.1: Psychological needs (SDT) linked to game design elements. Figure used with the permission of Dr. Zamzami Zainuddin. From his work (Zainuddin et al., 2020)



The flow theory, put out by (Csikszentmihalyi et al., 2005) is the second-most popular hypothesis in gamification research. The state of mind known as "flow" is characterized by a person's complete immersion in a sense of energised focus, full involvement, and enjoyment in the process of the activity. According to the theory, three conditions must be met to achieve a flow state. These are: a clear set of goals that channel attention towards a specific purpose, a balance between perceived challenges and the user's perceived skills and clear and immediate feedback. Another study discovered that the presence of game features in gaming activities has the potential to put players in a state of flow, given that the challenge is acceptable for the player's level of expertise (Rachels and Rockinson-Szapkiw, 2018). The ability of game features to promote a more fun and engaging player experience and enhance their gamified learning experience is also implied by this.

Models such as the Learning Mechanics and Game Mechanics model (LM-GM), linked game mechanics to specific learning activities (Natucci and Borges, 2021). However, current models

give little insight on the optimal combination of game elements and lacks insight to create a gamified environment (DiSalvo et al., 2010). Hence, there is gap in the literature that can be solved by conducting research into the the incremental effects of certain game elements in educational context. However, these theories give a general idea on implementations and recommendations, while specific aspects are yet to be researched. Most game-elements, whereas specific game-elements could be further understood by matching the specific game-element towards a related psychological theory.

2.2.2 Social comparison theory and leaderboards

People will, every now and again, have the urge to compare themselves to others. Social comparison should, according to (Gerber et al., 2018), be recognized as inevitable. In our daily lives, we might want to, for example, compare our athletic performance, psychical attractiveness or even intelligence with each others. The social comparison theory was first proposed by Leon Festinger in 1954 (Festinger, 1954), whom suggested that people have a innate drive to evaluate themselves. He argues that this evaluation is often done in comparison to others. In theory, if a social comparison derives its (motivational or engaging) value as an evaluation tool, than this would suggest that the leaderboard should also be understood as being mainly an evaluation tool.

Who people want to compare themselves to, and what the effects are when people compare themselves, have been studied ever since his original work and is summarized in the meta-analysis of (Gerber et al., 2018). Their analysis finds that people predominantly want to compare themselves with other who they believe are, in some way, superior to them. As most studies on this subject have been done on students, these students wanted to compare their grades with better performing students. These comparisons significantly effected the Mood and Ability Estimate of the students. The Ability Estimate is how good the students believed to be at a certain subject or skill, whereas Mood gives an estimate on how the student felt after the comparison. When the other student (the 'target') had performed worse, both Ability Estimate and Mood showed to increase. When the other outperformed the student, both showed a similarly effective decline. Long-term effects on mental well-being and the feeling of insecurity that have been published throughout the years showed to be insignificant after taking publication bias into account. If the leaderboard were to be created towards the wishes of the user, this information would suggest that better performers should not be hidden from the user, and that other solutions then not showing this information should be explored to compensate for the possible negative effects.

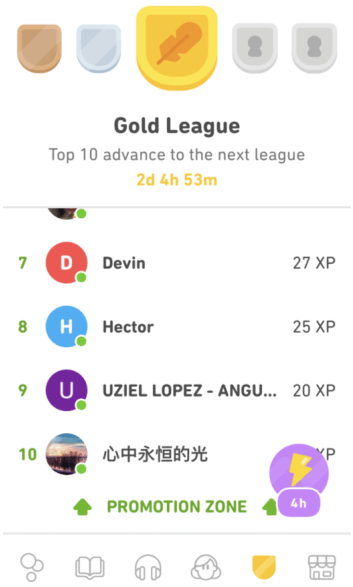
In order to find how the negative effects can be diminished, we first have to understand what causes the comparison's effects to be greater or weaker. According to (Gerber et al., 2018), the effects of a social comparison is greater when the target is within the persons proximity. Proximity, in this case, refers to closeness in physical or mental form. This is in line with the idea that a comparison is more influential when it influences the self, which is more likely to happen with people within one's social circle. The closer, the greater the impact after a comparison. The effects of a comparison are also diminished when a person is already knowledgeable with the topic or skill in which he or she is compared on. For example, a professional cyclist that is knowledgeable about his or her own average performance, can use this knowledge to mentally nuance a 'bad race', diminishing the negative effects on the Mood and Ability Estimate. If this also applies toward a digital environment is yet to be studied.

2.2.3 Prior research on leaderboards

The leaderboard is a game-element defined as a "visual display that ranks players according to their accomplishments" (Christy and Fox, 2014). Leaderboards are one of the most widely used game elements in gamification (Hamari et al., 2014b) and are considered one of the ten "main ingredients" of game design (Reeves and Read, 2009).

Depending on the design of a leaderboards, it can emphasize the users continues performance, the users performance in comparison to other users (Christy and Fox, 2014), status reporting or competition (Butler, 2013). They are used to enhance engagement through social comparisons (Jia et al., 2017). According to (Zichermann and Cunningham, 2011), we can distinguish leaderboards as being either absolute or relative. Absolute (also called 'infinite') leaderboards display all users and their scores, often starting at the top. This causes the top players to experience a broader sense of accomplishment in comparison to players at the bottom. Relative (also called 'no-disincentive') leaderboards place the user in the middle of the leaderboard, only showing the users above and below them. Figure 2.2 gives an example of an absolute leaderboard.

Figure 2.2: Example of an absolute leaderboard from Duolingo



How information is presented on the leaderboard can influence the users reaction. In the experiment of (Butler, 2013), both the experienced fun and user behaviour were effected by the presentation. In the experiment, users were presented with a leaderboard that showed either (A) only people that performed better, (B) only people that performed worse or (C) no other people at all. According to his research, the amount of points received had no impact on the experienced fun of the user. Instead of points received, it was the leaderboard position in comparison to others that impacted both fun and behaviour. Unlike the expectations of the researchers, the perception of fun was not increased when users were presented with a winning condition, which instead polarised the reported fun towards being either highly negative or positive. In relation to the neutral condition, users were 50 percent more likely to replay the game in order to increase their score. The losing condition centralized the perception of fun, causing responses in the middle of

the scale. The losing condition caused users to replay the game 100 percent more often in relation to the neutral condition. (Butler, 2013) concludes that the affect of the player is not always as expected, and that the response to the leaderboard is not the same for each player. This shows that no matter the position on the leaderboard, users are more likely to try again. These findings are similar to the empirical study of (Ortiz-Rojas et al., 2019), in which the relative and infinite leaderboard were both used as a solution to enhance the learning performance (LP) of students. In comparison to the neutral condition without a leaderboard, both leaderboard implementations had a significant positive impact on learning performance, leading to significantly increased average learning performance.

Result on leaderboard implementation have often found to be mixed, and researchers have tried to differentiate the user group based upon player type, gender, and context. While these differentiation's would influence the results of the experiment, they do not give a thorough understanding of the inner workings of the leaderboard. Instead, in order to understanding the leaderboard, the psychology of how people compare themselves in a natural situation should first be understood, as well as the effects of a comparison. When this is discovered, it could be assumed that the same mechanics play a role in a digital environment, that has yet to be studied.

2.3 Gaps in gamification research

Gamification's popularity and widespread implementation has been criticized for the lack of supportive evidence and thorough scientific understanding of its inner workings (Dichev and Dicheva, 2017). In their review, they conclude that the focus of research in gamification is as diverse as the outcomes of these studies, with too broadly stated research questions and conclusions. In the systematic review of (Zainuddin et al., 2020), 46 papers were summarized for their aims and effect, which found that 37 of these papers focused on motivation. The analysis concludes that gamification showed to achieve positive outcomes in motivation and engagement. However, in the research of (Broer, 2014), concludes that the current focus of gamification on badges, levels and leaderboards is too narrow to be useful in the long run, as many components of motivation are ignored. The research of (Toda et al., 2017) focused specifically on the negative effects that can be caused by gamification, and found that loss of performance was the most occurring negative effect, most often cited in relation to the leaderboard. The leaderboard was most strongly associated to many negative effects mapped in their work.

According to (Dichev and Dicheva, 2017), there is currently too little knowledge on how to avoid gamified scenario's that could harm learning. Research may conclude a positive result with a slight average positive effect on the sample, while disregarding a great negative effect on a few. If gamification is to be implemented for everyone, a solution should be found for the current unlucky few. The main challenge in gamification might be to move away from mere behavioral data, and use other methods to discover their inner workings.

> "Today, the main challenge has become to work against the grain of existing preconceptions of gamification (be they apocalyptic or utopian), established by evangelists, critics, industry practices, and mass media reporting." - Deterding (2014a)

Chapter 3

Methodology

The methods that were applied to answer the research questions, including the design of the experiment, are presented in this chapter. Figure 3.1 shows an overview of the five parts of the study design. In the first two steps, the participant is required to sign the research consensus and fill out the pre-experiment survey. Then, participants will complete a quiz once a week that will test their learning knowledge in a university course. After each quiz, for a total of three weeks. Each week, the participant will receive their score on a ranked absolute leaderboard with other students. After the quizzes, the post-experiment survey answers how participant perceived the leaderboard and how it effected them. Finally, two focus groups were planned to discuss the results and to explore the research questions further.

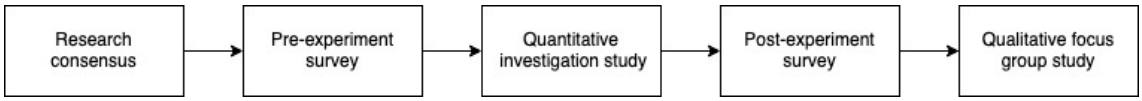


Figure 3.1: A summery of the research design

3.1 Participants recruitment

For this experimental study, university students were recruited from Utrecht University. This was done to perform the experiment in an educational setting, and due to convenience. The assumption was made that students enrolled in a game-related course would be more interested to participate in this study. Due to this, participants were recruited from the master course "Serious Gaming". This decision was made in agreement the course coordinators. The course is a non-compulsory course lasting 8 weeks with 61 students enrolled. Considering this study as a pilot for this topic, it was assumed that recruiting 30 participants (about half of the group) would be both realistic, and sufficient.

3.2 Participants incentive & quiz rational

As similar studies had trouble in finding a sufficient amount of participants, precautions were made that were meant to increase participation. As this leaderboard study allows for any action to be performed on which to score a participant, it was decided to create a knowledge-quiz about the course content in the assumption that students following the course would be interested in being quizzed about the content of the course. To ensure correctness and understandably of the questions, the knowledge questions were made in co-creation with the course coordinators.

There was no reward or prize incentive during this quizzes, as this could influence the results on how the quiz and leaderboard itself were perceived. However, a small prize incentive (chocolate or ice-cream) was awarded to those who completed the post-experiment survey or participated in one of the two focus groups, as these parts require a greater time and energy investment and were believed to be less valuable to the participant.

Finally, it was assumed that some participants might forget about the quiz or to perceive their time to be insufficient to complete it. These issues were addressed by allowing the quizzes to be made at the beginning of a lecture, where the lecture remained on-hold until the quizzes were completed. All participants received an email at the beginning of the lecture to start the quiz, regardless of their presence, to allow those not present at the lecture to complete the quiz. An example of this e-mail can be found in appendix 8. In the scenario where participants would not have completed the quiz, the participant would receive up to two reminders in their personal email.

3.3 Materials

Surveys and quizzes: The research consensus, pre-and post test survey and the three quizzes are built using the survey software tool Qualtrics, as it is the recommended and licensed tool by Utrecht University. Qualtrics complies with current GDRP regulations and allows the surveys to be completed on both mobile devices, laptops and computers.
Leaderboards: The leaderboards were created using the online leaderboard maker *keepthescore.co*.
Announcements and reminders: The reminders will be sent to the participants personal email using Outlook.
Analysis The built-in software of Qualtics was used to explore the data. Python 3.1 was used to perform the quantitative analysis. The qualitative part of the study was analysed using Nvivo 12.

3.4 Design

The five parts of the study design will be elaborated upon further in this section.

The first qualitative part is similar in setup to other research, and consists of a knowledge quiz in which the participants are scored and ranked on a leaderboard. The study follows a between-subject design, exposing half of the participants to a leaderboard showing students from their project group which were considered to be in direct proximity of the participant. The other half will receive a leaderboard with people not in their project group, considered outside of their proximity. This split is visualized in figure 3.2. Based upon the maximum group size mentioned in the course as five, this will be the amount of people displayed on the leaderboard.

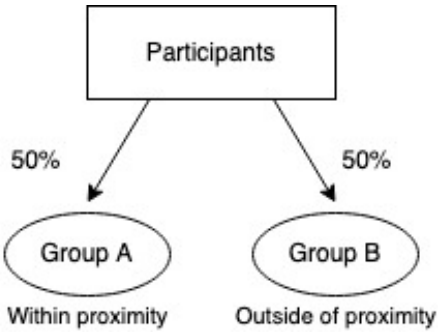


Figure 3.2: Diagram showing the split between groups A and B

The results of the quantitative research will give both behavioral data and information collected from the surveys. In the second part of the research, the results of the first part of the research will be discussed in two focus groups. This allows a reflection on the quizzes and leaderboard on their own and a more through interpretation of the results found in the first part of the study.

3.4.1 Research consensus

Precocious measures were taken to ensure the safety of participants with the requirement of the informative research consensus. The consensus emphasized the voluntary nature of the study and the ability to withdraw from the research at any given time, in which case the data would be deleted. Participants had to accept being ranked to others students in order to participate and were specifically instructed to contact the researcher in the case of any questions or concerns. At the end of every quiz, the contact information of the researcher was repeated with the same message. The research consensus can be read in appendix 8

3.4.2 Pre-experiment survey

The pre-experiment survey asks participants for demographic data, the amount of time they believed to invest in the course and their grade estimation at the end of the course. The complete list of all questions in the pre-experiment survey can be viewed in appendix 8.

3.4.3 Quizzes and leaderboards

Participants performed a scored quiz every week for a total of three weeks. This number is chosen to measure a possible change in effect when the experiment would be run for multiple weeks. Three weeks was believed to be sufficient to evaluate these effects, given the duration of the course. The scores were placed on a ranked absolute leaderboard. The scores reset each week in order to give participants a new chance each time.

To study the effects of social comparisons, the study follows a between-subject design, exposing half of the participants to a leaderboard with participants in their direct proximity (within their course-group) while exposing the other half with a leaderboard that only shows participants outside of their direct proximity (outside of their current course-group). Within the course, students are required to create groups of five for the mandatory group project. Based upon the maximum project group size mentioned in the course, five people will be the amount of people displayed on the leaderboard. With 30 participants in total, and five participants per leaderboard, there will be a total of six groups.

The leaderboard design within the experiment is purposefully build to make a social comparison more likely to occur to study its effects. As such, the leaderboard shows the full first and last name of the participants on the board, as well as the exact amount of points received and their current rank in comparison to the others on the board. The visual theme of the board is called "Pixel", which is chosen for being both clear and because of it's game-like appearance. The participant ending first is visually emphasized by a trophy-symbol next to the name. Figure 3.3 shows a fictitious example of what the leaderboard looked like during the experiment.

Figure 3.3: A fictional example of the leaderboard

3.4.4 Post-experiment survey

The post-experiment survey is an online survey with 61 questions. The majority of questions are multiple-choice (50) in order to reduce the required time-investment for participants. For the same reason, most (9) open-ended questions are skip-able. The survey itself is divided into four sections. These are; Motivation (1), Social Effects (2), How the Leaderboard is Used (3), and Technical Difficulties and Feedback (4). The first section is aimed at finding the participant's general motivation to aim for high grades and study and the course material. To find the motivation of the student, the validated Science Motivation questionnaire II (Glynn et al., 2011) was used as the main inspiration for the questions in the survey. Modifications were made to the questionnaire to fit the current research. The first alteration was in the word-usage, which was transformed to fit the context of the current research. For example, the first statement in the SCII was "Getting a good science grade is important to me", which has been altered to have the word "science" removed, making the first question "Getting a good grade is important to me". Another example is question 2: "I like to do better than other students on science tests", which was altered towards "I like to do better than others on tests". As the SCII is extensive and only a general sense of motivation is required in this research, the second alteration was to make the questionnaire shorter. Questions that were removed were seen as either irrelevant (such as "I will use science problem-solving skills in my career") or questions too similar to others questions in the context of this research (such as "I believe I can master science knowledge and skills" and "I am sure I can understand science"). Questions that scored high on reliability in the factor analysis of (Glynn et al., 2011) were preferred. On average, two questions from each motivational factor (Career, intrinsic, grade or other motivation) were included in the questionnaire with a reliability factor of 0.68 at the lowest. With two highly reliable questions of each motivational factor, the shortened questionnaire was believed to be sufficient to give an indication of the participants general motivation for this research. The second part of the survey was about what social effects the leaderboard may have had, such as comparing scores and the judgement of others. The third is how the leaderboard was

used and how it effected the participant, while the final section is aimed at finding any technical difficulties that may occur during the experiment as well as giving the participant the option to give feedback.

3.4.5 Focus groups protocol

The aim of the focus groups was to find a deeper understanding of the influence of the leaderboard on people. This was done by discussing the findings discovered in the experiment phase. The questions were meant as guidelines, while the main goal was to find the reasons that certain effects can occur. The insights are structured based on the research questions. The protocol and analysis is further explain in chapter 5. The main guiding questions were:

- 1. How would you describe your experience making the quizzes and leaderboard?
- 2. (How) did the leaderboard effect you? Was this a positive or negative experience?
- 3. How would you describe the social effect of the leaderboard?
- 4. If a leaderboard would be implemented within an education context; What requirements should be considered?

Using the methods described, we conducted the quantitative study and focus groups. The next chapter will present the results of the quantitative study.

Chapter 4

Quantitative investigation on the effects of the leaderboard through proximity

The results of the experiment and focus groups are presented in this chapter. It is first explained how the analysis is performed and how data has been handled. Then, the quantitative results are presented and summarized, followed by the results of the semi-structured focus groups. The insights are structured based upon the research questions from chapter 1.

4.1 Analysis

The survey responses and participation are analysed as quantitative data. All data, including the participation rate, has been merged into one file and explored using Qualtrics built-in analysis tools and were confirmed and tested in Python R.

The recorded focus group sessions were first transcribed using Otter.AI voice to text recognition software. The text was then edited to filter out mistakes made by the software and to anonymize the conversations. Then, a Thematic Analysis was performed to find the most prominent patterns from the focus groups. An inductive method to find patterns in the data, based upon the research questions, was applied.

Important parts of the conversation were then manually highlighted. Next, the thoughts and opinions were summarized for each research question together with reoccurring topics that are believed to be relevant for the research. Information that was considered not relevant enough for the scope of the research is placed in the future work section. This information, together with the prior background information, was used to discuss and conclude the initial research questions.

4.2 Data checking, screening and pre-processing

Duplicate data and entries with mostly missing data were removed from the data set. When duplicate data was found, the most complete entry was kept. Quizzes that where started but not completed were kept if one question was answered and no duplicate existed. Unanswered questions were scored as “I am not sure”, resulting in 0 points. On average, seven entries were removed per survey. During the analysis of the pre-and post-survey, only participant that answered both were included in the data set. The first quiz had the most entries containing missing data, while the final quiz had the largest amount of (removed) duplicates. Table 4.1 shows a detailed overview of the removed data for each part of the research.

Removed data	Concensus	Pre-test	Quiz 1	Quiz 2	Quiz 3	Post-test
Duplicates	5	3	5	5	9	3
Missing data	2	2	4	1	0	1
Total removed	7	5	9	6	9	4

Table 4.1: Removed data per research part

4.3 Response rate

Out of the sixty-one students that signed up for the Serious Gaming course, 70.5% (N=43) signed the research consensus to participate in the study. Out of this group, 76.7% (N=33) would continue to complete the first quiz. The second quiz had the largest amount of participants (N=36), while the final quiz had the lowest participation rate (N=30). Table 4.2 shows the amount of participants for each part of the research (after the duplicates and incomplete answers were removed).

Research part:	Consensus	Pre-test	Quiz 1	Quiz 2	Quiz 3	Post-test	Focus Group
Responses:	43	37	33	36	30	34	8

Table 4.2: Responses per research part

A total of 27 participants filled in both the pre and post-test survey. Five participants did fill out the pre-test, but not the post-test survey. Two participants did fill out the post-test, but not the pre-test. Out of the participants that made the first quiz, 81.8% (N=27) participated in every quiz. Some participants skipped over only one quiz, with the most common reason mentioned being "lack of time" and "Forgot to make the quiz" (both mentioned three times). Table 4.3 shows the number of participants that skipped over only one quiz, labeled as "unique".

Participation:	Quiz 1	Quiz 2	Quiz 3
Unique:	4	1	2

Table 4.3: Number of participants that skipped over a single quiz

4.4 Participants

Table 4.4 shows the demographic data about the participants. The prior knowledge level was reported to be similar between participants, with an average of 2.91 out of 5. The male/female ratio is close to equal, with some participants having reported not wanting to share their gender. The average age was close to 24, with an age variance of 10 years.

Male	Female	Knowledge level	Average age	SD age	Age variance
47%	44%	2.91/5.00	23.7 years	3.24	10.5

Table 4.4: Participant demographics

4.5 Quiz reception

Figure 4.1 shows the questions and given answers of participants on quiz and communication related questions. Participants reported not having prepared for the quiz (N=33, 97%). The main reason that was given was because the quiz-score did not effect the final grade, and that there was no other reward present. While making the quiz, most participants did try their best while making the quiz (N=26, 76.5%) and often either liked making the quizzes (N=18, 52%) or did not like nor dislike making them (N=10, 29%). The communication and questions were most often reported to be clear (N=20 , 58.8%) or very clear (N=9, 26.5%). On average, participants were indifferent if the quizzes helped them to focus on the course, with the majority of 35.3% (N=12) remaining neutral and an almost equal divide between agreement and disagreement.

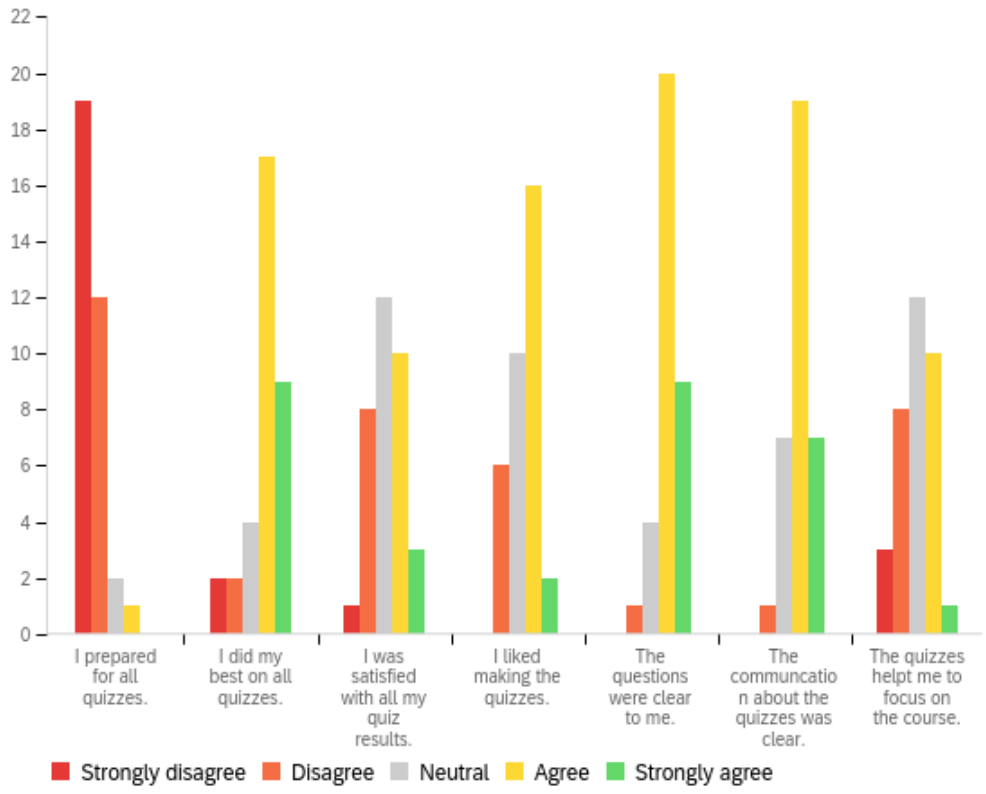


Figure 4.1: Answers to the questions on how the quiz was received

4.6 Exploring how the leaderboard is used

What makes a leaderboard interesting, how is it used, and what aspects are believed to be important? This chapter provides answers to research question 1, 'how is the leaderboard used?'. The results of exploring what information participants look for when opening the leaderboard and which aspects influence time spend studying are described in this chapter.

4.6.1 What information are people looking for on the leaderboard?

When send out, the leaderboard was looked at by 88% of participants (N=30). After opening the leaderboard, 67% (N=23) reported looking at the performance of others and using that information to evaluate their own performance. Figure 4.2 shows how participants answered each question.

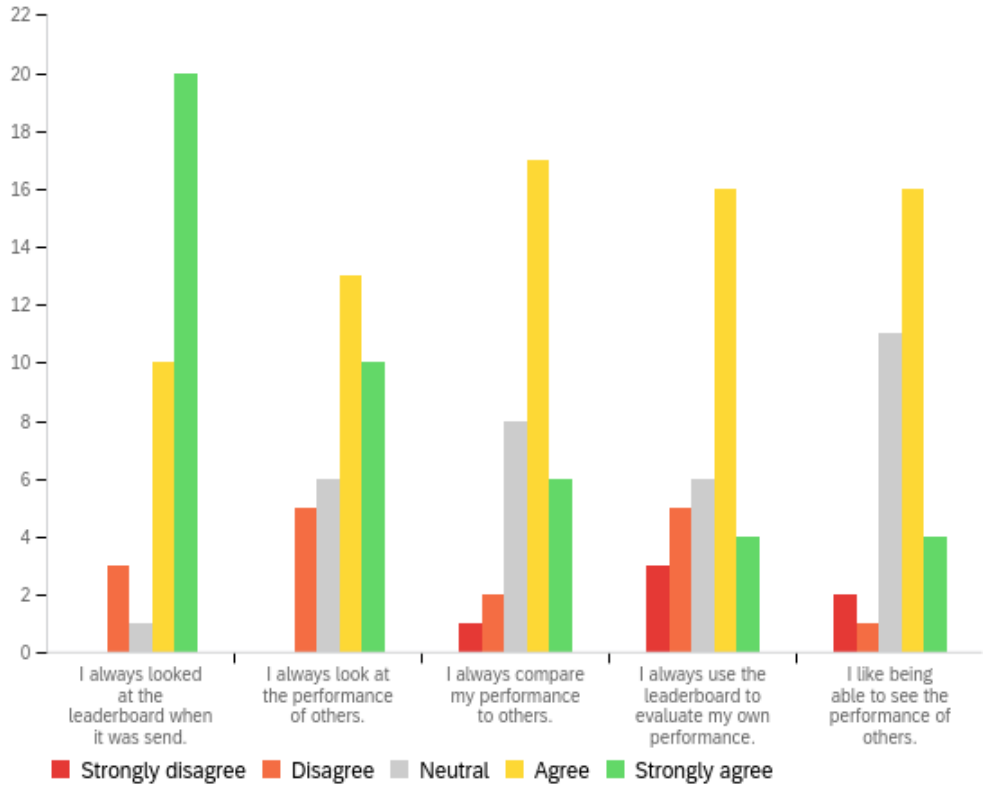


Figure 4.2: Answers to the questions about looking at the leaderboard

The most important reason to look at the leaderboard was to evaluate their own performance (56%, N=19). Not a single participant reported that they were most interested to see who had performed worse then them. Figure 4.3 shows the given answers.

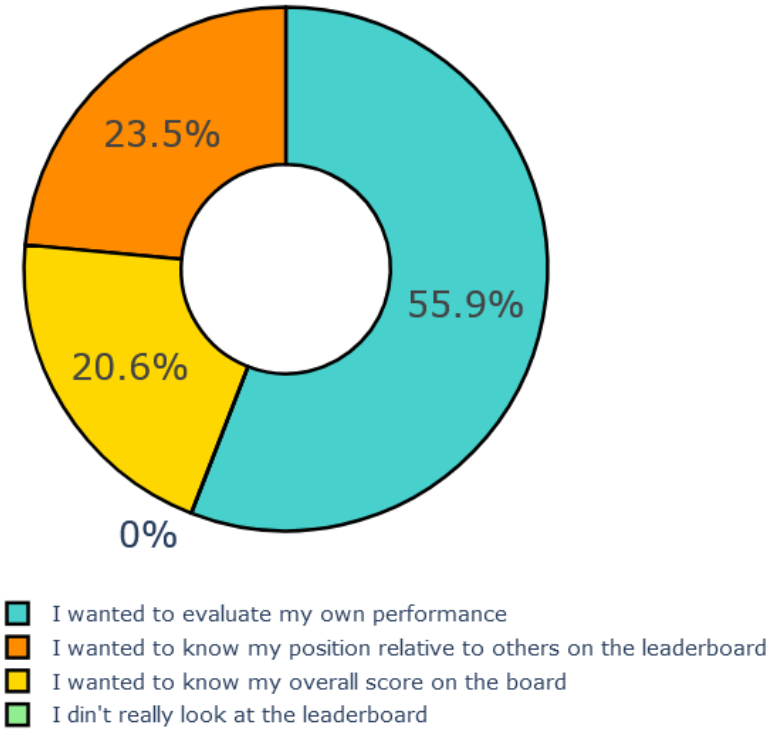


Figure 4.3: Answers to the post-experiment survey question: "What was your main reason to look at the leaderboard?"

When evaluating their own performance, participants reported that the most interesting aspect of the board was to see how other people performed that the participant knew (41 %, N=14), and seeing who had performed better then them (26 %, N=9). Many reported not being interested in the performance of others (32 %, N=11). The given answers are visualized in Figure 4.4.

4.7 RQ4: How are people (emotionally) effected by the leaderboard?

What effect did the leaderboard have on participants? The effects are explored and divided into the categories of score, motivation, emotion and social effects. This chapter answers research question 2, 'How are people effected by the leaderboard?'.

4.7.1 Which aspects of the leaderboard influence time spend on studying?

Figure 4.5 shows which aspects of the leaderboard were believed to increase or decrease the amount of time spend studying. The two main factors reported to lead to an increase were being able to see their previous ranking (N=18, 28.2%) and adding a reward that is related to the score (N=18, 28.2%). Other factors that would increase study time are the ability to see who performed better then them (N=15, 44.1%), if people close to the participant would see the leaderboard (N=12, 35.3%) and being able to see the performance of others that the participants knows (N=18, 28.2%). Few reported a decrease on time spend studying, with the visuals of the leaderboard and the ability to see who performed worse then the participants (both N=3, 9.4%) having the highest chance, followed by the ability to select who is on the leaderboard (5.9%, N=2).

The most common answer was that certain aspects had no effect on the time spent studying. Participants reported that the visual aspects, being able to see who performed worse and how many people would see the leaderboard had no effect.

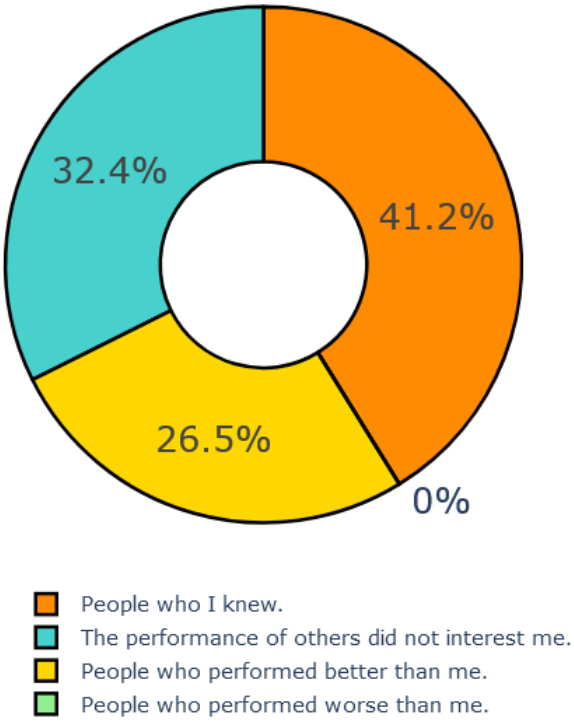


Figure 4.4: Answers to the post-quiz question: "What interested you the most?"

4.7.2 Score

Did participants like seeing each others score? 58.8% (N=20) liked being able to see the performance of others (with 12% strongly agreeing), while 32% remained neutral. Only 26.5% (N=9) liked that others could see their score, while the majority of 41.2% (N=14) did not agree nor disagree. 35.3% (N=12) disliked that others could see their score, out of which 14.7% (N=5) strongly disliked it. 53% (N=18) would feel proud when they achieve a high score with their name next to it. 32.4% (N=11) would feel ashamed when a low score would be shown with their name next to it.

4.7.3 Motivation

Were participants more motivated because of the board? 47% reported feeling more motivated due to being compared with others, while 30% disliked it (12% strongly disliking it). 30% reported being motivated for the quiz because others would be able to see their score, while 53% disagreed with this statement (24% strongly disagreed).

47.1% (N=16) of participants liked the quizzes more because of the leaderboard. 35.3% (N=12) did not agree nor disagree, while 17.6% (N=6) preferring to have made the quizzes without the leaderboard. 20.6% (N=7) reported that they would not have participated in the quizzes if the leaderboard was not present. 67.7% (N=23) would still have made the quiz without the leaderboard.

18% would have liked the quizzes more without the leaderboard, while 35% remained neutral on this and 47% liked it more with the leaderboard (15% strongly disagreeing). A majority of 67% would still have participated without the leaderboard (41% of participants strongly disagreed with this statement), while 12% was not sure.

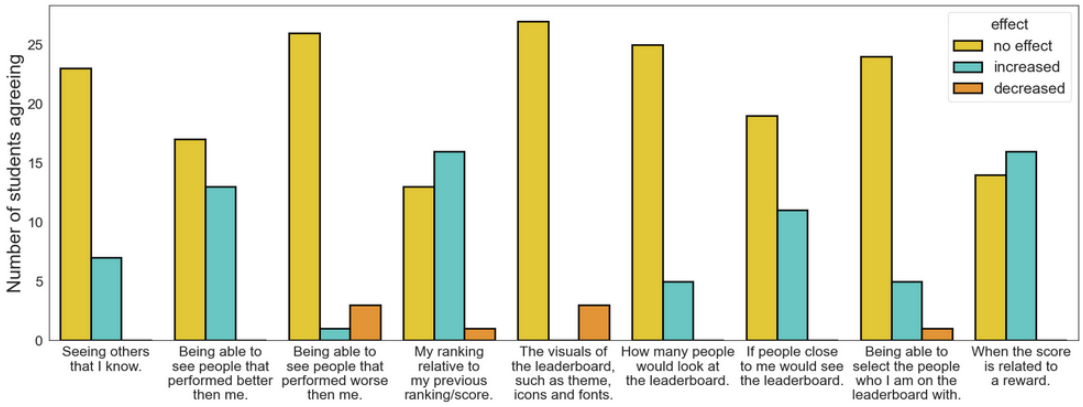


Figure 4.5: Aspects that increase, decrease or have no effect on the time spent on studying.

4.7.4 Emotional state

How did the leaderboard effect the emotional state of the participant? 17.7% (N=6) reported that the leaderboard made them anxious, while 23.5% (N=8) felt excited by the board. 35.3% (N=12) reported that the leaderboard made them question their own abilities, while the same amount of people reported they felt more confident in their abilities due to the leaderboard. 50% (N=17) reported that the leaderboard made them feel as if they had to perform.

When running a pairwise correlation test, the ranking performance showed a slight positive correlation with reported anxiety, with a correlation coefficient of 0.4377.

4.7.5 Social

What are the social effects of the leaderboard?

20.6% (N=7) of participants talked about the leaderboard with the people who were on their board, while 50.0% (N=17) never talked about it within their group. 23.5% (N=8) Talked about their score with people not on their board, while 61.8% (N=21) did not. Only 14.7% (N=5) talked about the leaderboard within their project group. The post-experiment survey asked in an open question with whom the leaderboard was discussed, if at all. The results from this question is that most talked about their score with people from their project group (41.2%, N=14), while others talked about it with friends from class (20.6%, N=7) or other friends (17.7%, N=6).

29.4% (N=10) likes to know how others in their project group performed, while 32.35% dislikes it. 20.6% (N=7) believed that seeing the scores of others on the leaderboard influences how they think about them. Participants did not believe that people with a low score are lazy (97%, N=33), while 35.3% (N=12) did believe that those with a high score are hardworking. 29.4% (N=10) was afraid to be judged by others.

When running a pairwise correlation test, the ranking performance showed a slight negative correlation towards believing others are lazy, with a correlation coefficient of -0.3121. Performance ranking was similarly negatively correlated with liking that others could see their performance with a value of -0.3995.

4.8 Can Social Comparison theory be used to understand leaderboards?

In social comparison theory, people were significantly effected in both their ability estimate and mood. Can the same effects be found in the case of leaderboard? If so, does proximity then have a similar effect? This chapter answers research question 3, 'can social comparison theory be used to understand the leaderboard?'.

4.8.1 Ability estimate

Is there a relation between the score of the participant, and their own estimation of the grade they will receive at the end of the course? Figure 4.6 shows the average score of the participant on the X-axis, while showing the reported difference in expected grade on the Y-axis. The graph shows that most participants did not change their grade estimation between the pre-and post-survey, and that there seems to be no relation between score and grade estimate. When running a pairwise correlation test, the score of the participants was slightly negatively correlated towards questioning their own capabilities, with a correlation coefficient of -0.3792.

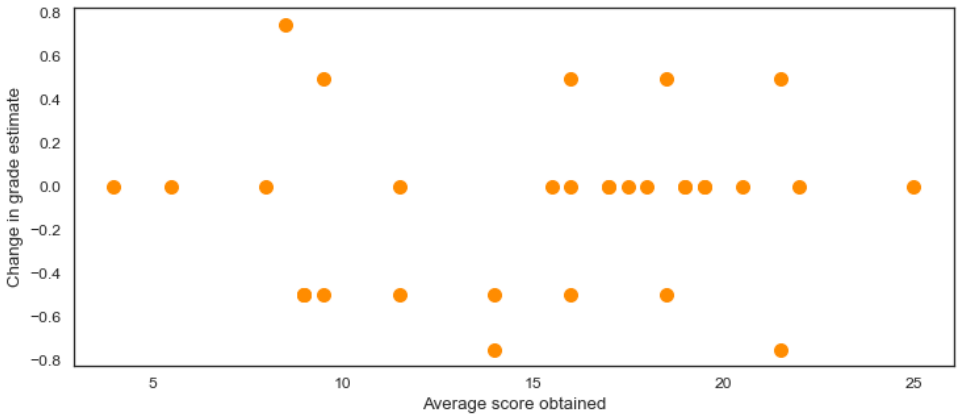


Figure 4.6: Scatter-plot showing the average change in grade estimation and average obtained score

The relation between the leaderboard position and the grade estimation is shown in Figure 4.7. It shows the average leaderboard position of each participant on the X-axis, with the difference in grade-estimate on the Y-axis. While most participants reported no difference in grade-estimate, the place on the board does seem to correlate with their grade-estimation. When running a pairwise correlation test, the ranking of the participants was negatively correlated to their believes in their own capabilities, with a value of -0.48 when all test results are included. When 0-values, indicating the absence of a change in grade estimate, are removed, we find a correlation coefficient of -0.69.

4.8.2 Proximity

Are the effects of the leaderboard strengthened by proximity? Figure 4.8 shows the emotional effects of the leaderboard, divided between group A and B. The graph shows that the within-proximity group felt both more excited and anxious due to the leaderboard. They felt more pride with a high score, while also feeling more pressure to perform, and being more often afraid to be judged.

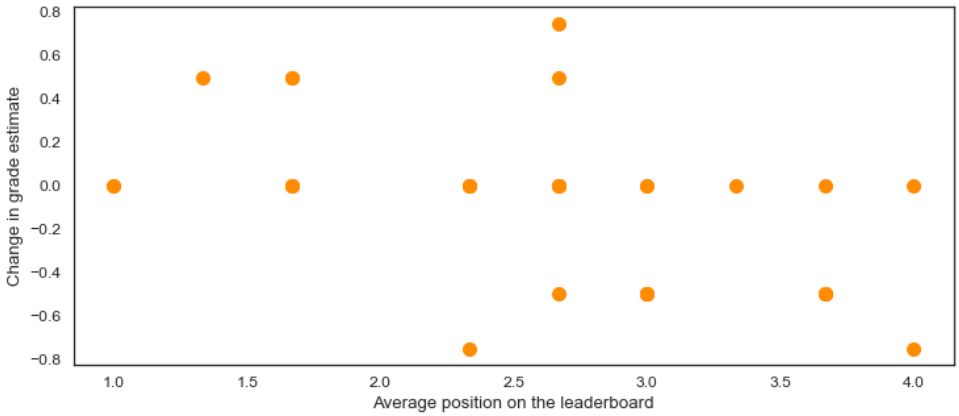


Figure 4.7: Correlation between change in grade estimate and position on the leaderboard averaged over all three quizzes.

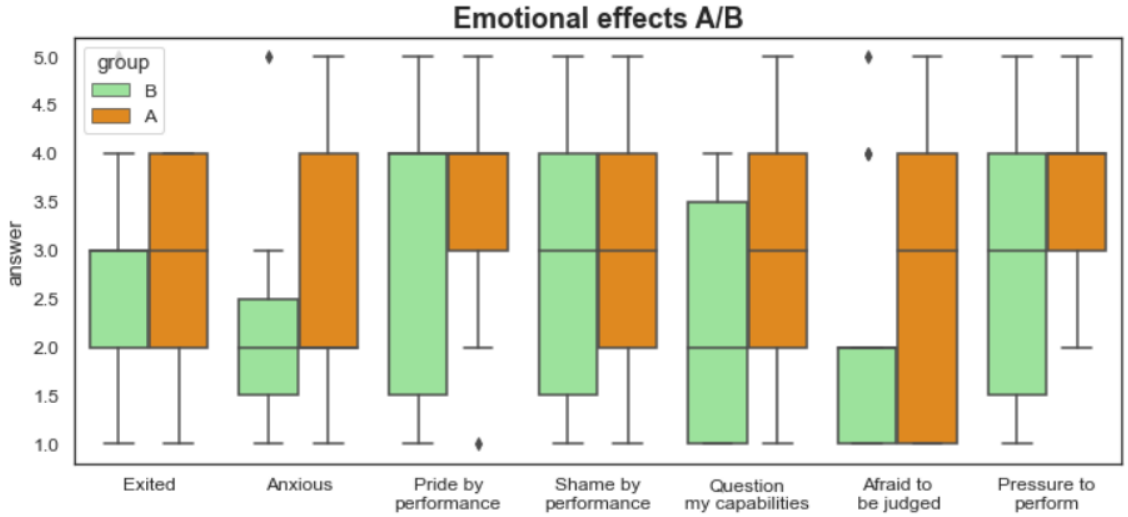


Figure 4.8: Emotional effects of the leaderboard

Figure 4.9 shows the social effects of the leaderboard, divided between groups A and B. The within-proximity group reported to like to seeing the scores of others more often. They were much more likely to talk about the leaderboard with people on their board. Group B was more likely to talk about the leaderboard with people who were not on their board.

4.9 Requirements for the educational context

This chapter answers how the leaderboard can be implemented into an educational context, while emphasizing the positive effects and minimizing the negative effects. It answers which negative effects are believed to be likely to occur gives some mentioned solutions. This chapter answers research question 4, 'What requirements make the leaderboard applicable to an educational context?'.

4.9.1 What positive effects can occur and how can these be emphasized?

When asked about the positive effects of the leaderboard, the most common answer was that it was a fun addition and that it helped participants to practice (N=10). Participants reported that more

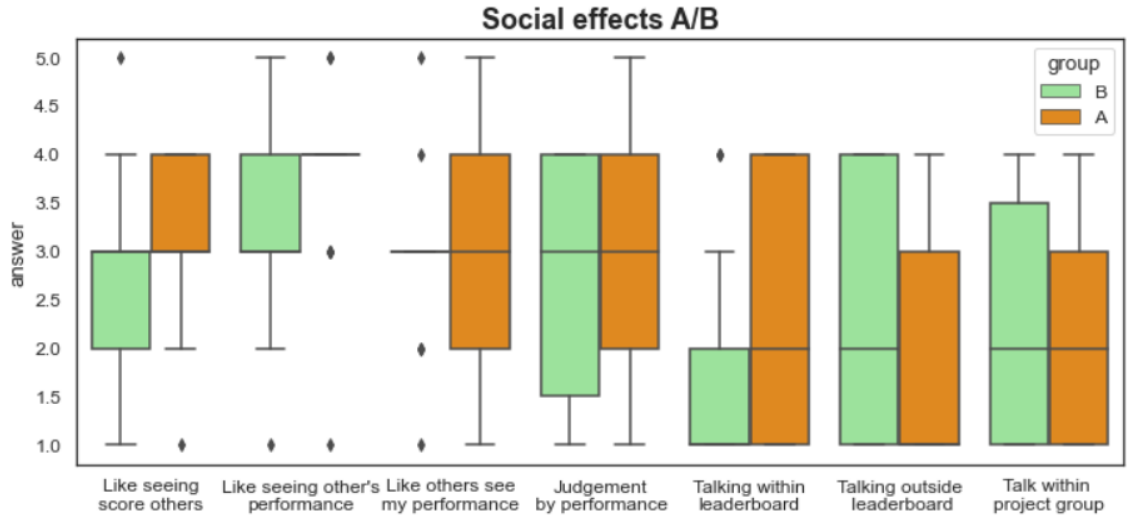


Figure 4.9: Social effects of the leaderboard

positive effects could occur if more detailed feedback was given, such as showing the difference in score in comparison to the previous week (N=3). Participants also mentioned that making the leaderboard an opt-in solution could be beneficial, as well as making it anonymous.

In the general feedback section (open questions), participants reported that more detailed feedback would be the most wanted feature that was missing in the current experiment setup (N=11). Participants mentioned wanting to know which question they got right and wrong, as well as showing the percentage of people that did get that question right. Other details mentioned were the mean of scores and an arrow indicating if somebody had risen or dropped in comparison to the previous week.

The leaderboard itself could be improved by displaying more participants (N=4), as some wanted to see how the group in general performed. This feedback was given only by high-performers. During the focus-group session, top performers explained that they felt the urge to know if they were the best overall, or just from their group. They thought that it could be more interesting to be put on a leaderboard with others from similar backgrounds, as some felt they might have had an unfair advantage because of their prior knowledge. A possible solution mentioned was to give participants a total top 3 of the whole class to satisfy this urge.

Low-performers were more interested in detailed feedback, and mentioned that showing only their own improvement over last week would have been interesting to them.

One participant would have liked it if the theme of the quiz could be altered by the participant itself.

4.9.2 What negative effects can occur and how can these be avoided?

When asked about the negative effects of the leaderboard, the most often mentioned negative effect is the demoralization of lesser performers (N=13). Participants were often afraid that it could increase stress and anxiety for pupils if not implemented correctly (N=9). Fear of bullying (N=4) was also mentioned, as well as different treatment of students based upon their score (N=2). One participants believed it could increase the chance of unwanted behaviour such as cheating.

Possible given solutions were to not show the last names, or to make the leaderboard anonymous (N=3).

4.10 Reported technical difficulties and feedback

The communication about the quiz and the clarity of the presented information on the leaderboard were both reported to be excellent, with only one participant reporting it being unclear. The main reasons to not have made one or more of the quizzes was lack of time, being busy with other things, or having forgotten to make the quiz. One participant reported wanting to prepare for the quiz before making it, and then forgetting about it.

4.10.1 Feedback

The feedback section allowed participants to give feedback that they wanted to share before the end of the survey. The positive and negative effects of the leaderboard and the recommendations are added to earlier questions about these topics. Two participants reported that the researcher did a good job, while two others reported that there were some grammatical errors in the survey (2). Another mentioned that five questions was a good amount of questions for a quiz, and that one participant was mainly motivated to help the researcher and not interested in the research itself.

Building upon the findings, we conducted focus groups following the protocol described in section 3. The next chapter presents the results from the focus groups.

Chapter 5

Qualitative focus group study

This chapter presents the results of the qualitative focus group study. The aim of the focus groups was to find a deeper understanding of the influence of the leaderboard on people. This was done by discussing the findings discovered in the experiment phase. The questions were meant as guidelines, while the main goal was to find the reasons that certain effects can occur. The insights are structured based on the research questions. The main guiding questions were:

1. How would you describe your experience making the quizzes and leaderboard?
2. (How) did the leaderboard effect you? Was this a positive or negative experience?
3. How would you describe the social effect of the leaderboard?
4. If a leaderboard would be implemented within an education context; What requirements should be considered?

5.0.1 Participants

Two focus groups were planned. The session was conducted on Utrecht University Master students (N=8, 8 female) that had participated in the experiment phase of this study. It was expected that the first focus group would create a better understanding of the experience from the perspective of the student, and that useful feedback on the results of the study would be collected. The second group was chosen to verify earlier findings and to gain further understanding in the possible wide variety of experiences that may have occurred with a more narrowed focus. In terms of performance during the experiment, the first group consisted of one high performing student, one average performing student and two students who had performed at the bottom end. The second had two students who performed at the top end and two who had performed at the bottom end. The backgrounds were Dutch (N=4), Spanish (N=1), Russian/American (N=1), German (N=1) and Chinese (N=1). Participants were all part of Group A (within proximity) due to convenience sampling.

5.0.2 Procedure

When starting the session, participants were first required to consent to record the session which was done vocally. Next, the ground rules of the focus group were explained. Participants were told that all experiences are unique and all experiences are equally interesting, to give others time when they are talking, and to try to be respectful towards each others privacy even when the focus group had ended. Next, participants were provided information regarding the background of the study,

as well as the session's goals and outline, and an introduction about the main contradictions in the leaderboard research. Participants were asked individually to talk about how they experienced the leaderboard and how it effected them or others. They were also asked to elaborate on their answers given on the post-experiment survey. Special attention was given to the causes of certain experiences. For instance, if according to one participant the leaderboard created more motivation, but the overall experience was explained as a negative, this seemingly contradictory information was elaborated upon. Participants were encouraged to share their thoughts that came up during the session. After the rounds of personal experiences, the participants were asked as a group to elaborate on the results from the survey and to share their explanation of the results. After the elaboration round, the final round focused on the broader question of how the leaderboard should be implemented in educational context, and what requirements are believed to be necessary to avoid any unwanted effects. The total duration of the focus groups were each about 1 hour and 15 minutes.

5.0.3 Analysis

The recorded session were transcribed using Otter.AI voice to text recognition software. The text was then edited to filter out mistakes made by the software and to anonymize the conversations. The transcriptions were then imported into Nvivo 12. Next, a Thematic Analysis was performed to find the most prominent patterns from the Focus Groups, based upon the related research questions following an inductive method. The effect the leaderboard can have on an individual is divided into the categories that influenced this effect according to the participants (see 5.0.5. This information, together with the prior background information, was used to discuss and conclude the initial research questions. Information that was considered not relevant enough for the scope of the research is placed in the future work section.

5.0.4 Exploring how the leaderboard is used

Similar to the results from the survey, the interest to look at the leaderboard comes from the interest to self-evaluate, together with general curiosity. Figure 5.1 shows the reasons that were mentioned during the focus group, divided between curiosity and self evaluation. The figure shows that participants were curious about their score, how others scored in relation to them and seeing if they were not at the bottom of the board. The latter was due to not wanting to stand out (in a negative sense) towards their peers. In terms of self-evaluation, participant said that validation for understanding the course material, checking if the course knowledge was understood and seeing whether or not they were a top performer interested them.

When looking at the leaderboard, prior leaderboard position changed what information was looked at on the board. Bottom-end performers wanted to see if they performed better in relation to previous weeks and often scanned through the names of better performers, while top performers wanted to know if they were still on top, without looking at worse performers.

5.0.5 The effect of the leaderboard

The effect of the leaderboard is divided into the categories of behavioral, social and emotional effect. In this research, *behavioral* describes how the behavior of an individual changes due to the leaderboard, by for instance increasing time spent studying. The *social effects* are how the leaderboard (score, ranking) influences how others are perceived (judged, looked at differently)

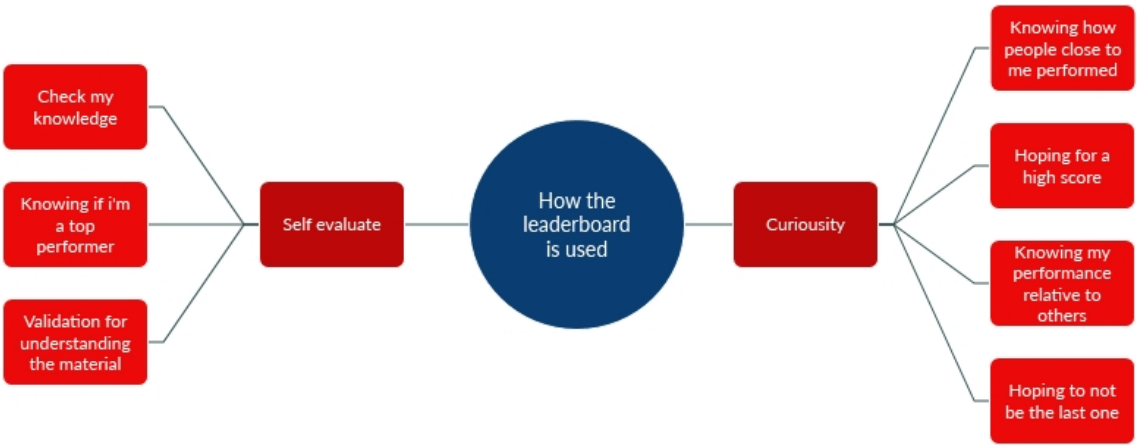


Figure 5.1: Mind-map of why and how the leaderboard was used

and if the leaderboard was talked about within their social environment. The *emotional response* is how a person felt in response to the leaderboard, for example, feeling stressed or joyful. Finally, influential factors for these responses are elaborated upon at the end of this section.

5.0.6 Behavioral effects

The behavioral effects were not felt strongly during the experiment, as the scores and ranking had no effect on the final grade and were not directly related to a reward. It was mentioned that the behaviour of a participant could change based upon their score over multiple weeks. Participants that performed well over a longitude of time mentioned to keep wanting to perform well to fit the created expectations about them. Some of the participants that did not perform well mentioned to try to figure out what they are doing incorrectly, by asking how others study, learn from them, and change the way they study. However, participants believed that if the performance would not improve over multiple weeks after these efforts, they would become demotivated and discontinue trying to compete.

> "I can ask the people like, Hey, what is your system? How you doing this? How did you manage to learn these things?" - *Participant 1*

5.0.7 Emotional effects

The emotional effects are mentions of, for instance, pressure, anxiety, happiness or joy. When social and emotional topics overlap, the emotional category is chosen when social events are merely imagined, without actually taking place. For instance, if there is an increase in anxiety due to the fear of being bullied without this bullying taking place, it is regarded as an emotional effect ('Social anxiety due to the presence of the leaderboard'). If the bullying would take place, it is regarded as a social effect ('Social action caused by the presence of the leaderboard').

> "The problem for me is that knowing that I'm going to be judged, just puts out pressure, even while I'm making the quiz. I started doubting myself." - *Participant 1*

An overview of the emotional responses, divided between positive and negative reactions, are shown in figure 5.2. The chart shows how the emotional effect of the presence and the visibility of the leaderboard were perceived. The positive effects are increased trust in their capabilities,

described as a "small ego boost", as well as a fun experience with the group and being able to score high. The positive experiences were mainly created due to a good group dynamic, good performance, and the scores not influencing the final grade. On the negative side, there are feelings of annoyance, anxiousness and losing the believe in being able to perform. These feelings were mainly created by bad performance, the possibility of being judged by others and prior negative experiences. Participants agreed that the negative effects were only felt slightly, due to the quiz not having any other influence on the course grade.

> "I'd say I don't really care about the scores. Because we are not in competition, we're not competing for something like rewards, so it doesn't really have much impact." - *Participant 4*

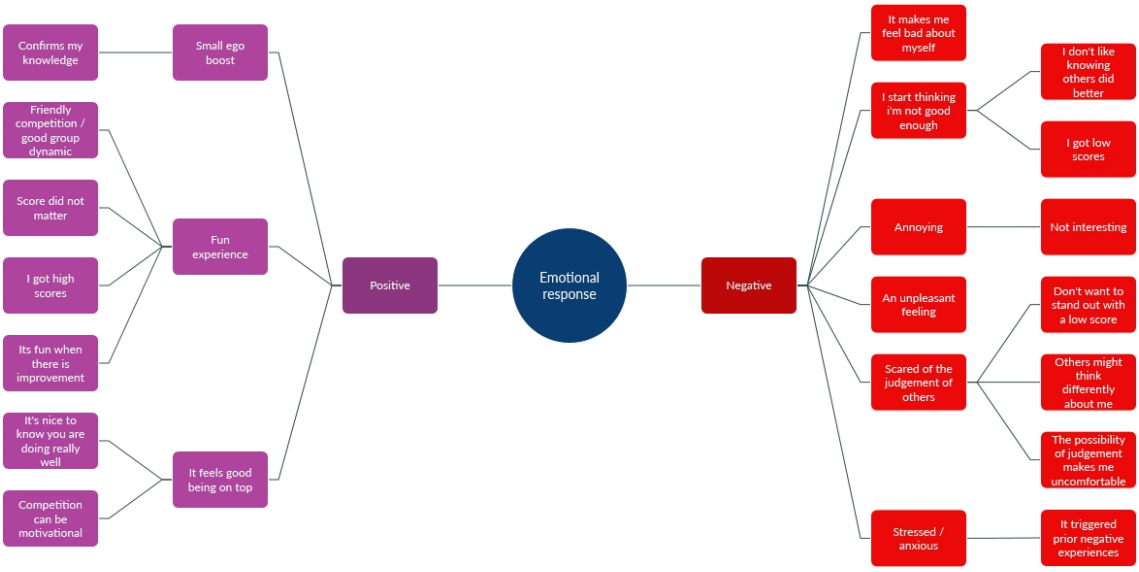


Figure 5.2: Mind-map of the emotional responses and their explanation

5.0.8 Social effects

The social effects are all effects caused by the group. For example, mentions of an increased ability estimate due to outperforming others falls under this category. Discussing the leaderboard and thinking about others differently due to the leaderboard score or ranking, is also categorized under social effects.

In terms of social interactions, the leaderboard was not discussed often. Top performers did not discuss the board with others on their board, as they believed it could be perceived as "boasting" or "bragging". They did talk about their score with others that they knew from class that had scored similarly well, challenging them personally. Performers on the top 3 on the board sometimes mentioned their score in friendly competition to better performers on the board, while bottom-end performers did not talk about their score at all. Bottom-end performers generally mentioned they did not feel the urge to talk about their score and did not care about the score "too much", as they had not prepared.

In terms of the opinion of others, top performers mentioned that they did not think differently about others who performed worse then them. The reason was that they only looked at the board to

see if they were again a top performer while disregarded the other information, even if they knew others on the board. Middle and bottom-end performers did look at the performance of others, being able to remembering (rare) low score of specific people. They feared others to look at them differently if the low scores continued for multiple weeks, a possibility which created stress and anxiety.

> "I don't think I would judge other people too much. I do remember that one person had a 'one' once. But I feel like maybe they just didn't really care about filling in the quiz then. But I was really glad that I never finished last because I feel like that would have affected me more." - *Participant 2*

5.0.9 Influential factors

The influential factors are the factors that increased or decreased the effect the leaderboard had on the participant. In this research, the main focus is on the ability estimate and proximity. Figure 5.3 shows a mind-map overview of the mentioned influential factors during the focus groups. The mentioned topics are elaborated upon further in this section.

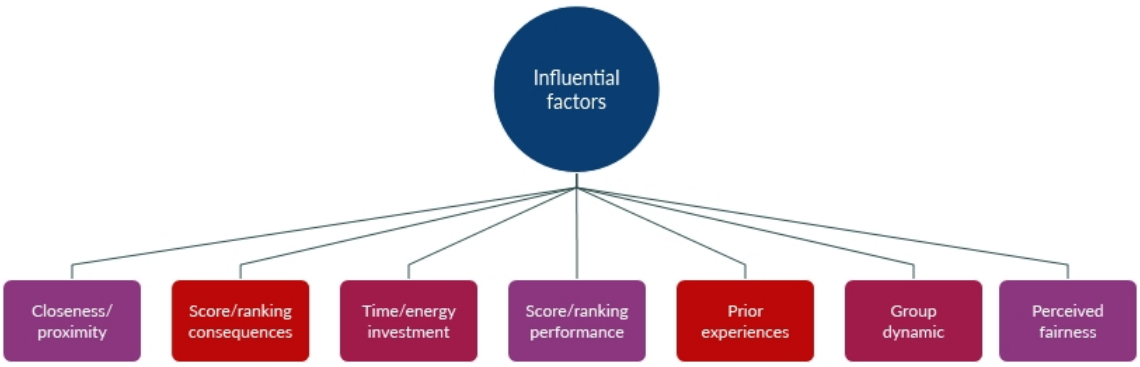


Figure 5.3: Mind-map of the mentioned influential factors

Closeness: Being compared to others who are unknown, or hardly known, makes the comparison less impact-full then with people close to us. Participants mentioned that an insufficient (bad) grade could generally create disapproving thoughts, such feeling disappointed in themselves or questioning their personal ability, but that these thoughts or feelings can cease to exist when those close to the participant scored equally or similarly bad. Likewise, getting a high (good) score can feel much less impressive if somebody close to us scored even higher. Closeness was mentioned to be one of the most influential aspects.

> "I've definitely had situations where we both failed an exam and it was like a very small difference and we just ended up laughing about it." - *Participant 5*

Score/ranking consequences: When the score or ranking has no impact on the individual, be it with a reward or final course grade, the impact of the scores and ranking are mentioned to be greatly reduced. Participants mentioned that this creates a more playful environment, lowering the bar to participate. However, this also causes some to lose (all) interest in the scores and ranking. When the consequences of the score and ranking are increased, the positive and negative effects of the leaderboard can similarly be effected. The strength of this effect was believed to be based upon the perception of the impact of the consequences.

> "I will be like, grounded, if I will get a seven, as my parents believe that to be a low grade. And I was not allowed to get out of the house to play. So yeah. My life you know, that I do for some grade that is worth really nothing at the end?" - *Participant 1*

Time/energy investment: Regardless of how participants performed, the prior time and energy investment greatly influenced how they were effected after seeing the leaderboard. A bad performance was mentioned to be easier to cope with if there was no or little prior time or energy investment in learning. A similar decrease in impact was mentioned when performing well without prior time or energy investment, as the high score could now in some cases be perceived as mere luck.

> "I think the difference is that if you didn't really learn at all, it's like 'nice!'. If you learn a lot you think: I did good. But if you learn a lot and get low score, you think, what did I do wrong, what happened?" - *Participant 5*

Score/ranking performance: The score and ranking was mentioned as a great influence on the impact on the individual. Good performance on the board was generally regarded as rewarding and fun, while some only looked at their own score. Participants believed that continues high performance could create an expectation pattern that could possibly feel like a burden. Middle performance was regarded as fine, while bottom-end performance could change how a participants felt and behaved. To avoid ending in the bottom-end again, participants would first check what there mistakes were and if they could have done it right, requesting feedback and reviewing their answers. Participants said that if their score would not improve after having tried multiple times, they would lose the believe in being able to improve and stop trying.

> "I feel I'm too far away from the top. So it's like, no matter how much effort I put in, I won't get to there. So I somewhat gave up, but that's not a very good experience, I would say." - *Participant 4*

Prior experiences: While participants mentioned that the leaderboard in the experiment had only a minor impact, many related the ranking and score to prior experiences with social comparisons and competition. Positive prior experiences were fun competitions, having little impact when not performing well and receiving great support to improve. Negative experiences were mentioned of parents with high expectations, a 'toxic' group dynamic and unfair teachers. The initial (emotional) response to the leaderboard was created by prior experiences with similar comparisons and scores. This initial reaction was mentioned as the reason to start to perform, or not to perform. To talk in friendly competition about the scores, or not want to talk about the scores at all.

> "I got bullied a lot when I was younger. I was really glad that I never finished last on the leaderboard because I feel like that would have affected me more. Because I do feel judged by other people very, very quickly, even though they don't. I feel like if this leaderboard would exist in that time, then I would have been bullied more in the math area, specifically, probably." - *Participant 2*

Group dynamic: The effect of the leaderboard is believed to be greatly influenced by a good or bad group dynamic. A positive group dynamic would be support towards each other, help each

other with good feedback, without being too competitive and having an overall good 'vibe'. A negative group dynamic was believed to be harsh towards each others (bad) performance, non-supportive and filled with judgement, also described as being "toxic".

> "If you can motivate each other positively, like help each other out to get both get a higher rank, I think then competition could be a positive addition, because you really want to help your friend out of like, maybe a bad grade. Because she knows you and maybe if you know like that your skills are the same, but one got a better score than the other, it would motivate me to work harder, basically." - *Participant 7*

Perceived fairness: The effect of the leaderboard is greatly effected by how the score and ranking are calculated and rewarded. If the leaderboard position was rewarded for unjust reasons, such as favoritism, "sneaky" or unjust correct answers and unfair metrics, participants would no longer try to perform or compete. Whether or not the leaderboard position was awarded fairly was important mostly for middle to low performers, as top performers did not mention it as something of importance when receiving a good score.

> "In another class, a lot of favoritism came in. So like, if a teacher likes you personally you would get high grades. So I was like, well, this is stupid. I am not going to try to fight for this because there is no chance for it. You know, it doesn't mean that I don't know the material, which I definitely know." - *Participant 1*

5.0.10 Requirements for educational context

Special attention was spent discussing how the leaderboard could be implemented in educational context that would emphasize the positive effects and minimize the negative effects. The influential factors and possible designs were discussed. This lead to five requirements that were believed to create a positive motivational effect on people, while reducing the negative consequences as much as possible. These five requirements are shown in figure 5.4 and will be explained later in this section.

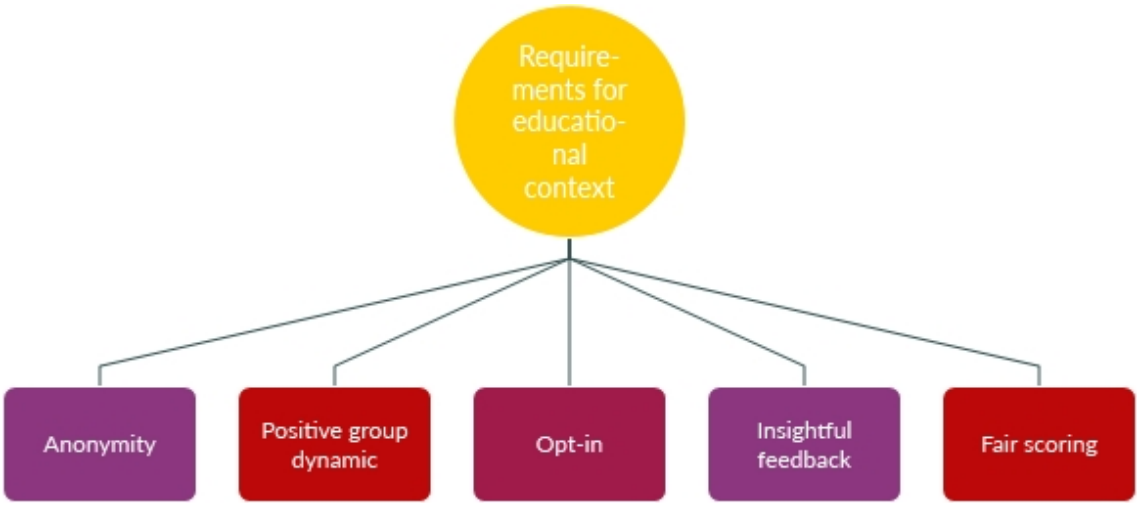


Figure 5.4: Mind-map of the requirements for education

Anonymity: Anonymity protects bottom-end performers from bullying or judgement from higher performers and was most often mentioned as a necessity. It allows them to decide by

themselves if they want to share their scores with others or not, lowering the barrier to participate. However, complete anonymity could reduce the motivational effects for high performers and those with a competitive mindset. Letting people decide names for themselves, allowing for random ("funny") names that can be changed if wanted, was believed to be anonymous and visible enough for both the top and bottom performers. This kind of anonymity does not necessarily deplete the competitive motivation from top performers, as top performers are believed to make themselves visible regardless and are generally uninterested in the names of worse performers.

> "I would be like I'm the panda! Yeah. But only if the score was good, otherwise, it just wouldn't say my fake name." - *Participant 7*

Positive group dynamic: See group dynamic in section 5.0.9

Opt-in: The possibility to opt-in allows those confident in their ability, or those who enjoy competition, to compete against others that want to compete as well. Opt-in protects the individuals from a negative response due to prior negative experiences with competition or leaderboards (see 5.0.9). Simultaneously, it allows participants to judge other influential or important aspects to be sufficient, such as group dynamic and anonymity.

Fair scoring: See perceived fairness in 5.0.9

Insightful feedback: Insightful feedback is feedback that allows the participant to improve upon their score. Improvement is especially necessary for bottom-end performers to keep competing, as continues bad performance will terminate people to keep competing and deplete their perceived ability on a subject or skill.

> "If you want to get better grades, but it is not working and I am trying my best. That's wrong. I have to try something different, you know, so I can compete with them. If you score like, badly, five times in a row, that you would at some point, say, I think I will stop with this competition, because it is not for me" - *Participant 1*

The next chapter will present a discussion of the findings from the two studies.

Chapter 6

Discussion

The aim of this research was to explore the inner workings of leaderboard and find reasons for its effect on the learning process. This contribution to the knowledge gap was identified during the literature review, where the need for a deeper understanding of specific Gamification elements was mentioned in several meta-analyses to be necessary. In this chapter, the will interpret the results, discuss limitations, list the research contributions and provide recommendations for future work. The research is summarized in chapter 6.

6.0.1 How the leaderboard is used

Our findings suggest that the leaderboard is used out of curiosity and as a tool to self evaluate. This is different from studies such as (Jia et al., 2017) and (Zainuddin, 2018), that categorize the leaderboard as an effective way to motivate users through competition. However, while top performers did care to remain a top performer, they were not interested or looked at those who performed worse, even if these people were known to them. This shows that the ranking is used in a similar way as the score itself, answering how well an individual performed. While not everyone uses the performance of others to judge their own performance, most (67%) did use the leaderboard for this purpose. The most interesting aspects of the board are the participants own score and ranking, being able to see who performed better then the participant and how others scored that the participant knew or is close to. It is likely that leaderboards showing only ranking, without showing how others performed such as those in (Ortiz-Rojas et al., 2019), would remove some of the main interesting aspects of the board and part of its effect. Our findings suggest that the leaderboard becomes more interesting if the scores are related to a higher final grade or reward. During the focus-group session, top performers explained that they felt the urge to know if they were the best overall, or just from their group. This finding adds to the research of (Jia et al., 2017), who attributed the personality type *extraversion* to this behavior instead of performance.

6.0.2 The effect of the leaderboard

The effect of the leaderboard had been studied in prior research, often focusing on specific effects such as learning performance, (intrinsic) motivation and engagement as mentioned in the meta-analysis of (Zainuddin et al., 2020). Others have put their focus on specific contexts and personality traits, such as (Jia et al., 2017). However, research in this field had often been conducted based upon behavioral data, without a deeper understanding of why these behaviors occur. This research identified seven influential factors that could influence a positive or negative response to the leaderboard, as perceived by the participants themselves. The most notable additions to earlier

research is what aspects influence these effect outside of the software and design itself. These are, closeness to the people on the leaderboard, consequences caused by the rewarded score, prior time and energy investment, score and ranking performance, a (positive) group dynamic, perceived fairness of the score and prior (good or bad) experiences with comparisons and competition. This research showed that motivation and learning performance can increase, while simultaneously creating a less desirable learning environment that can, for some, be demotivating. Future research is therefore advised to keep the mentioned aspects of influence in mind, before generalizing results towards a broader audience when conducting research on leaderboards.

6.0.3 Can Social comparison theory be used to understand leaderboards?

This research suggests that social comparison theory can be used to understand part of what creates the effect of the leaderboard, as ability estimate and proximity both greatly influence the impact of the leaderboard on the individual, similar to the findings of (Gerber et al., 2018). However, social comparison theory does not give specific recommendations on how to take these effects into account for a practical contexts to improve motivation or learning performance. The research at hand adds to this existing knowledge in the form of recommendations within a specific educational context, studied within the digital form of the leaderboard, specifically aimed to increase its beneficial aspects.

6.0.4 Requirements for educational context

This research found five requirements when implementing a leaderboard into an educational context, namely; anonymity, positive group dynamic, the option to opt-in, insightful feedback and fair scoring. The research suggests that these requirements minimize the possible negative effects the leaderboard can have, without removing the positive benefits of its implementation. To maximize performance, all five requirements for educational context are recommended to carefully be taken into account, requiring the necessary time and energy investment from those involved in the learning process, such as the teacher (for the fairness of the exam and how its scored, as well as insightful feedback and managing group dynamics), the group (in terms of group dynamic), the student (in terms of energy investment), combined with a system that minimized the chances of demotivating behavior (with the option to opt-in and anonymity) and maximizing the motivating aspects (allowing top performers to keep being challenged and to improve and allowing some social comparison), the implementation of the leaderboard in education can be a welcome addition that can improve learning performance and overall motivation. Additionally, the implementer is advised to consider if the 'unnatural' increased importance to perform and improve created by game-elements like the leaderboard is justified by the benefits it will have for the live of the individual, as the implementer is believed to have some responsibility as to not drive them towards goals against their better judgement.

6.0.5 Recommendations

To implement the leaderboard in educational context, the implementer needs to carefully consider the consequences of the implementation, which this research has shown to be more complex then motivation and engagement alone.

Next to the factors influencing successful implementation, the leaderboard element is often mentioned to lure participants into comparing themselves with others within their group, which

has its own implications. This effect, while perhaps beneficial for motivation and overall performance, could still deviate from getting the most out of each individual, or even to get them to a certain level of education. This research shows that a low grade, or even a insufficient grade to pass, can for some lose its value when those within the social circle of the individual performed similarly badly. Group performance can in some situations completely detach any emotional response to personal performance, as mentioned during the focus group. In theory, this would mean that groups that consist of high performers will judge themselves based upon other high performers, which could create a very high average standard to uphold. In contrast, groups filled with low performers could create a much lower standard, and thus, less energy and time investment required to spend on learning to remain 'part of the group'. For example, as some participants in the focus group mentioned feeling motivated to not perform at the bottom of the group, this goal will be much more competitive in a high performing group. As the leaderboard made top performers less interested in feedback and improvement, this could mean that top performers of low-average groups will not improve towards their maximum capability due to their self reflection being dependent upon the groups performance. As such, when implementing the leaderboard, the top-performers require a greater challenge to keep improving upon their already good performance. This also came up during the experiment, as top performers searched out for other top performers on other leaderboards to keep experiencing friendly competition, while spending no time discussing their performance with others on their board. Bottom-end performers on the other hand require insightful feedback to improve, and should constantly notice how their efforts are translated into an improvement in score.

6.1 Limitations

Participants were recruited from the master course "Serious Gaming" at Utrecht University and cannot be generalized to a general audience without further research. As noted by some of the participants, it is possible that the results would differ when the study would be performed on participants who are not interested in (serious) games. Most participants were of similar age, and other (learning)contexts and factors need to first be evaluated on their effect. Notable limitations of the study are the relatively short duration and the limited number of participants.

Context and quiz In the current study, participants had to perform in a knowledge quiz in order to end up high on the leaderboard. The knowledge quiz was beneficial for the participant, as the knowledge was related to the knowledge of the course. This could have influenced how answers were given in relation to why people looked at the leaderboard, as they could have solely wanted to use the leaderboard to understand their own knowledge of the course and only used the leaderboard to check their own score. It was not possible the see the score without looking at the leaderboard. This could partially explain the high amount of participants that mentioned not being interested in the scores of others and only using the leaderboard to see their own score.

Focus group The focus group was limited in that, due to convenience sampling, all participants were female and came from group A (within proximity). Specific attention was given to change this outcome, without success. It is possible that men would have given different answers during the focus group.

Experiment setup The first experiment limitation accrued during quiz 2. While only one answer was correct in quizzes 1 and 3, the second quiz had questions in which two answers were scored as being correct. This was done after the quiz had been made in agreement with the supervisor, as the answers were too similar. This caused the scores of quiz 2 to be higher, on average, then the other quizzes. There is no visible indication in the data that suggests this had an effect on the results.

The second limitation was that the experiment was done inside a lecture-room at Utrecht University under supervision of the researcher and one of the supervisors. However, some whispers and soft discussions could be heard, and it is believed that some may have cheated. Others spend an amount of time on the quizzes that far exceeded the average (sometimes by a multitude of 4), which may be an indication of looking up answers. No actions have been taken against this behaviour, and no effect of this behaviour was found in the data.

The third limitation is that the quizzes were presented at the beginning of a lecture to improve participation, which indeed caused nearly all present students to participate. However, not every students is present during each lecture, and the amount of students present during the final quiz was notably lower then during the first two quizzes. This seemed to have had a direct negative effect on the amount of participants that filled in the third and final quiz. The reasons given to not be present were lack of time due to upcoming deadlines and the exam. One more reminder then usual was sent out to improve the situation, creating an additional three entries for the final quiz, while still remaining lower then others. This explains the low participation rate of the third quiz.

The final limitation during the experiment was that the first quiz was made on a Friday, with the first leaderboard sent out on the following Monday. This created a time-frame of 3 days for participants to make the quiz. Due to scheduling re-arrangements, the other quizzes were made on Wednesday's and had a shorter time-frame of 24-hours to answer. Some of noted this time-frame as being too short, and some have made the quizzes after the deadline. These were not included in the data-set, but may have slightly effected the participation rate.

Participants One participant reported that he 'figured out' that the leaderboard scores were fake. This was because he received points on the leaderboard, without participating in the quiz itself. After some research, it appears that another individual may have incidentally answered (some) of the quiz questions on the wrong name, before exiting the quiz early (perhaps to try the quiz again using the right name). As a few questions were answered on the incorrect name, this score was presented on the leaderboard. After asking in the focus group, it seems that this participant did tell others that the scores are fake, but that others disregarded this comment. There is no indication in the data that suggests this had an effect, but it could have influenced factors such as the willingness to invest time to improve the score next week or general motivation for the quiz, leaderboard and research.

Chapter 7

Conclusion

7.1 Main findings & contributions

The results of this study provides a deeper understanding of what aspects cause, influence and effect the Gamification element Leaderboard and provides requirements and recommendations when implementation the leaderboard into an education context. It shows that the social comparison theory can be used to understand parts of the effect of the leaderboard and that the implementation of leaderboards can lead to an increase in while simultaneously creating a less enjoyable and learning environment for some. From the results and discussion of this research, the following main conclusions can be drawn:

- Seven aspects were found that influence the effect of the leaderboard on participants. These are; closeness to others on the leaderboard, the consequences related to the score and ranking, prior time and energy investment, the performance of the participant, prior experiences with competition and social comparisons, a positive group dynamic and the perceived fairness of the received score and ranking. For more information, refer back to chapter 5.

- Five requirements were found for implementing the leaderboard into an educational context. These are; anonymity, (positive) group dynamic, insightful feedback, fair scoring and the possibility to opt-in. The implementation of the requirements can vary and should be determined by a person involved in the course design, such as a lecturer. For example, a positive group-dynamic and insightful feedback might be encouraged with peer-reviews and social activities. For more information, see chapter 4

- The leaderboard is most often used out of curiosity and as a tool to self evaluate. This adds to the field of HCI research, that often describes the effect of the leaderboard as mainly a competitive tool. The performance of others was used by 67% of participants to self evaluate, showing that a substantial part of the group is not interested in the performance of others. More research is required to discover what causes the participant's performance relative to others to be used for this purpose. This research suggests that the leaderboard should be designed around the goal of self evaluation. For more information, see the recommendations in 6.

- The most interesting information on the board from the viewpoint of the participant, is who performed better and how others socially close to the participant scored. This research suggests that this information should be shown to the user if implemented. Anonymous

leaderboards and the option to opt-in could be used to avoid negative comparison effects, including those related to people close to each other.

- Time spend studying could be increased if the score on the leaderboard would influence the live of the participant, such as when the score is related to the final course grade, or when a reward is added. The performance could be increased if the score would be accompanied by personalized, insightful feedback, that would allow the participant to improve. Time spend studying could decrease when the participant does not improve after multiple weeks of trying, showing the need for feedback, and to experience improvement. Low and middle performers were most interested in insightful feedback. High performers were very not interested in feedback and improvement, showing that high performers could require a greater challenge to continue to improve.

- Social comparison theory can be used to understand part of the effects of the leaderboard. The ability estimation and emotional response are greatly effected by the performance of the group. This effect is stronger when the comparison is made with people who are close to the participant. This stronger effect can be recommended as it can be both beneficial and fun, under the condition of a positive group dynamic and the ability to opt-in.

- Motivation can show an increase while simultaneously creating a less enjoyable learning environment. This increased motivation could be beneficial when quick training is required. For example, in training scenario's for fireman or policemen. In the educational context, a more enjoyable learning environment for the entire group could be recommended. To take benefit from the motivational increase for part of the group, a positive group dynamic and the ability to opt-in are considered especially important.

- We have contributed to the field of HCI with the literature review, proving a foundation of knowledge.

- This research contributes to the field of HCI in the form of empirically gathered data. The data consists of the scores, participation, and answers in the questionnaire.

- This research contributed to the methodological design in which a cohort of students is engaged in one common activity, split into two groups of within and outside of proximity, and tested over multiple weeks on their understanding of the activity. This method has shown to find differences identify factors that influence how the game element Leaderboard effects the user.

7.2 Future work

There are many directions in which future research could take place. How the leaderboard is best used as a tool to self evaluate is not currently studied, and could be studied in future work. Each of the mentioned aspects that influence the effect of the leaderboard could be studied in detail, by re-running the experiment and focusing on each of these aspects individually. Next to the influences themselves, understanding in which contexts these aspects have a greater or lesser effect, as well

as understanding the role of a personality of these aspects could prove to be equally valuable to study. Other directions of possible future studies are explained and elaborated upon in this section.

Study into other contexts and actions: This study was performed on Utrecht University in The Netherlands. Other educational contexts, such as primary and high school, could give new insights into the effects of the leaderboard. Participants mentioned that the leaderboard was not often discussed as the master students only had 'professional' contact, mostly online, with other students in their group. This is not the case in high schools, as children are often required to spend their days in close physical proximity to one another every workday for multiple years. Future research could implement the leaderboard in different contexts to find the differences in the measures effects and improve upon the recommendations and requirements or even provide guidelines.

Study into the possibility of multiple aspect scoring:
This study showed that a received grade and the group performance can have a great influence on the individual. An interesting idea mentioned during the focus group was that the issue might be that only one type of score is shown. Games, such as Call of Duty or Mirror's Edge, show a leaderboard that has multiple points on which the player is scored. When a player scored low on one part of that score screen, it was mentioned to be less influential when other parts showed higher numbers. In games, people might be scored for teamwork, being quick, accurate, etc. It might be beneficial if students could be scored on other aspects as well, similar to what happens in games. Which kind of information the final result could be mixed with to reduce the impact, could influence the impact of a (bad) score on the user. What information could be shown and the implications of that information are left to be discovered.

Study in the mental blockade of prior experiences:
During the focus group sessions, the difference in how people responded to the leaderboard was effected greatly by prior experiences in their lives in relation to leaderboards, competition and social comparisons. Others who liked to perform, had great experiences with friendly competition that was fun and enjoyable, often with a good group dynamic. This shows that a negative response to the leaderboard and even the lack of willingness to compete is not directly related to the leaderboard or the competition itself, but is partially or entirely created by a mental blockade. In theory, this mental blockade can, at least for some, be removed or reduced. Future research could focus on finding recommendations for teachers to help students with this mental blockade, by for instance talking with them about their negative experiences and seeing if they are willing to give it another chance. New positive experiences could just as well aid this goal. A digital solution that would motivate students to give competition another chance might be a possibility that could further increase the effectiveness of the implementation.

Study into the implementation of anonymity:
This study shows how anonymity is essential to protect participants from negative consequences such as bullying and increased experienced pressure. However, there are still gaps into how this anonymity should be implemented. If the leaderboard is 'too' anonymous, it will decrease the beneficial, motivational effects for those who want to perform. When the leaderboard is not anonymous enough, the same negative aspects of the board mentioned earlier could return.

Final statement:
As numerous aspects influence the effect of the leaderboard, it is recommended to implement this element carefully considering the influential aspects mentioned in this research. Studying the leaderboard on itself is not enough, and careful attention needs to be paid to the other aspects that influence its effect to implement it successfully. More research is necessary to add to these findings and to further advance the discipline.

Chapter 8

Appendix A

8.1 Research Consensus

Introduction: You may have experienced how difficult it can be to stay motivated when studying. For some, the educational system can be more of a motivational challenge rather than an intellectual one. The digital education industry has widely implemented game-elements (such as points and leaderboards) in their products to increase the engagement and motivation of students. This process is called "Gamification". However, game-elements are currently implemented without a thorough scientific understanding of its inner workings. The focus of this research is to create a deeper understanding on what influences how people experience the game-element "Leaderboard" to make the implementation of this element more successful.

The aim of this study is to further understand how the "Leaderboard" influences motivation and learning performance.

If you participate in the study: Each week, for a total of three weeks, you will be asked to 'play' (fill out) a digital quiz. The quizzes are scored and compared with other students on a Leaderboard. The Leaderboards reset each week, creating a new possibility to end on top each week. The quizzes will take about five minutes to complete and can be completed on every modern computer, tablet or smartphone. You will also be required to fill in a pre- and post-test questionnaire.

What you will gain from it: The questions in the quiz are derived from the course material and are inspected and agreed upon by the course professors. By participating in the study, you are more likely to stay on track with the course material and gain feedback into your current knowledge level. The study itself could be interesting for those who will, at some point, make their own master thesis as it provides both an example as well as an experience from the perspective of the participant. We also hope that you are interested, in general, to contribute to scientific discovery and progression for yourself, your environment and the generations to come. Finally, if you completed all parts of the study, you will have a chance to win a delicious reward at the end of the study.

Participation for this study is voluntary. You can withdraw from the research at any given time (until the point of data analysis) without providing a reason and without consequence. Your data will be stored securely, deleted upon request and will not be sold or used for reasons other than this research. The scores do not influence your grading or any other aspect of the course. Your data or scores will not be used, viewed or otherwise analyzed by the professors from the

course.

If you have any questions, concerns or if you just want to talk about the research feel free to send a message to s.a.andredelaporte@students.uu.nl

o I consent to participate in this study under the listed conditions

Please fill in your full name:
Please fill in your student e-mail address:

8.2 Pre-test survey

Utrecht University

Information about the study

Please select your full name

Demographics

What is your age?

What is your gender?

- Female
- Male
- Other
- Prefer not to say

Previous knowledge

How would you rate your knowledge on Serious Games or other related topics that may aid you in this course?

	No / little previous knowledge		Some previous knowledge / similar courses completed previously		Expert knowledge / very similar courses completed previously
	1	2	3	4	5
Estimated prior knowledge level					

How many hours per week are you planning to spend on the course Serious Games?

- ◯ 0 - 4 hours
- ◯ 5 - 8 hours
- ◯ 9 - 12 hours
- ◯ 13 - 16 hours
- ◯ 17 - 20 hours
- ◯ 21 - 24 hours
- ◯ 25 - 28 hours
- ◯ More then 29 hours

What grade do you (realistically) expect to receive at the end of this course?

- ◯ Below 5.5
- ◯ 5.5 - 6.0
- ◯ 6.1 - 6.5
- ◯ 6.6 - 7.0
- ◯ 7.1 - 7.5
- ◯ 7.6 - 8.0
- ◯ 8.1 - 9.0
- ◯ 9.1 - 10.0

Relation

How well do you know the others listed below? Use the slider.

	We've just met	We've met before this course	We've worked together before	We see each other socially	We see each other socially frequently
	1	2	3	4	5
${e://Field/t1}	◯				
${e://Field/t2}	◯				
${e://Field/t3}	◯				
${e://Field/t4}	◯				

Do you currently work together with one of the people listed above in another course or activity? Please fill in the name(s) of the student(s) and the activity / course.

8.3 Quiz 1

Introduction: Welcome to the Serious Gaming Quiz! You will receive +5 points for each right answer and -2 points for each wrong answer. The option "I'm not sure" gives +0 points. More points are rewarded to those who are quick. Select your name from the list below to get started.

Question 1: "Flow Theory" is often seen as a key feature of well-designed games. According to Fullerton's theory, appropriate levels of (...) and (...), can cause a player to achieve a "Flow" state.

1. Motivation & Ability
2. Difficulty & Skill
3. Anxiety & Boredom
4. Skill & Motivation
5. I'm not sure

Question 2: Picture 1 shows an influential way to promote sales in online webshops. Which of the following principles of influence by Dr. Robert Cialdini is most present in the image?

1. Liking
2. Authority
3. Social proof
4. Unity
5. I'm not sure

Question 3: What is the difference between 'Play' and 'Game'?

1. Play requires explicit rules, a Game does not
2. Play is extrinsically motivated, whereas gaming is intrinsically motivated
3. Play does not require a purpose, a Game does
4. A Game requires explicit rules, Play does not
5. I'm not sure

Question 4: Which sentence would create the strongest feeling of product scarcity, according to Dr. Robert Cialdini principles?

1. "Buy one, get one free!"
2. "Doctors and medical experts recommend this item"
3. "Only a few units left at this price."

4. "Two of your friends want this item"

5. I'm not sure

Question 5: Re-create the map of the decision making process by putting them in the right order. Put the process that costs the least amount of thinking on the left as 1. Put the process that requires the most thinking on the right as 5.

1. Conscious, cost-benefit calculations

2. Habits

3. Heuristics

4. Intuitive responses and assessments

5. Intuitive check of self-concept

8.4 Quiz 2

Introduction: Welcome to the second Serious Gaming Quiz! You will receive +5 points for each right answer and -2 points for each wrong answer. The option "I'm not sure" gives +0 points. More points are rewarded to those who are quick. Select your name from the list below to get started.

Question 1: What best descripes the term 'participation design'?

1. A design approach that involves the stakeholders

2. A design approach variant utilizing interviews and focus groups

3. A design approach that finds solutions in co-operation with the end-user

4. An interactive prescriptive design approach

5. I don't know

Question 2: The characterizing goals of (serious) games can be matched to competence domains. One example can be found in the centuries-old board game Chess, in which the winner of a match was believed to be more intelligent (or a better strategist) then the other. Which domain of competence was Chess believed to be part of?

1. Cognitive & perceptual competences

2. Sensory-motor competences

3. Personal competences

4. Emotional & voliational competences

5. I'm not sure

Question 3: Bloom's taxonomy of cognitive objectives describes learning in six levels of complexity. At the first (lowest) level, people only remember information. What is considered the highest level of thinking?

1. Analyze: Draw connections among ideas
2. Evaluate: Motivating or justifying a stand or decision
3. Understand: Explaining ideas or concepts
4. Create: Producing new or original work
5. I'm not sure

Question 4: Interesting emergent behaviour makes a game more enjoyable. What could be seen as a good measure of how emergent the game-play is?

1. The interplay between novel and known actions
2. The ratio between strategic and basic actions
3. The amount of meaningful achievable goals
4. The variance in game-play objectives
5. I'm not sure

Question 5: Which description gives the definition of the term game-mechanics?

1. Game Mechanics are the objects, attributes and states of a game
2. Game Mechanics are the story and visuals that create the game experience
3. Game Mechanics are the capabilities and limits of the underlying technology of a game
4. Game Mechanics are the interactions and connections between the player and the game
5. I'm not sure

8.5 Quiz 3

Introduction: Welcome to the final Serious Gaming Quiz! You will receive +5 points for each right answer and -2 points for each wrong answer. The option "I'm not sure" gives +0 points. More points are rewarded to those who are quick. Select your name from the list below to get started.

Question 1: Which of the following is NOT a Game Mechanic?

1. Actions.
2. Emotions.
3. Skills.

4. Chance.

5. I'm not sure.

Question 2: Picture 1 shows the ATMSG model, but the term Serious Game is missing. Where should it be placed?

1. In the model, a serious game is an (instructional) activity.

2. In the model, a serious game is a tool.

3. In the model, a serious game is the object or motive.

4. The term serious game should not be in this model.

5. I don't know

Question 3: The Self-Determination Theory (SDT) was coined by Ryan & Deci in 2000. They describe three basic psychological needs that encourage human motivation. What are these three needs?

1. Interest, compliance, reward.

2. Autonomy, competence, relatedness.

3. Competence, compliance, inherent satisfaction.

4. Que, routine, reward.

5. I'm not sure.

Question 4: Which of the following specifically increases extrinsic motivation?

1. By making actions easier to perform.

2. Avoiding boredom and stress, aiming for a perfect flow state.

3. Increasing the experience of safety by compelling or controlling one's behaviour.

4. Showing the benefits and progress of the action to the user.

5. I'm not sure.

Question 5: What is not true about the attributes of game objects?

1. Attributes are always dynamic

2. All objects have at least one attribute

3. Attributes are categories of information about the object in the game

4. Attributes are the "adjectives" of game mechanics

5. I'm not sure

Focus group session: Would you like to participate in the focus group session? The focus group will take 30 to 60 minutes. We will contact you to see what works best for you in terms of time and location. Note that no points are awarded or deducted for this question.

1. Yes, I would like to participate.

2. No, sorry.

Explanation: Attributes can be static and dynamic.

8.6 Post-test survey

Identification

Welcome to the post-test survey! The survey is devided into 5 parts and should take between 10 and 15 minutes to complete. Due to the time investment, there is a treat waiting for you after you're done. We recommend you to make this questionnaire on your laptop, as there are a few open-ended questions. Note that this is the final part of the research. Thanks a lot for your help! **Select your name below to get started.**

Each part and it's expected duration:

1. Motivation (3 minutes)
2. Comparison with others (5 minutes)
3. Leaderboard aspects (2 minutes)
4. Your experience (3 minutes)
5. Technical difficulties and feedback (2 minutes)

Select your full name

Motivation questionnaire

Part one: motivation. 19 multiple-choice questions. Expected duration: 3-5 minutes.

Q1: What grade do you (realistically) expect to receive at the end of the course?

- Below 5,5
- 5.5 - 6.0
- 6.1 - 6.5
- 6.6 - 7.0
- 7.1 - 7.5
- 7.6 - 8.0
- 8.1 - 9.0
- 9.1 - 10.0

Questions about motivation. **10 questions.**

	Never	Rarely	Sometimes	Usually	Always
Getting a good grade is important to me.	○	○	○	○	○
I like to do better than others on tests.	○	○	○	○	○
I put enough effort into learning the course material.	○	○	○	○	○
I enjoy learning.	○	○	○	○	○
Learning for this course will help me get a good job.	○	○	○	○	○
I am confident I will do well on the upcoming exam.	○	○	○	○	○
I spend a lot of time learning for this course.	○	○	○	○	○
Understanding the course material will give me a career advantage.	○	○	○	○	○
I think about the grade I will get for this course.	○	○	○	○	○
I believe I can master the required knowledge and skills before the end of the course.	○	○	○	○	○

Questions about The Quiz. **7 questions.**

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I prepared for all quizzes.	○	○	○	○	○
I did my best on all quizzes.	○	○	○	○	○
I was satisfied with all my quiz results.	○	○	○	○	○
I liked making the quizzes.	○	○	○	○	○
The questions were clear to me.	○	○	○	○	○
The communcation about the quizzes was clear.	○	○	○	○	○
The quizzes helpt me to focus on the course.	○	○	○	○	○

Q19: If you didn't make one or more of the quizzes, what was the reason?

The effect of Social Comparison & Proximity

Part two: Comparison with others. 28 multiple-choice questions. Expected duration: 5-7 minutes. (Note: this is the biggest part.)

Comparison with others (part one). **14 questions.**

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I always looked at the leaderboard when it was send.	○	○	○	○	○
The information on the leaderboard was clear to me.	○	○	○	○	○
I always look at the performance of others.	○	○	○	○	○
I always compare my performance to others.	○	○	○	○	○
I always use the leaderboard to evaluate my own performance.	○	○	○	○	○
I like being able to see the performance of others.	○	○	○	○	○
I like that others could see my performance.	○	○	○	○	○
I always talk with others who where on my leaderboard about their score or my own.	○	○	○	○	○
I always talk with others who where **not** on my leaderboard about my score (or their score).	○	○	○	○	○
Comparing my performance with others motivates me to perform better.	○	○	○	○	○
A high or low score could influence how others look at me.	○	○	○	○	○
I'm motivation to make the quiz because others can see my score.	○	○	○	○	○
Without the leaderboard, i'm not sure if I would have made the quizzes.	○	○	○	○	○
I would have liked the quizzes more without the leaderboard.	○	○	○	○	○

Comparison with others (part two). **14 questions.**

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
15. The leaderboard made me anxious.	○	○	○	○	○
16. People with a low score are lazy.	○	○	○	○	○
17. The leaderboard made me question my own capabilities.	○	○	○	○	○
18. The scores of others on the leaderboard influence how I think about them.	○	○	○	○	○
19. I feel ashamed when others see a low score with my name next to it.	○	○	○	○	○
20. I feel proud when others see a high score with my name next to it.	○	○	○	○	○
21. People who perform well in the quiz are hardworking.	○	○	○	○	○
22. I was afraid to be judged by others.	○	○	○	○	○
23. The leaderboard made me feel as if I had to perform.	○	○	○	○	○
24. I've always talked with people in my project group about the leaderboard (score).	○	○	○	○	○
25. The leaderboard made me excited	○	○	○	○	○
26. I like to know the score of others in my project group.	○	○	○	○	○
27. I am more curious about score of people in my project group then outside my project group.	○	○	○	○	○

28. The leaderboard made me feel confident about my abilities.	○	○	○	○	○

What is the leaderboard to you?

Part 3: Leaderboard aspects. 3 multiple-choice questions. Expected duration: 2 minutes.

Q1: What was your main reason to look at the leaderboard?

- ○ I wanted to know my position relative to others on the leaderboard.
- ○ I wanted to know my overall score on the leaderboard.
- ○ I wanted to evaluate my own performance.
- ○ I didn't really look at the leaderboard.

Q2: What interested you the most?

- ○ People who performed better then me.
- ○ People who performed worse then me.
- ○ People who I knew.
- ○ The performance of others did not interest me.

Q3: How would you cathogorize these aspects of the leaderboard to increase time spend on studying?

Items	Increased time spend on studying	No effect on time spend studying	Decreased time spend on studying
Seeing others that I know.			
Being able to see people that performed better then me.			
Being able to see people that performed worse then me.			
My ranking relative to my previous ranking/score.			
The visuals of the leaderboard, such as theme, icons and fonts.			
How many people would look at the leaderboard.			
If people close to me would see the leaderboard.			
Being able to select the people who I am on the leaderboard with.			
When the score is related to a reward.			

Would there be another aspect of the leaderboard that would influence your time spend on studying that is not mentioned above? If so, what aspect would effect you?

The effects of the leaderboard

Part 4: Experience. This part will ask your opinion in 7 open-ended questions. No expected duration is given.

Q1/7: If you discussed your score with others, with whom did you discuss it and why? You don't have to give actually names, only your relation and reason (older brother, friend from my project group, etc).

Q2/7: Would you have preferred if your leaderboard showed scores of others in the class who were now excluded?

Q3/7: Would you have liked to select the people that are on your leaderboard? If so, who would you have wanted on the leaderboard?

Q4/7: How would you describe the effect the leaderboard had on you?

Q5/7: Would you like to have a leaderboard in other courses? Please explain why (not).

Q6/7: What negative effects could the leaderboard have in general,

in your opinion?

Q7/7: Can you think of a scenario in education where you would not want a leaderboard?

Technical difficulties and feedback

Part 5: The final part about technical difficulties and feedback. 2 open-ended and 2 multiple-choice questions. Expected duration: 2 minutes.

Q1/4: Could something have been improved that would have made the leaderboard more interesting for you?

Q2/4: Do you have any feedback about the research or other tips

that should be mentioned?

Q3/4: Do you want to be part of the focus group?

- ◯ Yes
- ◯ No

Q4/4: Would you like to receive the final research paper?

- ◯ Yes
- ◯ No

8.7 Reminder e-mail

Dear Serious Gaming participant,

We need your participation in the final Serious Gaming quiz!
It will only take 2 minutes to complete. The deadline is on 15:00 today.

Click here to start the quiz now!

Thanks a lot!

Kind regards,

Sjoerd André de la Porte

Figure 8.1: Example of a reminder email

Bibliography

Abbasi, A. Z., Ting, D. H., and Hlavacs, H. (2017). Engagement in games: developing an instrument to measure consumer videogame engagement and its validation. *International Journal of Computer Games Technology*, 2017.

Barata, G., Gama, S., Jorge, J., and Gonçalves, D. (2017). Studying student differentiation in gamified education: A long-term study. *Computers in Human Behavior*, 71:550–585.

Broer, J. (2014). Gamification and the trough of disillusionment. *Mensch & Computer 2014-Workshopband*.

Buil, I., Catalán, S., and Martínez, E. (2020). Understanding applicants' reactions to gamified recruitment. *Journal of Business Research*, 110:41–50.

Butler, C. (2013). The effect of leaderboard ranking on players' perception of gaming fun. In *International Conference on Online Communities and Social Computing*, pages 129–136. Springer.

Cagiltay, N. E., Ozcelik, E., and Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education*, 87:35–41.

Caponetto, I., Earp, J., and Ott, M. (2014). Gamification and education: A literature review. In *European Conference on Games Based Learning*, volume 1, page 50. Academic Conferences International Limited.

Christy, K. R. and Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, 78:66–77.

Csikszentmihalyi, M., Abuhamdeh, S., and Nakamura, J. (2005). Flow. *Handbook of competence and motivation*, pages 598–608.

Csikszentmihalyi, M. and Csikzentmihaly, M. (1990). *Flow: The psychology of optimal experience*, volume 1990. Harper & Row New York.

Deci, E. L. and Ryan, R. M. (2012). Self-determination theory.

Deterding, S. (2014a). The ambiguity of games: Histories and discourses of a gameful world.

Deterding, S. (2014b). Eudaimonic design, or: Six invitations to rethink gamification.

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15.

Dichev, C. and Dicheva, D. (2017). Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International journal of educational technology in higher education*, 14(1):1–36.

DiSalvo, C., Sengers, P., and Brynjarsdóttir, H. (2010). Mapping the landscape of sustainable hci. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages

1975–1984.

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, 7(2):117–140.

Fogg, B. J. (1998). Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 225–232.

Gerber, J., Wheeler, L., and Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological bulletin*, 144(2):177.

Glynn, S. M., Brickman, P., Armstrong, N., and Taasoobshirazi, G. (2011). Science motivation questionnaire ii: Validation with science majors and nonscience majors. *Journal of research in science teaching*, 48(10):1159–1176.

Hamari, J., Koivisto, J., and Pakkanen, T. (2014a). Do persuasive technologies persuade?-a review of empirical studies. In *International conference on persuasive technology*, pages 118–136. Springer.

Hamari, J., Koivisto, J., and Sarsa, H. (2014b). Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences*, pages 3025–3034. Ieee.

Jia, Y., Liu, Y., Yu, X., and Voida, S. (2017). Designing leaderboards for gamification: Perceived differences based on user ranking, application domain, and personality traits. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 1949–1960.

Khalil, M., Wong, J., de Koning, B., Ebner, M., and Paas, F. (2018). Gamification in moocs: A review of the state of the art. In *2018 IEEE global engineering education conference (educon)*, pages 1629–1638. IEEE.

Kors, M., Van der Spek, E. D., and Schouten, B. A. (2015). A foundation for the persuasive gameplay experience. In *FDG*.

Linehan, C., Kirman, B., Lawson, S., and Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1979–1988.

Natucci, G. C. and Borges, M. A. (2021). Balancing pedagogy, emotions and game design in serious game development.

Ortiz-Rojas, M., Chiluiza, K., and Valcke, M. (2019). Gamification through leaderboards: An empirical study in engineering education. *Computer Applications in Engineering Education*, 27(4):777–788.

Plass, J. L., Homer, B. D., and Kinzer, C. K. (2015). Foundations of game-based learning. *Educational psychologist*, 50(4):258–283.

Przybylski, A. K., Rigby, C. S., and Ryan, R. M. (2010). A motivational model of video game engagement. *Review of general psychology*, 14(2):154–166.

Rachels, J. R. and Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1-2):72–89.

Reeves, B. and Read, L. (2009). Total engagement: Using games and virtual worlds to change the way people work and businesses compete harvard business school publishing: Boston.

Seaborn, K. and Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of human-computer studies*, 74:14–31.

Shih, L.-H. and Jheng, Y.-C. (2017). Selecting persuasive strategies and game design elements for encouraging energy saving behavior. *Sustainability*, 9(7):1281.

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., and Akinkuolie, B. (2021). Gamification in mobile-assisted language learning: A systematic review of duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, pages 1–38.

Thiel, S.-K. (2016). A review of introducing game elements to e-participation. In *2016 conference for E-democracy and open government (CeDEM)*, pages 3–9. IEEE.

Toda, A. M., Valle, P. H., and Isotani, S. (2017). The dark side of gamification: An overview of negative effects of gamification in education. In *Researcher links workshop: higher education for all*, pages 143–156. Springer.

Zainuddin, Z. (2018). Students' learning performance and perceived motivation in gamified flipped-class instruction. *Computers & education*, 126:75–88.

Zainuddin, Z., Chu, S. K. W., Shujahat, M., and Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, 30:100326.

Zhonggen, Y. (2019). A meta-analysis of use of serious games in education over a decade. *International Journal of Computer Games Technology*, 2019.

Zichermann, G. and Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc.".