MASTER'S THESIS

Markus Schneider 2020

Copenhagen Business School
Department of Management, Politics and Philosophy
Master in Organizational Innovation and Entrepreneurship (M.Soc.Sc.)

Markus Schneider

Gamification: The effects of the game element countdown-timer on the engagement of European students

1st Supervisor: Dr. Claus Springborg

Copenhagen, 14th of May 2020

Number of pages: 80 pages

Number of characters: 180.000 characters with spaces

ABSTRACT

Gamification: The effects of the game element countdown-timer on the engagement of European students

Gamification research to date often lacks a coherent research practice that measures the effectiveness of gamification in motivating and engaging individuals. The present thesis addressed this gap in gamification research and applied the person-focused construct proposed by Landers, with the goal to measure the effect of one single game element. Within a post-positivist epistemology, a true experiment was performed to investigate the cause-effect relationship between the game element countdown-timer and the level of engagement of over 800 European students on a gamified online questionnaire. Within the independent-measures design, the treatment group was exposed to the countdown-timer while the control group experienced the identical questionnaire without the game element. Using a level of engagement-scale, the behavior of participants was measured. The presence or absence of the countdown-timer did not reveal significant differences in the level of engagement between and within the participants of treatment and control group. While gender as moderator was found to be indifferent to the incorporation of the countdown-timer, significant results were shown with age as moderator. Thus, European students belonging to Generation Z (≤24 years old) appear to be significantly more disengaged when not being exposed to gamification.

Keywords: gamification, game element, countdown-timer, engagement, experiment, gamified questionnaire, gender; age; generation z

ACKNOWLEDGEMENTS

It has been a long way with many delighting and challenging experiences. In this moment, I am just extremely grateful to those who have supported me during this chapter of life.

First and foremost, I would like to express my gratitude to Claus Springborg who supervised me for the present thesis: Thank you very much for your valuable advices and genuine support throughout the process. You inspired me to implement the research.

Furthermore, I am very thankful for having had the opportunity to pursue the master degree at Copenhagen Business School in Organizational Innovation and Entrepreneurship. It was one of the best decisions in my life and I am forever grateful to all the inspiration, drive and curiosity experienced together with my fellow students and lecturers of the OIE-program; as well as to the Danish society and the Danish people who I am able to call dear friends today.

My personal gamification journey started last year during my exchange semester at Queensland University of Technology in Brisbane, Australia. The lecture on gamification ignited the curiosity and passion within me to follow this path and will shape my future indeed. I will always be grateful for the knowledge gained, as well as experiences and connections made during this time of my life.

Moreover, the research results would not have been possible without the 884 participants who took the time to engage with the questionnaire, therefore I would like to send my greatest gratitude to students from all over Europe who contributed to the study.

Lastly, I cannot express in words how grateful I am for the love and support by my girlfriend Mirjam who supports me through my life to go for my dreams and reach for the stars. You know that I would not be at this point without you. Same goes for my beloved family, my mother and my father and especially for my beloved twin-sister Sandra who supported me greatly from the distance. When we have each other, we have everything.

Markus Schneider

Copenhagen, 14th of May, 2020

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

"Games are the new normal"

-- Former US Vice President Al Gore in 2011

Nine years ago, when the audience of the "Annual Games for Change" conference in New York City heard Al Gore put forward this remarkable quote, gamification as a concept was not widely known (Dymek & Zackariasson, 2016). Since then, over the past decade, gamification has become an emerging trend across various academic disciplines and industries. Defined as "the use of design elements characteristic for games in non-game contexts" (Deterding et al., 2011, p.9), it has been proposed that gamified practices, processes and application will become a key element in motivating individuals, "particularly those born in the digital age of online games" (Bai et al., 2020) while in fact; it is not a new concept. For example, gamification has been employed for long in educational institutions such as with levels, grades and degrees as well as in marketing endeavors with reward memberships or point cards (Nelson, 2012). The idea is that if the active ingredients that make games engaging are isolated, then practitioners can put those ingredients into an existing practice, and thus make them more engaging (Cugelman, 2013). Hereby, gamification is not itself a product, rather it consists of "[adding] game elements to change a process that already exists with [the aim] to change how that process influences people" (Landers et al., 2018, p.317).

Back in 2011, the research firm Gartner predicted that by 2014, over 70% of Fortune Global 2000 organizations would have adopted gamification by some means (Goasduff & Pettey, 2011), but that 80% of those efforts would ultimately fail to meet business objectives rooted in suboptimal design (Pettey & van der Meulen, 2012). These predictions were not too far from reality since still today, gamification is commonly misconceived in being capable of making everything fun (Burke, 2019). Consequently, jumping on this gamification bandwagon is a risky endeavor. However, it is not because gamification does not work, but rather, "because it is easy to get it wrong if [researchers and practitioners] do not understand what it is, know its limits, and make informed decisions on its application" (Cugelman, 2013; p.1). Yes, there are limits to what can be achieved with gamification – and that is why the current hype and broader trend needs a course correction (Burke, 2019).

The effectiveness of gamification has been tested in a variety of domains, generally showing positive results (Koivisto & Hamari, 2019) commonly in motivation and engagement (e.g. Mekler et al., 2017; Sailer

et al., 2016; Xi & Hamari, 2019). Gamification has been proven effective in contexts such as education and learning (e.g. Su & Cheng, 2018; Tsay et al., 2018), health and well-being (e.g. Matallaoui et al., 2017) and in the work place (e.g. Mitchell et al., 2018). While the results in general lean towards positive findings about the effectiveness of gamification, the amount of mixed results is remarkable (Koivisto & Hamari, 2019). Gamification often lacks a coherent research practice leading to interpretable and generalizable research findings (Koivisto & Hamari, 2019; Landers et al., 2018). The aim of the present thesis, therefore, lies in contributing to the field and the formerly described course correction by conducting a gamification study that is based on scientific rigor. It is the author's hope that the present thesis will change the current trajectory for gamification… even just a little bit.

1.1 Background and problem statement

Gamification research until today mainly consists of conceptual frameworks and theoretical constructs (Alsawaier, 2018; Kasurinen & Knutas, 2018; Seaborn & Fels, 2015), even though research has conducted and built up a remarkable amount of empirical research in the recent years (e.g. Kuo & Chuang, 2016; Sailer et al., 2017; Tsai et al., 2018; Xi & Hamari, 2019). Nevertheless, based on recent literature reviews, empirical research on gamification to this date has mainly focused on a few specific game elements, such as points, badges and leaderboards (Hamari & Sarsa, 2014; Seaborn & Fels, 2015) which consequently is the most prevalent criticism on gamification (Bogost, 2011; Kapp, 2012; Rehn, 2016). In addition to that, measuring the effectiveness of gamification becomes more complex the more game elements are combined in a gamified process. Gamification research to date, did not attempt to isolate the effect of e.g. a single game element (Armstrong et al., 2016). Furthermore, most of the empirical research on the topic was conducted without control groups and considerably few studies used identical or validated measurement instruments for particular variables that are likely to account for the effectiveness of gamification (Koivisto & Hamari, 2018). Landers et al.(2018), critically reflect on most empirical research studies, since variables such as mediators are mostly unspecified and unmeasured (Landers et al., 2018). Moreover, systematic literature reviews on gamification observed a disconnection between theory and applied research (Seaborn & Fels, 2015).

All in all, gamification studies often lack a coherent research practice that isolates the effects of single game elements and leads to interpretable and generalizable research findings (Koivisto & Hamari, 2019; Landers et al., 2018). Scholars have been calling for empirical studies with comparative and longitudinal

designs in gamification research for long in order to obtain a clearer picture on the effectiveness of gamification (e.g. Alsawaier, 2018; Deterding et al., 2011; Seaborn & Fels, 2015).

1.2 Research objective and question

Taking the described critical reflections and problems within gamification research into account, the present thesis aims at applying a gamification research practice that measures the effect of one single game element accurately, and thus leads to replicable and generalizable results. The four types of person-focused constructs studied within gamification science, according to Landers et al (2018). are game elements (predictors), targeted outcomes (criteria), intermediary individual changes (mediators), and personal and situational contexts (moderators) (Landers et al., 2018). Considering the critical reflections on gamification research to date and the lack of a coherent research practice as a solid starting basis for the investigation, the research objective of the study is formulated as follows:

> *To conduct proper gamification research grounded in the application of the four types of person-focused constructs in order to find out if gamification is effective in changing behavior within individuals.*

Considering this objective, the present research intends to provide contribution to both theoretical and practical implications and, the following defines the specific constructs for present study: The game element countdown-timer is selected as the initial causal force (predictor) of a distal target outcome (Landers et al., 2018). The countdown-timer is intentionally used to induce certain effects of interest which in the present study is an increased level of engagement within European students (criterion). The causal link between game elements and target outcomes is mediated by psychological and behavioral changes within the individual (Landers et al., 2018) which can be measured based on the students behavior within a questionnaire on the gamified landing page (mediator).

Consequently, the following research question was developed for the present study:

> *Can the use of countdown-timers on a questionnaire landing page increase the level of engagement of European students?*

In order to answer the research question, three hypotheses were developed based on the literature. H_2 and H_3 reflect on the fourth element of the person-focused construct, i.e. the moderator variables, which in the present study consist of differences in gender and age respectively. Accounting for these

moderators allows answering the research question more granular and enlarges the opportunities for future research. The three hypotheses are identified as follows:

> *H_1 = Users exposed to the game element countdown-timer show a higher level of engagement than users not being exposed to countdown-timers.*
>
> *H_2 = Male users are expected to show a higher level of engagement than female users when being exposed to the game element countdown-timer*
>
> *H_3 = Users younger than 25 years of age will show a higher level of engagement than users older than 25 years when being exposed to the game element countdown-time*

In order to test the described hypotheses and answer the research question with scientific rigor, the present thesis conducts a true experiment with an independent measure design by exposing European students (in the treatment condition) to the game element countdown-timer on a questionnaire landing page, while leaving it out for a control group of European students.

1.3 Delimitations of the study

Initial delimiting factors of this study are the aforementioned research objective and question. Furthermore, in consideration of resources available for the present study, the research is limited to measuring the effectiveness of one single game element, i.e. countdown-timers and does not empirically cover other possible game design mechanics and game elements. Narrowing down the scope of the present study, the researcher does not consider the cultural differences of the participants that may influence the level of engagement and motivation although participants of the experiment are originally from different countries and have different personalities.

Lastly, motivation and engagement are two complex concepts that are widely studied from different research perspectives in theory and practice. Therefore, the theoretical part of the present thesis is based on and limited to the following theories: self-determination theory (Deci & Ryan, 2000) and flow-theory (Nakamura & Csikszentmihalyi, 2014). In the context of gamification, these theories help to address the most essential factors that relate to changing the behavior of individuals based on higher levels of motivation or engagement.

1.4 Structure of the thesis

The present research is which is divided into eight chapters.

Chapter 1 represents the introduction of the thesis, where the background and problem statement of current gamification research guide the direction of the research objective and consequently, the formulation of the research question. The hypotheses to be tested as well as the delimitations are described in the introductory chapter.

Chapter 2 consists of a literature review providing a deeper understanding of existing research relevant to the research question of the present thesis. This theoretical chapter circumscribes the concept of gamification and its building blocks as well as a construct of gamification research applied to the game element countdown-timer.

Chapter 3 aims at outlining the researchers' methodological choices and explaining them in relation to the alternatives that were possible to adopt. Based on the post-positivist philosophy, the deductive research approach for theory development, with a mono method quantitative research design in the form of a true experiment will be explained and justified. Furthermore, the setup and execution of the true experiment will be described and potential threats to validity will be identified.

Chapter 4 presents the results and findings of testing the hypotheses in the experiment. Descriptive statistics will give the reader a clear and succinct summary of the data while inferential statistics will provide the reader with the results on to whether there are statistically significant differences between the groups and conditions of the experiment.

Chapter 5 advances with a discussion that is based on the empirical results and literature, which then leads to chapter 6 where limitations of the study are identified.

Chapter 7 is concerned with suggestions for future research while chapter 8 provides the practical implications of the study results. Chapter 9 rounds the thesis off with a short conclusion.

2. LITERATURE REVIEW

The literature review chapter aims at precisely defining gamification and its building blocks, addressing the current gap in research as well as introducing a construct of gamification research as the foundation of the empirical study.

2.1 Definition of the term 'gamification'

In the past decade, gamification has gained attention by academics, educators and practitioners from a variety of domains as a consequence of becoming an emerging trend within marketing and business in the first hand (Seaborn & Fels, 2015). The definition of gamification, however, still is under academic debate since the term has been used inconsistently in research. Nevertheless, a systematic literature survey conducted by Seaborn and Fels in 2015 showed an emerging definition by Deterding et al. based on the work of industry practitioners and academic which defines gamification as "the use of design elements characteristic for games in non-game contexts" (Deterding et al., 2011, p.9). Accordingly, with more than six thousand academic citations to date, it is one of the most cited papers in gamification research (Google Scholar, 2020). Deterding et al. unpack their definition by looking closely into the definitions' building blocks of *game, element, design and non-game contexts*. Furthermore, their research on the topic claims that gamifying includes a rule-bound and goal-oriented design that incorporates elements which can be found in most (not necessarily all) games (Deterding et al.). Alternative proposed definitions in academia regard the concept of gamification as "the process of game-thinking and game mechanics to engage users and solve problems" (Zichermann & Cunningham, 2011, p. 14), "the use of game mechanics and experience design to digitally engage and motivate people to achieve their goals" (Burke, 2016), or "a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation" (Huotari & Hamari, 2012, p.19).

For the purpose of the present thesis, I will coherently use the definition of Deterding et al. that is "the use of game design elements in non-game contexts" (Deterding et al., 2011, p.9). First of all, it is the definition that allows for the broadest application of gamification. Secondly, in this framework, the core of gamification is a design process (Deterding et al., 2011; Deterding, 2015) to enrich or augment a user experience with a specific change in the outcome (Landers et al., 2018). Consequently, it is strongly related to the research question of this present thesis, which is concerned about the effects of

gamification on a questionnaire landing page. Finally, and most importantly, as mentioned before, the definition by Deterding et al. is the most commonly used among researchers and practitioners (Seaborn & Fels, 2015).

In order to develop a deeper understanding of what gamification is, the following sections will unpack the definition of Deterding et al. in its building blocks of *game, game element, design and non-game contexts*.

2.2 Game-Design and Game Elements

A common denominator for all influential academic definitions on gamification is the "intentional use of game elements" (Seaborn & Fels, 2015, p.17), i.e. when gamifying, these elements are intentionally used by the designer with the "goal to create a specific change in a distal target outcome" (Landers et al., 2018, p.317). Game elements can be regarded as building blocks or features that are common or characteristic for games, such as the "ten ingredients of great games" (Reeves & Read, 2009). There is neither a strict nor boundless categorization, as Deterding et al. define it, rather a general limitation to elements that are found in most but not all games such as patterns, principles, objects and models (Deterding et. al., 2011; Seaborn & Fels, 2015).

Based on recent literature reviews, empirical research on gamification to this date has mainly focused on a few specific game elements, such as points, badges and leaderboards (Hamari et al., 2014; Seaborn & Fels, 2015) which consequently is the most prevalent criticism on gamification (Bogost, 2011; Kapp, 2012; Rehn, 2016). However, gamification is much more than points, badges and leaderboards, as pointed out by numerous scholars (e.g., Blohm & Leimeister, 2013; Chou, 2019, Kapp, 2014). It operationalizes game elements such as narratives, avatars, teams, time pressure, competitions, boss fights or constraints – the list could be continued for much longer and is only a small fraction of the differently categorized game elements in game science (Chou, 2015; Reeves & Read, 2009; Werbach & Hunter, 2012).

As the variety of elements is characteristic for the multiplicity of gamification (Seaborn & Fels, 2015), it is not possible for scholars and practitioners to agree on one universal framework for game elements, rather there is a variety of frameworks based on the differences in perspectives, such as the perspective of game developers using the Mechanics, Dynamics, and Aesthetics (MDA) framework (Hunicke et al., 2004) or the perspective of procedural content generation that categorizes game elements into six classes including game bits, game space, game systems, game scenarios, game design, and derived content (Hendrikx et al.,

2013). Consequently, game elements can be borrowed from countless frameworks within sub disciplines of game science.

The operationalization of game elements is continuously the same: They are used as "causes of effects of interest in processes that have been gamified" (Landers et al., 2018, p.321). More precisely, after gamification, more game elements are present in the specific process than before, ranging from one simple addition of a game element or the combination of many, however the latter rendering empirical research more complex (Armstrong et al., 2016). Generally, the boundary between "a game" and an "application with game elements" can be blurry, giving it subjective and socially constructed meanings (Deterding et al., 2011, p.11). Some individuals might "play" while others "use" a gamified application, depending on informal rules or shared goals. Therefore, in relation to the research question, the selected game element is chosen purposefully for the specific outcome of interest since it functions as predictor in gamification research (Landers et al., 2018), which will be elaborated more upon in chapter 2.4.

2.3 Gamification in non-game contexts

Per definition gamification applies game elements to contexts "other than games", meaning outside of entertainment, thereby leaving it open to proliferate in all kinds of contexts, purposes and scenarios (Deterding et al., 2011). Gamification could theoretically be used in the context of games, however, it then would simply be considered game design, not gamification (Deterding et al., 2011). All in all, the concept of gamification has been tested and applied in many different contexts, including *education and learning* (e.g. Caponetto et al., 2014; De Sousa et al., 2014; Dicheva et al. 2015; Dichev & Dicheva, 2018), *participation in government* (e.g., Bista et al., 2014), *health and well-being* (e.g., Johnson et al., 2016; Pyky et al., 2017; Zuckerman & Gal-Oz, 2014), *marketing* (e.g. Hamari, 2017; Hofacker et al., 2016; Huotari & Hamari, 2017), and *management* (e.g., Deterding, 2019; Mekler et al., 2017; Stanculescu et al., 2016).

A systematic mapping study looking at publication trends in gamification literature revealed that *education* was the most common overall theme of the 1164 papers studied (Kasurinen & Knutas, 2018). Proof-of-concept prototypes, theoretical papers on the application and papers examining eLearning tools such as massive open online courses were the most common types of research. Although being a recently published paper (2018) the data collected 'only' contains the status quo of research until 2015. Supporting these findings, a systematic survey of the academic literature on gamification by Seaborn and & Fels

(2015) showed that within applied research the concept of gamification is used multidisciplinary and across many fields while the top fields of research were again *education* and *health* (Seaborn & Fels, 2015).

In the present thesis, gamification will be used to incorporate the game element countdown-timer to an online questionnaire in the context of general self-reflection.

2.4 Effectiveness of Gamification

The effectiveness of gamification has been tested in a variety of domains, generally showing positive results (Koivisto & Hamari, 2019) commonly in motivation and engagement (e.g., Kuo & Chuang, 2016; Sailer et al., 2016; Mekler et al., 2017; Xi & Hamari, 2019) often due to an increase in the satisfaction of psychological needs (e.g., Xi & Hamari, 2019; Mitchell et al., 2018; Zuckerman & Gal-Oz, 2014) or flow (Chan et al., 2019; Eickhoff et al., 2012). Gamification has been proven effective in contexts such as education and learning (e.g. Larson, 2020; Tsay et al., 2018; Su & Cheng, 2018; Da Rocha et al., 2016), crowdsourcing (Morschheuser et al., 2016), health and well-being (e.g., Matallaoui et al., 2017; Pyky et al., 2017) and in the work place (e.g., Mitchell et al., 2018).

While the results in general lean towards positive findings about the effectiveness of gamification, the amount of mixed results is remarkable (Koivisto & Hamari, 2019). Gamification often lacks a coherent research practice that leads to interpretable and generalizable research findings (Koivisto & Hamari, 2019; Landers et al., 2018). In their systematic journey of academic literature on gamification Seaborn and Fels (2015) also observed a disconnection between theory and applied research (Seaborn & Fels, 2015). In order to obtain a clearer picture on the effectiveness of gamification, scholars for long have been calling for empirical studies with comparative and longitudinal designs in gamification research (e.g. Alsawaier, 2018; Deterding et al., 2011; Kuo & Chuang, 2016; Seaborn & Fels, 2015).

Even though research has conducted and built up a remarkable amount of empirical research in the recent years (e.g. Kuo & Chuang, 2016; Sailer et al., 2017; Tsai et al., 2018; Xi & Hamari, 2019), Koivisto & Hamari (2018) concluded after reviewing more than 270 empirical studies in their gamification research review, that most of the empirical research on the topic is conducted without control groups and that considerably few studies use identical or validated measurement instruments for a particular variables that are likely to account for the effectiveness of gamification (Koivisto & Hamari, 2018). Landers et al.

(2018), too, critically reflect on most empirical research studies, since variables such as mediators are mostly unspecified and unmeasured (Landers et al., 2018).

Taking the described findings and critical reflections into account, the present thesis and its empirical research employs variables that measure the effectiveness of gamification within an the experiment with control groups in order to investigate the research question based on scientific rigor and accuracy.

2.5 Gamification Research

2.5.1 Theoretical Concept of Gamification Research

The four types of person-focused constructs studied within gamification science, according to Landers et al., are game elements (predictors), targeted […] outcomes (criteria), intermediary individual changes (mediators), and personal and situational contexts (moderators) (Landers et al., 2018). The construct is illustrated in Figure 1. Game elements such as e.g. points, narratives or competitions are the initial causal force of a distal target outcome (Landers et al., 2018), they are "intentionally used" (Seaborn & Fels, 2015, p.17) to induce certain effects of interest. These effects of interest can be outcomes such as improved student learning (e.g. Su & Cheng, 2018; Tsay et al., 2018). The causal link between game elements and target outcomes is mediated by psychological and behavioral changes within the individual (Landers et al., 2018). It is related to the findings of Hamari et al. in their literature review from 2014, where they explained how gamification works, i.e. by psychological mediators being causally interlinked with behavioral mediators that in turn lead to changes in the outcome (Hamari et al., 2014).

As an example, the empirical research conducted by Su & Cheng showed that by incorporating gamification into the learning process, students showed improved learning and achievements (criterion) due to higher motivation (psychological mediator) (Su & Cheng, 2018). Furthermore, in an empirical study with undergraduate students, Tsay et al. (2018) found that by incorporating several game elements such as e.g. leaderboards, badges and feedback mechanisms in their learning activities, the intrinsic and extrinsic motivation of students increased based on self-determination theory (Deci & Ryan, 2000). This psychological state of being more motivated mediated the behavioral change of a higher participation and engagement rate which in turn mediated the overall increased course performance compared to the control group that didn't use the gamified but more traditional learning delivery (Tsay et al. 2018).

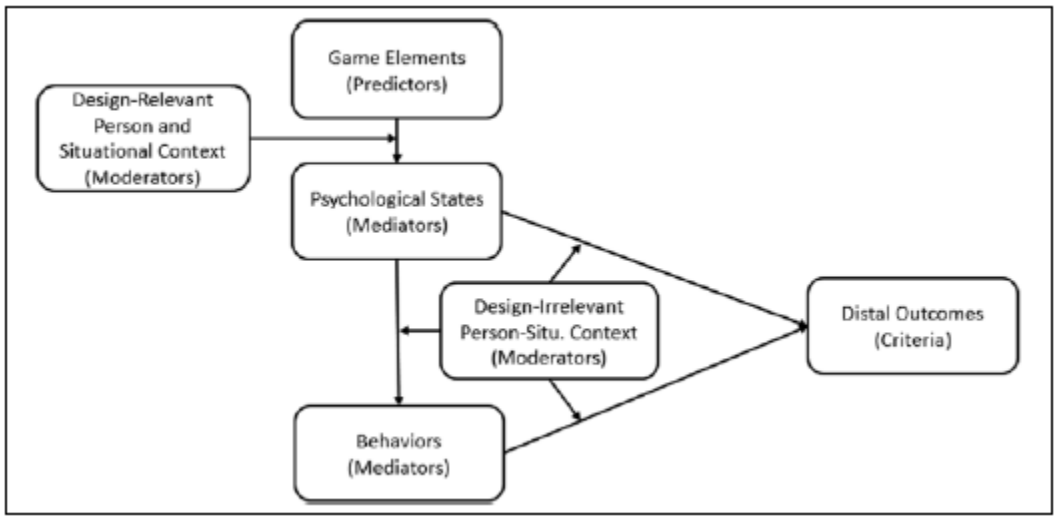


Figure 1: Theoretical causal relationships between constructs in gamification science (Landers et al., 2018, p.320)

As illustrated in Figure 1, each of these linked effects might be moderated by other variables, be it the general context of the study or variables within the individuals participating, which in turn influence the psychological or behavioral states as well as the target outcomes in vice-versa direction or strength (Lander et al., 2018). All in all, the described causal relationships between the four classes of constructs, predictors, criteria, mediators and moderators, are essential to reliably and successfully researching and applying gamification in various settings.

Thus, the following sub chapters will look into each of the four constructs and critically reflect on their application in research to date. Most importantly, these constructs will be investigated in relation to the research question. The game element countdown-timer (predictor) will be incorporated into a landing page to cause the outcome of interest measured in increased level of engagement (criteria) through the increased psychological state of engagement and the completion of the questionnaire (both mediators) within the personal and situational context of European students and different gender (moderators). Clearly defining and measuring the relationships between these constructs provides the foundation for reliable and valid gamification research (Landers et al., 2018).

2.6 Predictor Variables in Gamification: Game elements

Game elements, as previously defined, are elements that are *characteristic* to games which translate into elements that are used in most, but not all games (Deterding et al., 2011). In the context of gamification,

game elements are used with the intention to induce certain effects of interest, i.e. “to create a specific change in a distal target outcome” (Landers et al., 2018, p.317). Thus, based on Pearl’s work on causality, gamification research aims at drawing causal conclusions regarding the use of game elements (Pearl, 2000).

In general, after applying gamification, there are more game elements found in a process or application than before, ranging from one simple addition of a game element or the combination of many, however, incorporating many game elements renders empirical research more complex as casual conclusions become more difficult to establish (Armstrong et al., 2016). Gamification researchers conducting empirical research are concerned with investigating differences and continuities in outcomes after incorporating game elements into a specific context. Therefore, it is crucial to control for the difference in outcome with and without the use of specific game elements. Thus, game elements need to be clearly defined and used intentionally (Landers et al., 2018). For the present thesis, the game element countdown-timer will be of specific interest as the research question aims at finding out whether the incorporation of countdown-timers lead to a higher engagement level within European students. Therefore, in the following sub-chapter, the concept of the game element of countdown-timer will be defined and its application as predictor variable will be elaborated.

2.6.1 Game Element Countdown-Timer

In the framework of Werbach & Hunter (2012), game elements are segmented into smaller building blocks - game dynamics, game mechanics and game components - according to their decreasing order in the level of abstraction (Werbach & Hunter, 2012). Game dynamics constitute the highest level of abstraction and set the concept of the gamified process or application. In regard to countdown-timers c*onstraints* is the game dynamic that imposes barriers of interaction on users between the latter or as the rules of general enactment (Werbach & Hunter, 2012). The game dynamic of constraint can be linked to the lower abstraction game mechanic of challenge, competition and feedback, that requires users’ effort to solve certain tasks, that users can win or lose (e.g. against other users or against time); and inform users on how they are doing respectively. Game mechanics are defined as “basic processes that drive action forward and generate player engagement” (Werbach & Hunter, 2012, p.79), and can be manifested through the game component countdown-timer. Timers (invisible or visible countdown clocks in a digital game) “give

players specific durations in which they can achieve goals […] as well as regulate the pace and rhythm of a game" (Adamou, 2018, p.63).

Even though countdown-timers are addressed less in literature compared to other game elements, the incorporation of countdown-timers is found within several gamified applications and processes (Butler, 2015; Economou et al., 2015; Khaleel et al., 2017; Matenga et al., 2018; Pettit et al., 2015; Sun et al., 2018; Zhang et al., 2018). Economou et al. (2015), for example, used a countdown-timer to increase the level of challenge posed by each (application) scenario (Economou et al., 2015). By users having to 'play' against time to meet certain requirements, countdown-timers were suggested to be effective in enhancing the engagement of users, especially when being combined with a potential loss that users try to avoid such as negative marking and loosing points or badges that have been gained in the process already (Economou et al., 2015). In the study of Matenga et al. (2018), researchers supported the claim that a timer mechanic can affect time spent on tasks by users as the timer increased the average completion times of certain tasks related to programming (Matenga et al., 2018). Furthermore, several studies incorporated countdown-timers along with other game elements such as leaderboards, progress bars and reward points to achieve a specific change in participants' behavior, such as motivation and engagement (Khaleel et al., 2017; Sun et al., 2018; Pettit et al., 2015; Zhang et al., 2018), learning outcome (Jagušt et al., 2018) as well as for organizational and managerial purposes (Butler, 2015; Garbaya et al., 2019).

Outside of gamification research, studies concerned with the influence of countdown-timers found that the latter can influence driver behaviors in terms of decisions to stop or cross intersections (Limanond et al., 2009; Long & Yang, 2011). Furthermore, countdown-timers are often used within marketing and advertisement for the purpose of generating a sense of scarcity and urgency within potential customers and consequently change their engagement and buying behavior (Cialdini, 1993; Kim et al., 2020).

All in all, research to date has not addressed countdown-timers much in comparison to other game elements such as avatars or leaderboards, which is surprising in the perspective of Adamou (2018) since "timers are key in engaging players because they help to build emotion (such as anticipation, relief and urgency)" (Adamou, 2018, p.63). Having defined the concept of countdown-timer and its application within gamification, the focus will now shift towards the outcomes of interest in gamification caused by game elements in general and the relevant outcomes of interest related to the research question of the present thesis.

2.7 Criterion: Outcomes of Interest in Gamification

Gamification is applied in various contexts and domains as described previously (e.g. in education; Tsay et al., 2018); therefore, a variety of target outcomes can be pursued by gamification researcher-practitioners. Nevertheless, changes in the outcome of interest generally meet organizational needs, such as for businesses or educational institutions (Landers et al., 2018). Examples of empirically studied distal outcomes are return on investment (Conley & Donaldson, 2015), improvement in academic achievement (Su & Cheng, 2015) or increased online education retention rates (Krause et al., 2015). In theory, there are infinite possibilities of domains, contexts and thus, distal outcomes where gamification can be applied. All in all, as gamification researcher-practitioner, it is important a) to clearly define the distal target outcome of gamified interventions; and b) empirically investigate whether the incorporated game elements failed or succeeded in achieving the changes in the outcomes of interest (Landers et al., 2018).

2.7.1 Outcome of interest in the gamified landing page

The desired distal target outcome of the thesis project is to achieve an increased level of engagement within European students on a landing page, caused by the implementation of a countdown-timer, compared to a landing page that is not gamified, i.e. in absence of a countdown-timer or any other game element. Measuring the level of engagement based on the implementation of gamification provides valuable insights on how to more effectively engage with European students. In turn, new possibilities open up for educators, marketers and decision-makers such as in presenting information or invoking specific actions within target audience more effectively. Nevertheless, gamification researcher-practitioners need to be careful in evaluating the success or failure of gamification interventions based on the sole changes in the outcome of interest criteria (Landers et al., 2018). To make informed and correct evaluations whether gamification efforts were effective or not; and, therefore, to answer the research question most adequately, changes in mediator and moderator variables are of crucial importance and will be described in the following chapters.

2.8 Mediator Variables in Gamification

When gamifying a process or application, the designer intents to change human behavior to achieve the target outcomes, i.e. through cause-effect relationships between the predictor variable and outcome criteria previously described (Landers et al., 2018). Establishing these relationships between variables is

important, because correlation is a necessary condition for claiming that two variables are causally related (Preacher & Hayes, 2008). Furthermore, these cause-effect relationships evoke the idea of mediation, which is "the process by which some variables exert influences on others through intervening or mediator variables" (Preacher & Hayes, 2008; p.879). It is of large scientific interest to explain how or by what means (=mediating variables) a causal effect occurs.

Thus, for gamification researchers, it is of interest to understand how to most effectively alter human behavior, attitudes and other psychological states with the incorporated game elements (Landers et al., 2018). This is at the core of any gamification intervention, since gamification works by psychological mediators being causally interlinked with behavioral mediators that in turn lead to changes in the outcome (Hamari et al., 2014). For example, a gamification designer incorporates experience points (predictor) to increase engagement and motivation (mediator; psychological state), thus alters a learners' learning intensity (mediator; behavioral state) which in turn effects (positively) the outcome in a test (criteria). As shown in the described example and in Figure 1, mediators can and do occur in series (Hayes & Preacher, 2010). Being relevant to the research question, it is important to outline the possible direct and indirect relationships between predictors, mediators and outcomes (Landers et al., 2018):

1. Game elements may have a causal, *direct effect* on psychological states.
2. Psychological states may have a causal, *direct effect* on behaviors.
3. Both psychological states and behaviors may have a causal, *direct effect* on outcomes.
4. Game elements may have a causal, *indirect effect* on target outcomes via the intermediary causal effect of psychological states.
5. Game elements may have a causal, *indirect effect* on target outcomes via the intermediary causal effects of psychological states on behaviors and behaviors on outcomes.

Within gamification research, a multiplicity of mediating variables such as motivation (e.g., Kuo & Chuang, 2016; Sailer et al., 2016; Mekler et al., 2017; Xi & Hamari, 2019) or flow (e.g. Chan et al., 2019) has been studied as explanatory link between game elements and the attained effects. Nevertheless, Landers et al. critically reflect on most empirical research studies within the gamification, since mediating variables were mostly unspecified and unmeasured (Landers et al., 2018). Furthermore, the more game elements are

incorporated or even combined, the more complex it becomes to establish causal relationships between the variables (Armstrong et al., 2016).

Applied gamification science with empirical studies based on specified and measured mediating variables is a relatively young field and still in its initial stages (Landers et al., 2018). Thus, it is of interest to conduct a study based on the research question that defines and specifies the mediator variables as its effects depict the key pillar when investigating the effectiveness of gamified processes or applications.

2.8.1 Motivation and Engagement as Mediator Variables

Within gamification research, the psychological state of motivation as a mediating variable has been studied extensively being the explanatory link between game elements and the achieved outcomes of interest (e.g., Kuo & Chuang, 2016; Mekler et al., 2017; Sailer et al., 2016; Xi & Hamari, 2019). In order to apply the psychological theories underlying the mediation of motivation most effectively, they need to be deeply understood before.

Self-determination theory (SDT), as proposed by Ryan & Deci (2000) builds the foundation for understanding the interplay between extrinsic elements acting on individuals and the intrinsic motives and needs inherent in every person (Ryan & Deci, 2000). According to SDT, motivation requires the satisfaction of three basic psychological needs: autonomy, competence and relatedness, thus improving these needs of the user can increase their intrinsic motivation to complete a task (Forde et al., 2016; Ryan & Deci, 2000). Autonomy is found by giving users a sense of control and freedom over one's actions. Competence is achieved when users feel like they are progressing towards a meaningful goal and relatedness allows users to connect to an application as well as with other users (Richards et al., 2014; Sailer et al, 2017). The feeling of competence can also be connected to the theory of flow (Nakamura &Csikszentmihalyi, 2014) where the balance of challenge and skill provides the optimal condition to experience flow and therefore the satisfaction of competence need in gamified contexts (Chan et. al., 2019; Hamari & Koivisto, 2014).

Furthermore, according to the results of an empirical study conducted by Xi & Hamari (2019), i) immersion-related features are positively associated with autonomy need satisfaction; ii) achievement-related features are not only positively associated with all forms of need satisfaction, but also the strongest predictor of both autonomy and competence need satisfaction; and iii) in regard to social-

related features, it has positive associations with autonomy, competence and mostly relatedness need satisfaction (Xi, N. & Hamari, J., 2019).

In another study, Mekler et al. (2017) found that game elements did not decrease intrinsic motivation, but on the contrary improved tag quantity without increasing intrinsic motivation or competence need satisfaction. At the same time, in this particular study context, the game elements points, levels and leaderboards may have had a positive effect on extrinsic motivation of the participants (Mekler et al., 2017). Overall, research shows the importance of integrating and internalizing extrinsic motivation that is congruent with the users' intentions in order to achieve a positive effect on users, based on the theory of SDT (Ryan & Deci, 2000).

Furthermore, research implies that external rewards, such as points, leaderboards, badges and performance graphs can positively affect the perceived competence need satisfaction and thereby intrinsic motivation of the user. (Mekler et al., 2017; Sailer et al., 2017) Moreover, Landers et al. found in an experiment that leaderboards on tasks increase performance as well as that goal-setting theory moderates the success of leaderboards (Landers et al., 2017). Additionally, feedback mechanisms, such as reward elements i.e. points and leaderboards, provide users with a feeling of competence which increases intrinsic motivation, especially in the presence of autonomy, however only under the condition of being perceived rather informational as controlling (Forde et al., 2016).

2.8.2 Motivational Effect of countdown-timers in a gamified context

The motivational effects of countdown-timers can be related to the psychological need satisfaction of competence and autonomy which influence intrinsic motivation, to the feeling of flow as well as to extrinsic motivation.

Firstly, the psychological need of competence is satisfied when users feel like they are progressing towards a meaningful goal (Richards et al., 2014; Sailer et al, 2017). In this regard, completing a task such as filling out a questionnaire before the countdown-timer ends is suggested to provide users with a feeling of accomplishment. Thus, countdown-timers can act as incentive assuming the goal, e.g. filling out a questionnaire or completing a task, is perceived as meaningful by user (Richards et al., 2014).

Secondly, the psychological need of autonomy is found by giving users a sense of control and freedom over one's actions (Ryan & Deci, 2000). Since countdown-timers let users know exactly how much time

they have to finish a particular task, it increases anticipation; and consequently, users will feel more in control (Adamou, 2018). The latter can increase the intrinsic motivation of users based on the satisfied psychological need of autonomy. Nevertheless, it can make a difference on intrinsic motivation whether the non-static presence of countdown-timers is perceived as informational and not controlling feedback, as shown with leaderboards and points (Forde et al., 2016). If the incorporated countdown-timers are perceived as informational, the felt satisfaction of both autonomy and competence is magnified (Sailer et al., 2017).

Moreover, the motivational and engaging potential of countdown-timers can also be connected to the theory of flow by Csikszentmihalyi & Nakamura (2014) where the balance of challenge and skill provides the optimal condition to experience flow and therefore the satisfaction of the competence need in gamified contexts (Chan et. al., 2019; Hamari & Koivisto, 2014). Countdown-timers on a gamified landing page, therefore, need to optimally challenge users in order to convey a feeling of competence. Consequently, the more the average time for completion of the particular activity on the gamified landing page is perceived as "just-manageable", the higher will the perceived feeling of competence and the level of engagement be (Nakamura & Csikszentmihalyi, 2014, p.90).

Furthermore, time limits can "discourage deliberation and top-of-mind responses" (Amadou, 2018, p.64), which can be relevant for specific research approaches. As "Harvard Implicit Association Tests" (or IATs) have shown, these types of responses are more truthful as they reveal hidden biases that individuals might not even be aware of (Adamou, 2018; Greenwald et al., 2003).

Lastly, having set a realistic task completion time might reduce the possibility of a splintered focus, as shown with the productivity technique 'pomodoro' (Cirillo, 2006). Consequently, users would feel less inclined to procrastinate or waiver their attention, thus minimizing to drift between one task to another if they only have a certain window in which to complete a task (Adamou, 2018).

As literature suggests, the presence of countdown-timers is suggested to increase the level of engagement within users. Consequently, the first hypothesis of the study can be derived: *Users exposed to the game element countdown-timer show a higher level of engagement than users not being exposed to countdown-timers.*

2.9 Moderator Variables: Circumstances in Gamification Research

Gamification works differently in different contexts or on different individuals – the personal or situational circumstances exert influence on the outcome of interest. These influences can be understood through the lens of moderator variables. Per definition, moderator variables affect the direction and/or strength of the relationship between a predictor and outcome (Baron & Kenny, 1986). Thus, the effect of the moderator variable determines the effect of a specific predictor variable on the achieved outcome of interest (Cohen et al., 2013). As a consequence, taking moderation effects into account inform better practice in research, especially in regard to generalizability and external validity of empirical results (Fairchild & McQuillin, 2010). Examples for moderator variables are gender, age, personality or organizational conditions (Cohen et al., 2013). They can either be categorical (e.g. gender or race) or continuous (e.g. attitudes, personality), as well as either organismic (e.g. human perceptions) or situational (e.g. environmental conditions, economic crisis) (Cohen et al., 2013).

Figure 1 illustrates the moderator variables in a gamification construct and labels them as person and situational context (Landers et al., 2018). They can be either design-relevant or design-irrelevant. Design-relevant moderator variables influence the effect of game elements on targeted psychological states, such as that users who are over-exposed to leaderboards might have negative attitudes toward the concept of leaderboards, thus decreasing the users' engagement and altering the strength of the effect (Landers et al., 2018). For example, in an empirical study conducted by Mekler et al. in 2017, the researchers examined the effect on individual gamification elements on participants and found that general attitude was a moderator of the tag quantity generated by participants (Mekler et al., 2017). Furthermore, the addition most common game elements associated with gamification such as points, levels, badges "may help in some learning contexts, but harm in others" (Landers et al., 2015). Hence, the effectiveness of particular game elements is always moderated by the specific context and circumstances.

Generally, it is important to consider in any gamification design that the effect of game design elements seems to depend on how well the design was implemented in terms of aesthetics and quality, making both design-relevant moderators for gamification intervention, too (Sailer et al., 2017). Design-irrelevant moderators affect the strength of relationships after changes in the psychological state of users have been achieved. In an organizational setting, for example, gamification might have been effective by altering psychological states such as engagement, however, decreased leadership support for certain behavior

such as creativity might still alter the overall desired target outcome. Therefore, in order to conduct proper gamification research and not be misled, both design-relevant and design-irrelevant moderators need to be considered and are critical to assessing the effectiveness of gamification interventions (Landers et al, 2018).

2.9.1 Moderators in the present context

As gamification is still a novel phenomenon, several moderator variables have not been investigated in depth yet. For the present context, therefore, the categorical, demographic moderator variables of gender and age will be taken into consideration. Aptly so, both age and gender generally have not yet received great attention as moderating variables in both information technology literature (Gefen & Straub, 1997; Sun & Zhang, 2006) as well as in the context of gamification literature (Koivisto & Hamari, 2014). Both age and gender represent perspectives wherein variation and preferences have been long disregarded by the industry and by academics (Greenberg et al., 2010; Griffiths et al., 2003; Williams et al., 2008).

2.9.2 Gender as moderator variable

Generally, within literature to date, the inclination towards gamification is suggested to be independent of gender (Brauner et al., 2013), noting that great variety in motivation most likely exists also within genders (Carr, 2005; Kafai, 1998). However, in a more granular perspective, gamification is supposed to motivate and engage genders differently based on research on gender roles. The latter indicates that men display more instrumental behavior (Spence & Helmreich, 1980; Venkatesh et al., 2000), and generally, are more task- and achievement-oriented than women (Hoffman, 1972; Minton & Schneider, 1980). Woman, as research shows in contrast to the former proposition, are more motivated by affiliation (Hoffmann, 1972) and generally more interpersonally-oriented than men (Minton & Schneider, 1980; Spence & Helmreich, 1980). Thus, women are more concerned with social relations while men are oriented more towards competition and achievements (Venkatesh et al., 2000; Williams et al., 2008). These past research findings on genders differences are highly relevant to the influencing factors of motivation and engagement in the context of gamification. As the commonly applied theory of motivation in the field, self-determination theory (SDT) proposes that motivation requires the satisfaction of the three basic psychological needs: autonomy, competence and relatedness (Ryan & Deci, 2000). Within gamification research of motivation and engagement, the fulfillment of the psychological need of competence has been shown to influence

the behavior of men more than women while women are more influenced by the satisfaction of the psychological need for relatedness (Koivisto & Hamari, 2014; Oyibo, 2017; Yee, 2006).

The game element countdown-timer mainly touches the psychological need of competence since completing the questionnaire before the countdown-timer ends is potentially provides users with a feeling of accomplishment and competence, thus engaging them more in the experiment. The psychological need of competence is fulfilled through competition, winning and achievements. All these characteristics are shown to influence male behavior more than female (Koivisto & Hamari, 2014). Furthermore, the countdown-timer has no social extensions attached to it which seems to engage women more than men. Based on this proposition, male users are expected to show a higher engagement than female participants in the study to be conducted. Consequently, the following hypothesis can be derived: *Male users show a higher level of engagement than female users when being exposed to the game element countdown-timer*.

2.9.3 Age as moderator variable

In the present context of the thesis, age as a moderator variable will be distinguished between the two generations Generation Z and Generation Y, which are shown to have different characteristics and preferences, thus moderating the effectiveness of the incorporation of the game element countdown-timer differently.

Generation Z refers to individuals born from 1995 through 2010 (Bassiouni and Hackley, 2014; Fister-Gale, 2015; Seemiller & Grace, 2016), therefore ranging from 10 to 25 years of age in 2020 and making up the majority of higher education today (Škuta & Kostolányová, 2016). Members of Generation Z are often referred to as digital natives aptly to their world being completely shaped by technology (Seemiller & Grace, 2016; Twenge, 2017). Using the internet and social media has become part of their life and socialization (Ding & Yu, 2017); and technology is seen as instrument (Van den Bergh & Behrer, 2016).

Its predecessors, the Millenials Generation, or also called Generation Y, were born between 1980 and 1994 (Berkup, 2014; Seemiller & Grace, 2016). Millenials have had nearly a lifetime of exposure to technology (Seemiller & Grace, 2016) and seem to have an inclination towards technology, however, in comparison to Generation Z, Millenials were rather being accustomed to than born with it (Berkup, 2014).

On the other hand, a distinctive trait of individuals within Generation Z lies in the fact that they face difficulties staying focused and keeping attention (Bíró, 2014; Berkup, 2014; Eck, 2006; Seemiller & Grace, 2016). As previous research has highlighted, the application of game elements and gamification in

handling the specific needs of Generation Z is considered to be useful to infuse motivation, enhance learning and, most importantly, increase focus and attention (Biro, 2014; Geck, 2007; Sanmugam et al., 2016; Skinner, 2018).

Theory on the extent to which age or generational discrepancies do exist in the context of gamification are not clear and have been disregarded for long by the industry and academics (Greenberg et al., 2010; Koivisto & Hamari, 2014). Hence, age discrepancies in the context of gamification are assumed to be similar to those in other digital gaming contexts (Koivisto & Hamari, 2014). As digital gaming theory suggests prior experience with digital games could have a positive effect on the level of engagement as well as on performance within a gamified activity or experience (Bittner, 2013), younger age groups are likely to perceive more flow and enjoyment as well as usefulness with gamified experiences compared to older age groups (Bittner & Shipper, 2014). Furthermore, the findings of Bittner & Shipper (2014) showed that more experience with games can lead to a higher perceived control of gamified services (Bittner & Shipper, 2014). Hence, prior gaming experience potentially provides more familiarity with specific game mechanics such as countdown-timers. A higher degree of perceived control is related to the psychological need of autonomy (Ryan & Deci, 2000); and thus might increase the level of engagement with users.

Based on literature, students below the age of 25 years, i.e. members of Generation Z, are expected to have more prior experience with technology, games and gamification than students older than 25 years old (Generation Y). As a consequence, it is assumed that age positively moderates a difference in the perceived control over the experience, and thus in the level of engagement when being exposed to the game element countdown-timer. Hence, the following hypothesis needs to be tested: *Users younger than 25 years of age show a higher level of engagement than users older than 25 years when being exposed to the game element countdown-time.*

2.10 Summary of the hypotheses

The present research question aims at finding out whether or not the use of countdown-timers on a questionnaire landing page can increase the level of engagement of European students. In order to answer the research question, three main hypotheses were developed based on the literature: Firstly, it is assumed that users exposed to the game element countdown-timer show a higher level of engagement than users not being exposed to countdown-timers. Secondly, with gender being the moderator variable,

male users are expected to show a higher level of engagement than female users when being exposed to the game element countdown-timer.

Based on generational differences as moderator variable, the final hypothesis lies in the assumption that users younger than 25 years of age will show a higher level of engagement than users older than 25 years when being exposed to the game element countdown-time.

In the following, the researcher will describe the scientific actions to be taken to investigate the present research problem and test the hypotheses, as well as the background and rationale of the chosen research design.

3. RESEARCH METHODOLOGY

3.1 From Research Philosophy to Research Strategy

The present chapter aims at outlining the researchers' philosophical choices and explaining them in relation to the alternatives that were possible to adopt. Based on the philosophical choices, the research approach for theory development, the methodological choice as well as the research strategy will be explained and justified.

3.2 Research Philosophy

Research philosophy refers to a "system of beliefs and assumptions about the development of knowledge" (Saunders, 2009, p.124). Researchers make a number of types of assumptions at every stage, whether being consciously aware of them or not (Burrell & Morgan, 2017). These mainly include assumptions about the realities encountered in research, i.e. ontological assumptions, and about human knowledge, i.e. epistemological assumptions (Saunders, 2009, p.124).

Ontology as a branch of philosophy is the science of what is, of the natures and structures of reality (Guarino et al., 2009; Saunders 2009; Smith, 2012). Consequently, social ontology is concerned with the nature and structures of social actors (Bryman, 2016). Social actors can be considered as "objective entities that have a reality external to social actors" or as "social constructions built up from the perceptions and actions of social actors" (Bryman, 2016, p.28). Hence, positivism (objectivism) and constructionism (subjectivism) are referred to as two important positions within ontology (Bryman, 2016; Lincoln et al., 2011). All in all, the researchers' ontological assumptions shape the way in which one sees and studies one's research objects (Saunders, 2009, p.127).

Epistemology as a branch of philosophy deals with the sources of knowledge that questions how knowledge is acquired (Bryman, 2016). It concerns assumptions about knowledge, i.e. what does and does not constitute acceptable, valid and legitimate knowledge (Burrell and Morgan, 2017; Hallebone & Priest, 2009). An epistemological issue within the social sciences is whether the social world can and should be studied "according to the same principles, procedures and ethos as the natural sciences" (Bryman, 2016, p.24).

Both the view of reality (ontology) and the theory of knowledge (epistemology) are shaped by two main poles of philosophical positions in research: *positivism* and *interpretivism*. Within *positivism*, researchers are portrayed as 'scientists' on the quest for objective knowledge (Raddon, 2010, p.6). There is one true, granular and deterministic reality. The scientific method is employed to attain observable and measurable facts that enable law-like generalizations. Causal explanation and prediction is regarded as contribution and as acceptable knowledge (Saunders, 2009, p.136). Within *interpretivism*, on the other hand, researchers can be considered as 'detectives' following the quest for subjective knowledge (Raddon, 2010, p.7). Reality is viewed as complex and rich, socially constructed through culture and language. Therefore, interpretivism offers multiple meanings, interpretations and realities. Hence, the focus and emphasis lies in understanding rather than explaining (Marsh & Furlong, 2002, p.20; Saunders, 2009). New understandings and worldviews are regarded as contribution to knowledge.

Since the research question of the present thesis aims at investigating whether the use of the game element countdown-timer increases the level of engagement within European students, one is prone to view it through the philosophical lens of positivism. Firstly, the research questions does not aim at understanding the 'how' and 'why', i.e. the subjective perspectives, meanings or actions of social actors. Therefore, the application of an interpretivistic research philosophy can be ruled out. Secondly, the researcher rather aims at attaining objective knowledge by measuring a causal relationship between the game element countdown-timer and the engagement rate within European students.

Nevertheless, the researcher acknowledges that the findings of the present study can only "probably be true" (Lincoln et al., 2011, p.170); and never be without any error – which is associated with the research philosophy of *post-positivism,* credited to Karl Popper (1963). Post-positivism criticizes and rejects both the positivist epistemology and ontology, i.e. the objective nature of reality and the ability of science to discern that reality (Fox, 2008). From a post-positivistic, ontological view, "reality is only imperfectly and probabilistically apprehendable" (Lincoln et al., 2011, p.168). The conceptualization of reality of a post-positivistic researcher, therefore, is simply a layered way of seeing that reality and not the only one actual, true reality (Bryman, 2009, p.25). As a post-positivistic researcher, the author tries to minimize errors and bias and as well as to minimize subjectivity as much as possible by creating commensurate structures in the study. From an epistemological point of view, the researcher still aims at investigating causal mechanisms (Saunders, 2009, p.136). This way of knowledge accumulation through causal linkages and

generalizations is common between the positivist and post-positivistic philosophy. The latter also applies to the goodness or quality criteria of research: Both philosophies aim at maximizing the "conventional benchmarks of rigor: external and internal validity, reliability and objectivity" (Lincoln et al., 2011, p.170). The main difference hereby lies in the nature of knowledge, which lies within post-positivism in nonfalsified hypotheses that are probable facts or laws, while positivism establishes verified hypotheses as facts or laws (Lincoln et al., p.170).

All in all, the present gamification research is guided by the epistemological and ontological assumptions of post-positivism. Consequently, effects (e.g. of game elements) exist in the real world and the goal of the researcher lies in discovering, measuring and predicting these effects as much and as objective as possible, thereby getting closer to uncovering the truth. To draw meaningful conclusions, this research must be post-positivist, rather than positivist, "because of its focus on human behavior, which is reactive to science and interventions" (Landers et al., 2018, p.317); and thus the effectiveness of gamification interventions is likely to change over time, unlike in the natural sciences, based on positivism, where phenomena continue to exist regardless of being measured (Landers et al., 2018, p.318).

3.3 Research Approach

Theory development within research can traditionally be distinguished in two approaches: deductive or inductive reasoning (Saunders, 2009, p.144). The general difference between these approaches lies in the way new knowledge is generated and how thinking and action processes are carried across the research. Inductive reasoning refers to moving from the specific to general, while deductive reasoning starts with the general and results in the specific (DePoy & Gitlin, 2016, p.5).

Deductive reasoning implies that the researcher draws on what is known about a particular field and on theoretical frameworks in order to deduce a hypothesis which is then subject to empirical tests (Bryman, 2016, p.21). Consequently, the deductive "top-down" approach is primarily used by experimental-type researchers (DePloy & Gitlin, 2016, p.5). Inductive reasoning, on the other hand, is employed by researchers for naturalistic inquiries. There is no "truth" or general principle that is accepted "a priori", rather the researcher seeks general rules or patterns emerging from specific observations (DePloy & Gitlin, 2016, p.5). Thus, applying inductive reasoning aims at generating theories or hypotheses by primarily using qualitative research (Thyer, 2009, p.34). In addition to the two research traditions of inductive and

deductive reasoning, the approach of abductive thinking needs to be introduced. Abduction refers to an iterative process in naturalistic inquiry, where the data is analyzed for its own patterns and concepts in order to generate a new or modify an existing theory (DePloy & Gitlin, 2016, p.6; Saunders, 2009, p.145). Induction attempts to fit the data to a theoretical framework or generate concepts that emerge from the data while in deductive reasoning the data is "contained and controlled by the hypothesis" (DePloy & Gitlin, 2016, p.6).

All in all, the objective of the study determines which research approach is suitable. In the present thesis, the research emphasis lies in testing the effectiveness of gamification. Theory on gamification and its effectiveness does exist and has been evaluated; however, narrowing it down, the effectiveness of gamification in increasing engagement within European students by incorporating the game element countdown-timer has not been tested yet. Therefore, operationalization leads to the development of the concrete hypotheses assuming the before mentioned cause-effect relationship which then is tested, aiming at generalizing the results and thereby building new knowledge (Saunders, 2009, p.146). Consequently, the deductive reasoning approach was chosen for the present thesis.

3.4 Research Strategy

Research traditionally discerns three broad strategies that guide the collection, measurement and analysis of data with the goal to answers research questions and control variance (Aaaker et al., 2008; Phillips, 1966; Kerlinger, 1973; Wyk, 2012). These three strategies are exploratory, descriptive or explanatory (causal) in nature (Wyk, 2012).

Exploratory research is the most useful (and appropriate) for studies that are investigating a subject about which there are high levels of uncertainty and ignorance about the subject, and when the problem is not very well understood, i.e. very little existing research on the subject matter (Wyk, 2012, p.8). Consequently, exploratory research can answer questions of all types (what, why, how) and is usually characterized by a high degree of variance and flexibility (Saunders et al., 2012; Wyk, 2012, p.11). On the other hand, descriptive research aims at defining and accurately describing a phenomenon and is more structured than exploratory research while having no control over the variable (Blanche et al., 2006; Wyk, 2012). Descriptive research, therefore, is directed at answering the "what"-question and enables the subject to be viewed more completely than was possible without following this research strategy

(Ethridge, 2004; Fox & Bayat, 2007). Lastly, explanatory research attempts to determine the causal relationship between variables and is very structured in nature (Blanche et al., 2006; Wyk, 2012; Zikmund et al., 2012). Explanatory research, therefore, asks causal questions such as "why" and "how" and holds a high degree of control over the variable and other confounding factors, making it more complex to run. Experiments are the most common main data collection method in causal research strategies (Zikmund et al., 2012).

Research as a "purposeful, goal-directed activity" (DePloy & Gitlin, 2016, p.53) is conducted for a reason. Therefore, the purpose of the study drives the selection of the research strategy. As the literature research revealed, gamification research lacks a coherent research practice that leads to interpretable and generalizable research findings (Koivisto & Hamari, 2019; Landers et al., 2018), since most of the empirical research on the topic was conducted without control groups and without identical or validated measurement instruments for particular variables that are likely to account for the effectiveness of gamification (Hamari et al., 2014). The purpose of the present thesis lies in addressing the described lack within research and is directed at investigating and measuring the cause-effect relationship between a single game element, i.e. that of a countdown-timer, and user engagement. Consequently, the selected research strategy for the present thesis is of *explanatory (causal)* nature shaped by the deductive post-positivist philosophy. Nevertheless, the present thesis additionally shines light on a lack (or ignorance) within gamification research on rigorous empirical testing of its effectiveness and can be considered as exploratory research in that regard, too.

In accordance with the explanatory, causal research strategy and the present underlying post-positivistic paradigms of objectivity and control, the *experiment in its true form* is chosen as data collection method as all three prerequisites are fulfilled: Firstly, sufficient theory about a phenomenon exists in order to examine causality; secondly, the research question implies the intent of predicting and revealing a cause; and thirdly, the conditions permit the fulfillment of randomization, control group and manipulation (DePloy & Gitlin, 2016, p.140). The true-experimental design presents the most "objective" scientific approach and offers the greatest degree of control and internal validity (DePloy & Gitlin, 2016, p.138), which is the logical choice from an ontological and epistemological post-positivist lens.

Lastly, choosing the accurate research strategy in regard to the research question also entails orientation towards either a quantitative or qualitative strategy in the collection and analysis of data (Bryman, 2011,

p.32). Qualitative research, on the one side, emphasizes an inductive approach to the theory of knowledge, i.e. through the generation of theories. It entails an interpretative epistemology and constructivist, subjective ontology (Bryman, 2011, p.32). On the other hand, quantitative research emphasizes quantification in the collection and analysis of data which is linked to a deductive approach and associated with the test of theories. Furthermore, it has epistemologically incorporated the practices and norms of positivism and post-positivism as well as embodies a view of reality that is external to the actor and objective in nature (Bryman, 2011, p.33). Based on the described premises, the present thesis clearly entails the *application of a quantitative research strategy*.

All in all, the present thesis follows an explanatory, causal research strategy where quantitative data is collected through setting up a true experiment, based on a deductive research approach which is shaped by the post-positivist assumptions of the author.

3.5 Validity, Reliability and Generalizability

The aim of any experiment lies in producing results, i.e. causal relationships that are valid, reliable and generalizable to populations and contexts not included in the experiment (Bracht & Glass, 1968), by maximizing both internal and external validity (Barabas & Jerit, 2010). While *internal validity* is defined as "the validity with which statements can be made about whether there is a causal relationship from one variable to another" (Cook & Campbell, 1979, p.38), *external validity* describes the "extent and manner in which the results of an experiment can be generalized to different subjects, settings, experimenters and, possibly, tests" (Bracht & Glass, 1968, p.438). In simple terms, validity translates into measuring "what [one] thinks [one] is measuring" (Field & Hole, 2002, p.44). Therefore, in the present thesis, measuring engagement of the participants has been conducted as objective as possible and with as few systematic and random errors as possible. Moreover, reliability is the ability to replicate the same results under the same conditions (Field & Hole, 2002, p.47). Hereby, validity is a prerequisite of reliability – in order to be reliable a measure first needs to be valid. In order to achieve reliability in the present experiment, the researcher made sure to measure the dependent variable, i.e. the engagement rate, as precise as possible (Field & Hole, 2002, p.57). Generalizability is closely related to external validity and presents the ability to apply the results of a study to a larger population and to different circumstances than what the participants in the original study experienced. The best measure of generality is by empirical testing such as through replications of the experiment on other populations and other contexts (Field & Hole, 2002,

p.63). According to Landers et al. (2018), two of the key goals within gamification research that lead to a higher quality conclusions are "experimental design (i.e., internal validity), and generalizability (i.e., external validity)" (Landers et al., 2018, p.329). All in all, extensive efforts have been put into designing the experiment such that it maximizes both internal and external validity, and thus generates results that are valid, reliable and generalizable. These efforts in designing and running the true experiment are explained in the following chapter.

3.6 Research Method

The following 'Research Method' chapter provides the reader with a clear idea of what method the experimenter used and how he conducted the study. It aims at offering the possibility for future researchers to replicate the experiment based on the information provided in this chapter (Field & Hole, 2002, p.320).

3.6.1 Experimental Design

This section provides an overview of the formal design of the experiment. The experiment used a between-groups (or 'independent-measures') post-test only design where randomly allocated, separate groups of participants performed each in one of the different conditions only (as illustrated in Figure 2; Field & Hole, 2002, p.70). The post-test only experiment design was selected being most valuable when pretesting is not possible or not appropriate, which is the case since the experience of gamification changes a population which then becomes the new truth for this population (Landers et al., 2018, p.319; Thom et al., 2012).

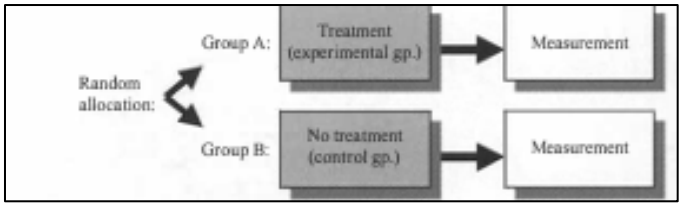


Figure 2: The 'post-test only/control group' design (Field & Hole, 2002, p.71)

Participants were asked how they perceive their current situation during the global virus-crisis as the context for the experiment. Both conditions were identically structured with the sole difference lying in

the fact that the treatment condition was exposed to the countdown-timer while the control group did not experience the countdown effect.

The participant's experiences within the given condition were comparable as the appearance of the landing page was designed to be identical on desktop, tablets and smartphones. In addition to that, the participant's experiences in both conditions were made as identical as possible by integrating a box on the landing page of the control group at the exact same measurements as the countdown-timer on the treatment-condition landing page (as shown in Figure 3), thus increasing the validity of the experiment. (Field & Hole, 2002, p.38)

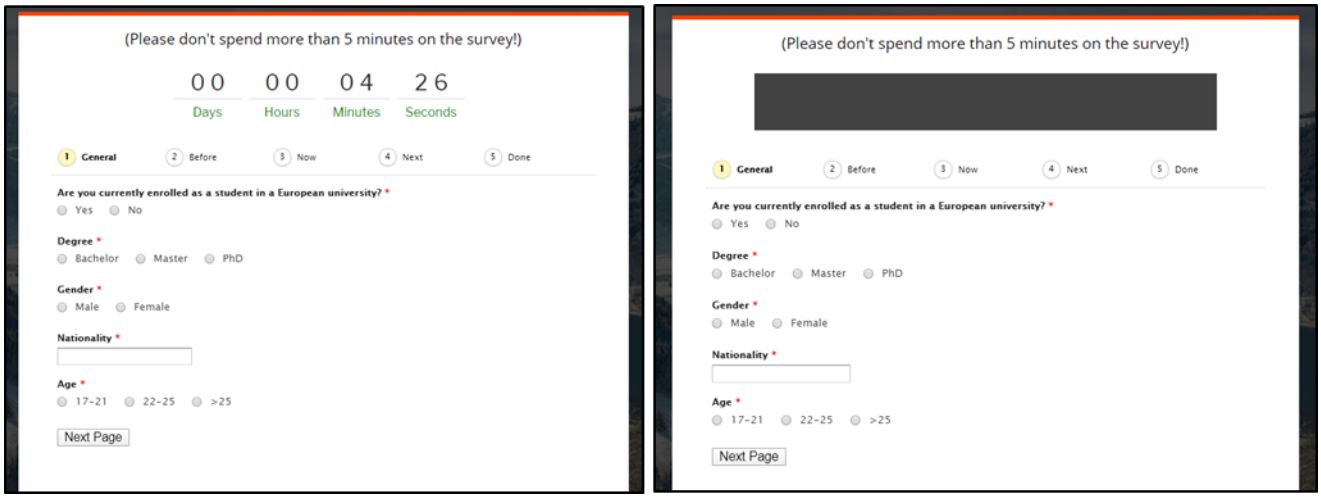


Figure 3: Identical look of experiment conditions except countdown-timer or black box respectively

There was one independent variable used, i.e. the game element countdown-timer (as shown in Figure 4) which was incorporated into the landing page. In the experiment, the countdown-timer starts counting down from five minutes after the participant presses the "start" button on the landing page (Figure 10) and gets redirected to the questionnaire page.

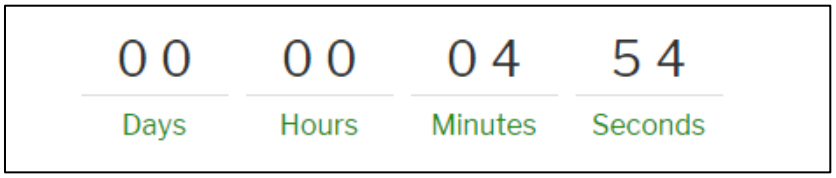


Figure 4: Independent variable: game element countdown-timer used on landing page

The primary outcome of interest, i.e. the dependent variable, was the level of engagement in the treatment and control group whose operationalization will be explained more specifically in chapter 1.5. Generally, in a study with a between-groups design, it is essential that participants are allocated randomly to the treatment or conditions (Field & Hole, 2002, p.74); the practice will be explained in the next subchapter “participants”. By rigorously and properly carrying out randomization of participants, a fairly unambiguous identification of cause and effect is enabled. Therefore, the present gamification researcher made sure as much as possible that the only systematic effect on participants' behavior is the researcher's manipulations of the independent variable (Field & Hole, 2002, p.71).

The advantages of the chosen between-groups design are its simplicity and the elimination of practice and fatigue effects. Setting up a between-groups experiment is simple and takes less time than a within-participants design. In addition to that, participation time is shorter for the participants, making recruitment easier for the experimenter. By making sure to allocate participants randomly to the different conditions, procedures such as counterbalancing are not necessary (Field & Hole, 2002, p.75). Furthermore, there is no possibility that participation in one experiment can affect the participation in another experiment in an irreversible and meaningful way, as each participant takes part in only one condition.

The disadvantages associated with the chosen between-groups design, however, lie in its expensive form and its insensitivity to experimental manipulations (Field & Hole, p.2002, p.77). First of all, between-groups experiments are linked to high expenses in terms of time, effort and participant numbers. In the present experiment, recruiting participants and running the experiment stretched over weeks and months; and constantly being required to acquire new participants is time-consuming and laborious while time simultaneously needs to be put into analyzing the data and writing the thesis, too. As a consequence, every additional variable doubles the number of participants necessary to test – assuming there are only two levels of each variable (Field & Hole, 2002, p.97). In addition to that, according to Field and Hole, between-groups designs are associated with an increased insensitivity to experimental manipulations in comparison to a within-participants design. It is therefore less likely to detect an effect of the experimental manipulations. Increasing the likelihood of finding significance in the effect, consequently, is linked to increasing the investment in recruiting participants and extending the time of running the

experiment, besides rigorously allocating participants randomly to the conditions (Field & Hole, 2002, p.77).

All in all, the between-groups experiment design was selected by the researcher to keep it as simple as possible and to run a well-designed and well-executed experiment with the goal to attain significance in detecting differences between the two conditions, rather than attempting a 'grandiose' experiment containing various conditions.

3.6.2 Participants

Overall, more than 884 individuals participated in the experiment. The demographic distribution between the two conditions will be illustrated more in detail in the results-chapter.

All participants voluntarily took part in the experiment and were not paid for their participation. They were recruited over LinkedIn through identical direct messages that were personalized with the potential participants' first name. Over a time span of 30 days, more than 3000 invitations and direct messages were sent out to students in the LinkedIn network. About 884 LinkedIn users accepted the invitation to take part in the experiment (29.5%). In order to be among the invited LinkedIn student profiles, users had to fulfill the following search criteria:

- Region: *Europe* (Figure 5)
- Title: *Bachelor-Student, Master-Student, Student-Assistant, IT-Student, PhD-Student, Student, Intern, Research-Assistant* (Figure 6)
- Work Experience: *Less than one year, between 1-2 years, between 3-5 years* (Figure 7)

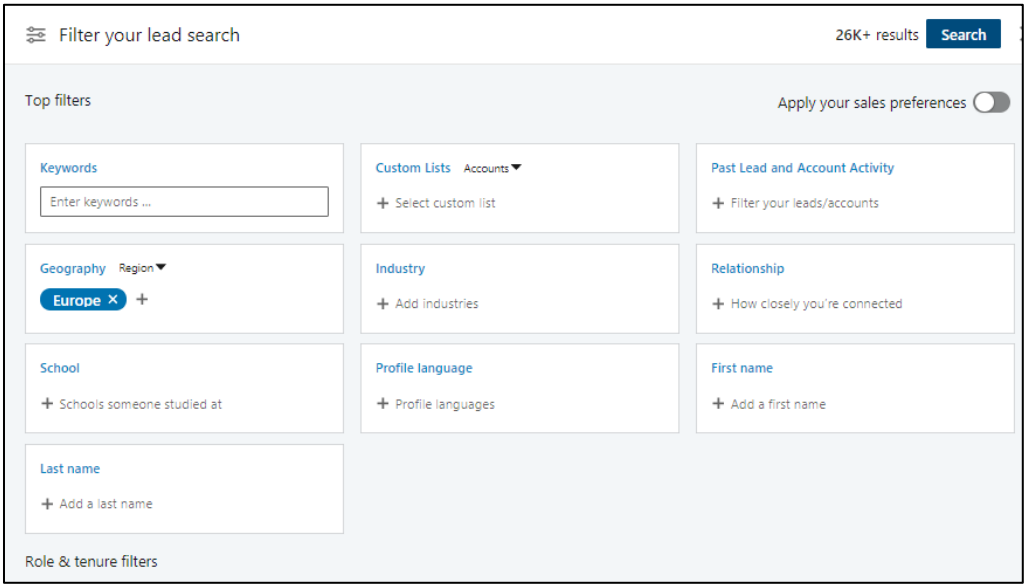


Figure 5: LinkedIn-Search Filter 1/3 - Region Europe

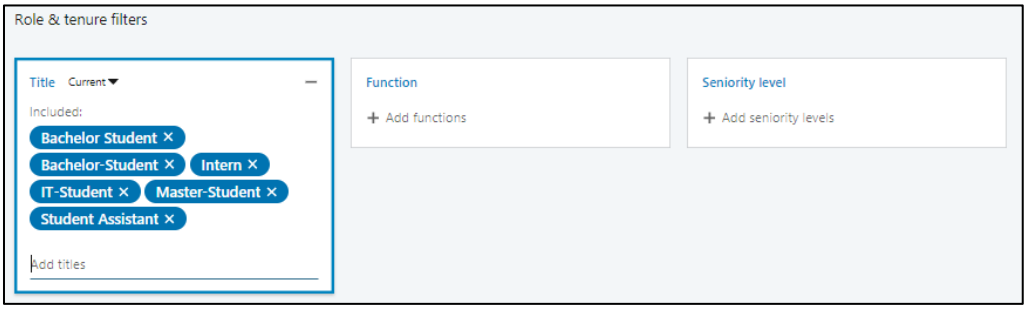


Figure 6: LinkedIn-Search Filter 2/3 - Possible Titles

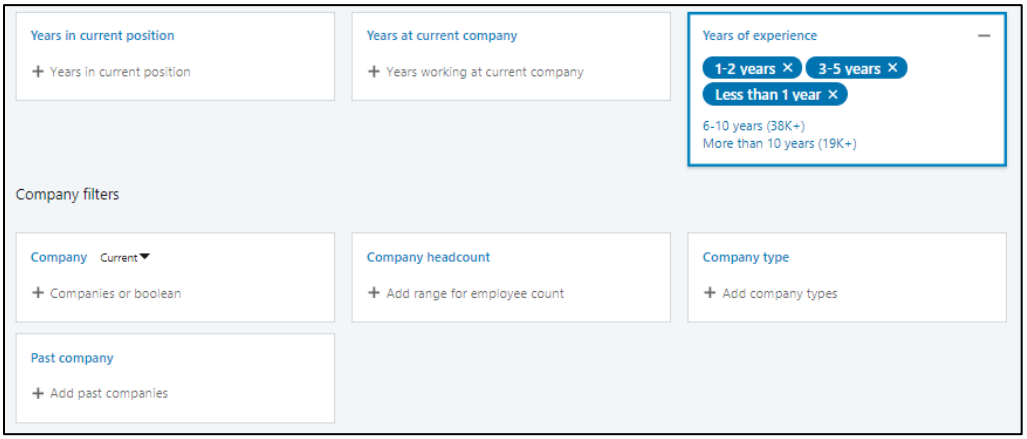


Figure 7: LinkedIn-Search Filter 3/3 – Years of work experience

All in all, LinkedIn indicated that more than 26.000 users match the criteria described above. Consequently, almost 15% of the overall reachable population for the researcher has been approached and invited to the experiment. LinkedIn only indicates 2nd and 3rd degree contacts of the searching person. 2nd degree contacts are connections of a user's contact while 3rd degree contacts are connections of a 2nd degree contact. Thus, 26.000 total users are the total reachable population for the researcher personally (Figure 8), not the overall number of users that match to the set criteria.

Figure 8: Total accessible target population on LinkedIn

Before recruiting participants, the search results of more than 26.000 profiles were saved in a new list. In the recruitment phase, participants were allotted randomly to the experiment conditions in order to ensure not producing bias in the results. In practice, 100 LinkedIn profiles within the target population were randomly contacted every day. It was carried out by sending invitations with a link to the treatment condition to all profiles on even search page numbers (50 users) ; and, consequently, with a link to the control condition to all profiles on uneven search page numbers (50 users as well).. This alternating manner was executed continuously on more than 30 days, meaning that all participants were invited to both conditions of the experiment on every weekday, instead of having the treatment group be tested on weekdays and the control group on weekend days which could potentially skew the results.

The standardized invitation message sent out to potential participants within the reachable population is shown in Figure 9 below. The only difference between the treatment group and control group invitation lied in the ending of the website link.

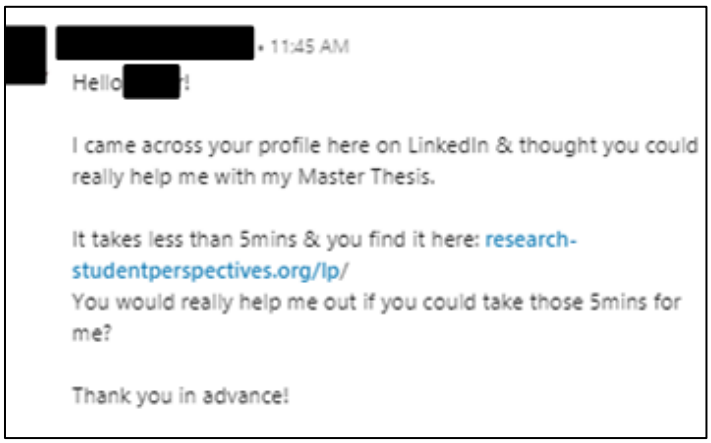


Figure 9: Standardized Invitation message for the reachable population on LinkedIn

Since only about one third of participation invitations got accepted, this randomized approach could potentially lead to systematic differences. Nevertheless, over the course of 30 days following the described randomization approach, the probability of unsystematic differences such as motivation, anxiety, irritability or receptivity to time-pressure are be diminished on average (Field & Hole, 2002, p.72). By rigorously and properly carrying out randomization of participants, a fairly unambiguous identification of cause and effect is enabled. Therefore, the present gamification researcher made sure as much as possible that the only systematic effect on participants' behavior is the researcher's manipulations of the independent variable (Field & Hole, 2002, p.71).

3.6.3 Ethical considerations

Generally, in the process of recruiting participants and conducting the actual experiment, ethical considerations were taking into account such as informed consent, debriefing, confidentiality as well as psychological protection (Field & Hole, 2002, p.98). When contacting potential participants, the latter had the possibility to voluntarily participate in the experiment. Furthermore, the contact messages informed the potential participants about the general purpose of the study (‘master-thesis’) and the topic (‘researching student perspectives’). The researcher considered informing potential participants on the specific subject, i.e. the influence of the game element countdown-timer on their behavior; however, taking into account the potential bias caused by participants being aware of the subject of study, also known as the ‘hawthorne effect’ (Campbell, 1957, p.308), the researcher decided to leave this information out. During the experiment, participants were able at all times to leave the experiment by changing the

website or closing their internet-browser. After participating in the experiment, participants obtained answers to all questions on the subject of study. The questionnaire was answered anonymously and every participant received a personal 'thank you'-message as gesture of appreciation.

In regard to personal data, the researcher was not interested in the participants' individual data; rather in the combination with those from other participants which protected the right to confidentiality of every participant at all times (Field & Hole, 2002, p.100). Lastly, the experiment was set up in a way that does not harm participants both physically and psychologically. The questionnaire was designed in a way that "avoided making participants feel stressed, embarrassed, depressed, anxious or fearful" (Field & Hole, 2002, p.101).

3.6.4 Apparatus

The '*Apparatus*' section gives details of how the experiment was set up in order to enable future researchers to replicate it (Field & Hole, 2002, p.322). For the experiment, two landing pages were designed and set up with the content management system Wordpress and its plug-in ThriveArchitect, with one landing page leading to the website that incorporated the countdown-timer while the other landing page lead to the website that contained a black box as placeholder instrument. The website domain "www.research-studentperspectives.org" was acquired at web hosting provider STRATO for the experiment to provide a more professional appearance and increase the overall response rate.

The participants' behavior was measured through a questionnaire that was designed with the service of Wufoo.com and embedded in the Wordpress landing page. Single-time participation in the experiment was made sure by counting individual IP-addresses only once. Measuring the level of engagement in the treatment and control group as dependent variable was chosen as "precise, unambiguous and objective [measurement]" (Field & Hole, 2002, p.57). With the described measurement in place, the researcher was able to directly and precisely count the number of times a specific behavior occurred which in turn leads to a higher reliability of the study (Field & Hole, 2002, p.44). For the completion of the questionnaire the only requirement necessary was internet access on computers, laptops, tablets or smartphones.

The questions in the questionnaires in the form of checkboxes, multiple choices and Likert-scales were employed as they are easily understood and give the participants more scope to express how they feel about something. Nevertheless, for the present thesis they were not relevant and were identical in both conditions of the respective experiment; rather these elements were built in to engage the participants

with the content and measure their behavior under the presence or absence of countdown-timers. Consequently, benefits and disadvantages, including content, criterion and factorial validity of different measurement elements were not at the scope of the experiment and will not be elaborated in the present thesis. (Field & Hole, 2002)

3.6.5 Procedure

The procedure section aims at providing details of how the experiment has been carried out and run in practice. As described before, participants were invited to the experiment through a personal direct message containing a link to the landing page. Different links were sent out to the treatment and control group respectively for the purpose of measuring and aggregating the data separately. Once arrived on the identical landing page, users were informed that the researcher was investigating the perspectives of European students in relation to their ideal future (illustrated in Figure 10). As explained in the "ethical considerations" section, the participants were not informed that the true subject to be investigated was the game element countdown-timer since it potentially would have caused skewed results. Moreover, visible in Figure 10, it was indicated on top of the landing page that the experiment should take less than 5 minutes. After reading through the introduction and description, participants were able to start the experiment by clicking on a blue button that contained the following call to action (as shown in Figure 10).

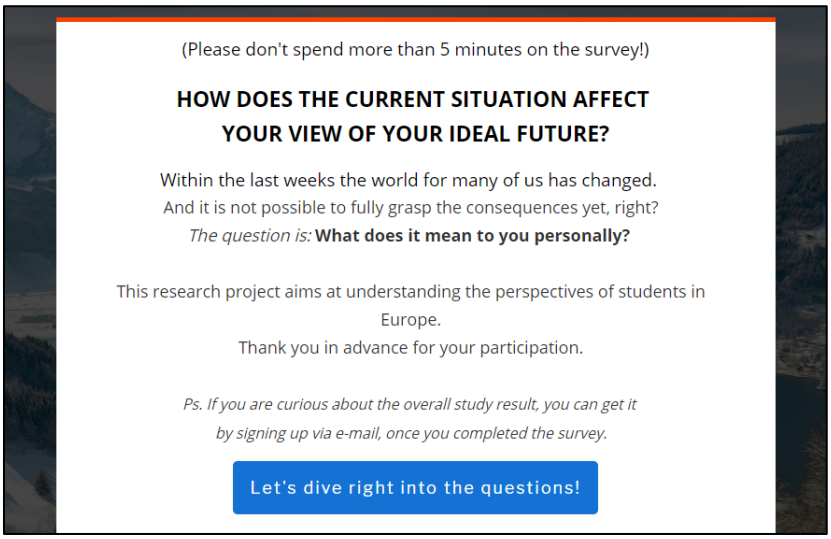


Figure 10: Identical landing page for both treatment and control condition

In the treatment group condition a countdown-timer (as visualized in Figure 4) appeared and directly started to count down from 5 minutes. The participants of the control group, however, faced a black box of the same size as the countdown-timer located at the exact same position (as shown in Figure 3). Both groups were asked to take the exact same questionnaire, leading both groups through 5 answer pages where all fields were required to be filled out to proceed to the next page (as visible below in Figure 11).

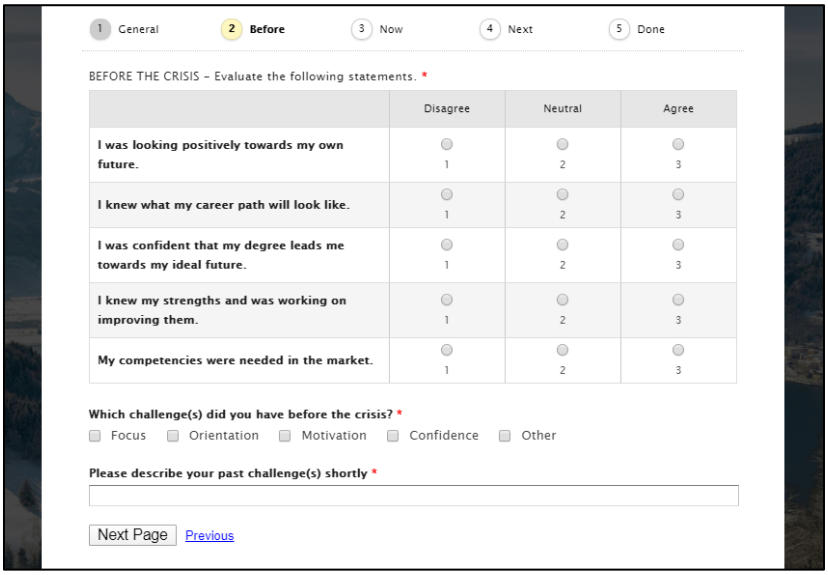


Figure 11: Questionnaire elements during the experiment

After having answered all the questions, participants received a "thank you-message" on the website for their participation. The questionnaire was tested before rolling out the actual experiment among a small sample of 30 participants within the target population to investigate the user experience, the ease of navigation through the questions as well as measuring the completion time. The test revealed highly positive answers among all relevant dimensions with an average completion time of 4:07mins. Consequently, 5 minutes were chosen for the timer to count down to for the actual experiment, in order to achieve a both challenging and realistic range of time for the participants. For participants that were not able to complete the survey within the given time, the experiment stopped and they received the identical "thank you-message" for participating.

3.6.6 Operationalization of engagement as dependent variable

Generally, the engagement of participants is the dependent variable of the present experiment. This outcome of interest, however, needs to be defined in a "precise, unambiguous and objective [manner]" (Field & Hole, 2002, p.57). Without unambiguous, precise definition, research does not allow for reliable and replicable results by future research. The present operationalization of engagement, therefore, defines the exact measuring method used in the experiment. Consequently it allows future researchers to follow exactly the same experimental procedure. All in all, the operationalization described in this subchapter leads to a higher robustness and reliability of the experiment.

In the present experiment, the engagement of participants was measured on a scale from 1 to 6; with 1 being the lowest and 6 the highest form of engagement (Figure 12 below).

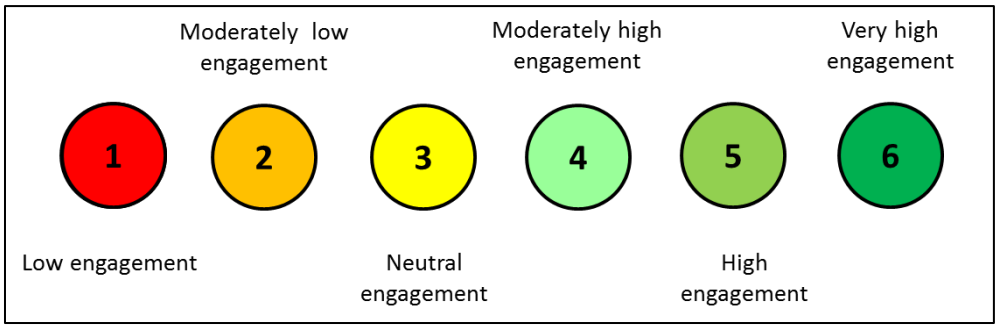


Figure 12: Operationalized scale of engagement of Participants

The operationalized scale of engagement can be decoded as follows. Once a participant decided to start the experiment, the questionnaire in both the experimental conditions takes the participants through 5 pages in total. Thereby, each completed page indicates the respective level of engagement. To be more specific, very low engaged participants would leave the experiment after the first page of the questionnaire while moderately high engaged participants would complete the survey, i.e. participate on all five pages. Since all fields in the respective pages were required to be selected or filled out in order to get to the next page, this method of measurement can be considered as very accurate. The highest level of engagement, level 6, was assigned only to participants that completed the questionnaire and entered their e-mail address in the signup field on the last page as invitation to receive a summary of the study results (shown in Figure 13). Overall, it was the only field in the questionnaire that was not required to be filled out. The action of providing a valid e-mail-address (checked by the survey tool), due to its personal

nature and obvious signal of interest, is hereby considered as strong indication of engagement and therefore as additional point to completing the respective pages of the questionnaire.

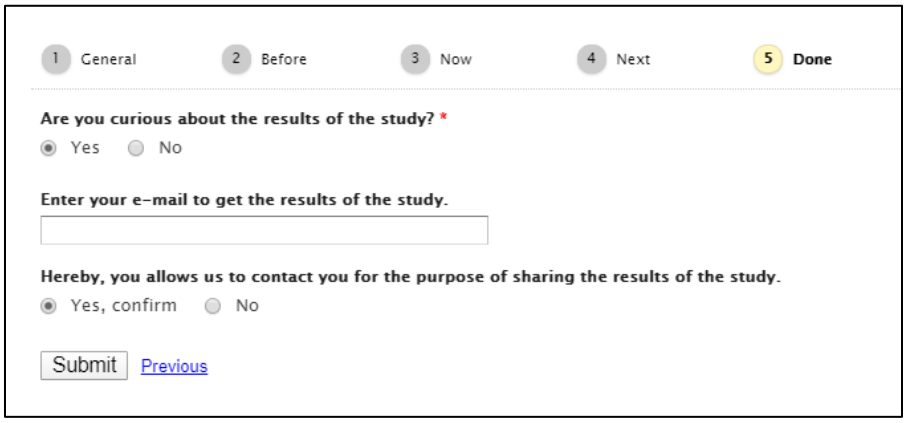


Figure 13: Important part of measuring the dependent variable: e-mail signups at the end of the experiment

Operationalizing the dependent variable of engagement in a more granular fashion enables the researcher to measure engagement more precisely compared to simply comparing the completing rates and e-mail sign-up numbers between the treatment and control group of the experiment. Being able to directly count the number of completed pages as well as the e-mail signups and its operationalization to different levels of engagement leads to a higher reliability of the study (Field & Hole, 2002, p.57). Furthermore, it enables future researchers to follow and replicate the current experiment in more robust and reliable manner. After collecting the data and assigning the respective levels of engagement, it will be analyzed in the results-chapter.

3.7 Internal and external validity of the experiment

The key to internal validity is good measurement and study design, and representative sampling (Bernhard & Bernhard, 2013). Within gamification research, internal validity describes the degree to which the predictor variable, i.e. the game element, was causing a change in the attitude or behavior of the user (Hamari et al., 2014; Landers et al., 2018). In the past, most empirical studies within gamification did not implement one game element only, but multiple game elements, which commonly led to a lower degree of internal validity since it gets impossible to deduce which game element or combination of game elements were responsible for changes in the outcome (Landers et al., 2018). As a consequence, the present experiment uses only one game element and isolates it to increase the internal validity of the

experiment. As a result of this experimental design, differences in the outcome of interest to a high degree can be causally attributed to the addition of the countdown-timer. The influence of extraneous variables on the outcome of interest, however, cannot be fully excluded and will be reflected in the following subchapter. As Finkel et al. (2017) state "no single study can accomplish everything […] and that there will always be alternative explanations for the effectiveness of a manipulation" (Finkel et al., 2017, p.2). It means that there will always be doubts about the internal validity of any experiment. Furthermore, resources are finite, and there is always a tradeoff between either bolstering internal or external validity (Brewer & Crano, 2014; Cook & Campbell, 1979; Finkel et al., 2017).

External validity, on the other hand, presents the degree to which it is possible to apply the results of the experiment to a larger population and to different circumstances than what the participants in the original study experienced (Bracht & Glass, 1968). The interactive effects of extraneous variables and of experimental arrangements affect the external validity or generalizability of experimental results (Campbell, 1957, p. 313). Reaching a high degree of external validity, therefore, is more complex and associated with a large investment of resources time wise and financially (Field & Hole, 2002, p.58). The potential threats to external validity of the present experiment are assessed after first evaluating the threats to internal validity in the following.

3.7.1 Threats to internal validity

It is important to seek control of the experiment by ruling out extraneous variables. Therefore, Campbell (1957) distinguished between seven extraneous variables which experimental designs in social science seek to control for: "history, maturation, testing, instrument decay, regression, selection, and mortality" (Campbell, 1957, p.311). In the following, the influences of these seven extraneous variables are investigated in order to assess the degree to which they might pose a threat to the internal validity of the present experiment.

History is defined as a threat to internal validity when other factors external to the participants in the study occur during the experiment which affects the experimental groups differently (Bernhard & Bernhard, 2013; Campbell, 1957; Field & Hole, 2002; Slack & Draugalis, 2001). As the experiment took place in the internet, distributed to participants throughout Europe, it was not possible to control external factors influencing the two groups differently. Nevertheless, by randomly assigning participants to either

groups, and through generating a representative sample, the influence of these factors can be minimized. *Maturation* refers to the fact that users become more experienced while the experiment is conducted as well as that individuals change biologically and psychologically over time (Bernhard & Bernhard, 2013; Field & Hole, 2002; Slack & Draugalis, 2001). Generally, these threats can be ruled out as both *history* and *maturation* are more of a concern for longitudinal studies (Slack & Draugalis, 2001).

Threats to internal validity due to *testing* occur when participants get tested repeatedly for results in dependent variables which may then change their responses naturally (Bernhard & Bernhard, 2013; Campbell, 1957; Slack & Draugalis, 2001). In the present case, the testing threat scenario would be manifested in participants being exposed to the same game element twice or to similar content twice. As a consequence, the participants' view on the time-countdown or content might be biased and part of the changes measured in the dependent variables could then be the results of maturation of users, which is a realistic scenario for gamification (Landers et al., 2018, p.319; Thom et al., 2012). Consequently, within the present research design, each participant is tested once only and for one of the conditions only, thus excluding the testing effect (Bernhard & Bernhard, 2013; Campbell, 1957; Cook & Campbell, 1979).

The *instrumentation* threat results from changes in instrument measurement or observer changes, when differences in the results might not be due to true treatment effect (Bernhard & Bernhard, 2013; Campbell, 1957; Slack & Draugalis, 2001). In the present study design, measurement criteria are precise and consistent throughout the two conditions of the experiments. There is no change in measurement and all the measured data is obtained objectively. Thus, the instrumentation threat to internal validity is not assessed as critical.

Regression to the mean is a threat to validity occurring when researchers conduct experiments on "groups that have extreme scores on a dependent variable" (Bernard & Bernard, 2013, p.96; Field & Hole, 2002). In order to prevent the outcome of a study to be confounded, it is necessary to control the groups for a high degree of stability within the population (Campbell, 1957), which has been achieved in the present experiment based on a large sample size and randomly assigned participants.

The next threat to internal validity lies in the *selection of participants*. If participants differ from each other to a large extent, it can skew the resulting outcomes. The selection threat is a critical concern if participants cannot be randomly assigned to treatment or control groups (Slack & Draugalis, 2001). It is necessary to control the groups being as equal as possible to relevant confounding variables, such as age

or gender (Slack & Draugalis, 2001). Therefore, participants were randomly assignment to the treatment or control group and potential moderators such as age and gender were measured, too. Thus, due to the large sample any differences among individuals in the population of both groups should be distributed equally. It is not possible to fully eliminate the possibility of selection bias (Bernard & Bernard, 2013); nevertheless, the present study design tries to reduce it to a minimum.

Lastly, a threat to internal validity can occur when a "biased subset of [sample] members have dropped out" (Campbell, 1957, p.300) which results in unequal study groups at the end of the experiment. It is known as *experimental mortality* in the form of withdrawals, dropouts or attrition (Field & Hole, 2002; Jurs & Glass, 1971; Slack & Draugalis, 2001). The effect of experimental mortality on the present study can be regarded as insignificant as the experiment does not last for a long time and the treatment condition is not particularly demanding (Jurs & Glass, 1971, p.65). Moreover, dropout proportions are not assumed to differ between treatment and control group.

All in all, there are various extraneous factors that can lead to changes in behavior, changes that can be confused with the effects of the intended manipulations, thus they present threats to the internal validity of the experiment (Field & Hole, 2002). Good experimental designs, therefore, guard against (control for) all of these competing explanations for the changes in the dependent variable, such that the "changes are a direct consequence of the experimental manipulations" (Field & Hole, 2002, p.62).

3.7.2 Threats to external validity

The threats to external validity fall in two broad classes: *population validity* and *ecological validity* (Bracht & Glass, 1968). *Population validity* deals with generalizations to populations of participants, i.e. whether it can be expected that the latter behave in the same way as did the sample experimental participants (Bracht & Glass, 1968). Populations can be distinguished between the experimentally accessible population and the target population (Kempthorne, 1961). The former group is available to the researcher. The target population is defined as "the total group of [participants] about whom the experimenter is empirically attempting to learn something" (Bracht & Glass, p.440). Applied to the present experiment, the researcher attempts to empirically better understand European students' behavior exposed to gamification and draw conclusions upon these findings. The experimentally accessible population circumscribes all European students randomly assigned to either the treatment or control group. The more representative the sample is, the higher can be the confidence in generalizing from the

sample to the target population. Randomized selection, therefore, is the foundation for generalizing the findings to a larger group of subjects (Bracht & Glass, 1968; Kempthorne, 1961).

Furthermore, the larger the sample the more likely is the research to achieve statistical significance which is the foundation for generalization (Field & Hole, 2002, p.154). Lastly, the over-use of participant groups can pose a threat to the external validity of the current experiment. In the present case, only volunteers participated whose behavior could potentially be different compared to the behavior of non-volunteers.

Ecological validity, on the contrary, refers to the generalization of obtaining the same effects of a study under other environmental conditions. Thus, the experimental effect is assumed to be independent of the experimental environment (Bracht & Glass, 1968). Generalizing the effects of countdown-timers on a landing page to offline situations would be associated with indeterminate risks, as these "real" situations are not similar to the experimental setting (Bracht & Glass, 1968). On the other side, however, other experimenters can replicate the experiment to a high degree on landing pages or websites in the internet. For this reason, the independent variable and the setup of the experiment are described in sufficient detail in the present methodology section; and, as a consequence, the scientific value of the experiment is increased (Bracht & Glass, 1968).

The next threat of ecological validity lies in the multiple-treatment effect, which arises when same subjects experience two or more treatments consecutively and, as a consequence, it becomes difficult to "ascertain the cause of the experimental results or to generalize the results to settings in which only one treatment is present" (Bracht & Glass, 1968, p.439). In the present experiment, the multiple treatment effect can be excluded as subjects will only participate in one experiment consisting of one treatment only. Additionally, the threat of pretest and posttest sensitization to the content of the treatment does not apply to the current experimental design. The threat of the *experimenter effect*, however, is relevant to the present experiment. More precisely, the behavior of participants might be "unintentionally influenced by certain characteristics or behaviors of the experimenter" (Bracht & Glass, 1968, p.439). Applied to the present experiment, the appearance of the LinkedIn profile as well as the style and nature of the direct message could influence the participants' behavior.

Furthermore, the expectations of the experimenter could also bias the administration of the treatment and the observation of the participants' behavior, however, this threat to external validity has been reduced by the precise and objective measurement criteria of the dependent variable (Bracht & Glass,

1968, p.439). Moreover, measurement of the dependent variable at two different times can produce different results (Bracht & Glass, 1968). In order to minimize this threat, experiment invitations were sent out at around the same time of the day for 30 days consecutively for both control and treatment group. The experimenter effect can be linked to the *'hawthorne' effect* which presents the threat of participants' behavior being influenced as to how one should or is expected to behave in the experiment (Bracht & Glass, 1968, p.439). The 'hawthorne' effect cannot be excluded in the present experimental setting; however, based on the fact that participants take part in the experiment anonymously and the 'true' subject of investigation is not known to them, it is assumed that the influence of the 'hawthorne' effect is minimal.

Good experimental designs, therefore, put effort in controlling for all of these threats to external validity, as shown within the present gamification experiment, in order to attain results that are generalizable to a larger population and to different circumstances than what the participants in the original experiment experienced (Bracht & Glass, 1968).

4. RESULTS

This section presents the results and findings of the experiment. After providing information about the pretreatment applied to the data, descriptive statistics will give the reader a clear and succinct summary of the data that was collected in the experiment. Finally, inferential statistics will present the reader with the results on whether there are statistically significant differences between the groups and conditions of the experiment.

4.1 Pretreatment of data

All in all, the experiment counted a total of 884 data entries or participants. However, not all entries provided meaningful data to be included in the analysis. Therefore, a count of 92 entries (10.40%) in total was eliminated and excluded from the analysis. The two reasons for exclusion were either double entries, probably because of technical difficulties and measured through the IP-address, or participants that indicated not being enrolled European students at the time the experiment took place. Data entries of participants that did not provide a complete set of data, i.e. did not finish the questionnaire, were counted based on their relevance for measuring the engagement of participants as illustrated before in the 'operationalization' chapter.

4.2 Descriptive Statistics

In total, data entries of 792 participants were meaningful and consequently included in the analysis. Both conditions of the experiment had an almost identical sample size: The treatment group counted 394 participants while 398 people participated in the control condition (Figure 14).

Figure 14: Sample sizes of treatment and control group of the experiment

As visualized in Figure 15, 255 (64.72%) participants were male and 139 (35.28%) female in the treatment condition, while 231 (58.04%) participants identified as male and 167(41.96%) as female in the control

group. Male participants are slightly overrepresented in the sample which is based on the fact that the majority of LinkedIn users are men and the random allocation reflects on that.

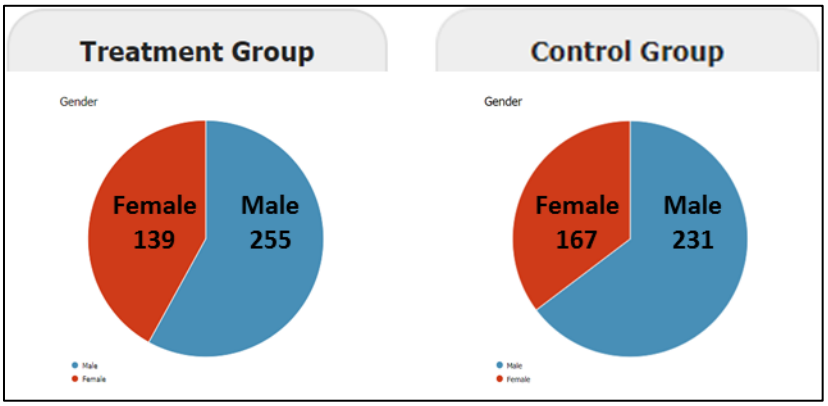


Figure 15: Gender distribution in treatment and control group

In terms of study degree, the treatment group reveals about 325 (82.49%) of the participants pursuing a master degree, compared to 65 (16.50%) bachelor and 4 (1.02%) PhD students. As Figure 16 shows, the control group had a similar distribution of about 305 (76.63%) master students, while 89 (22.36%) were enrolled as bachelor and 4 (1.01%) as PhD students.

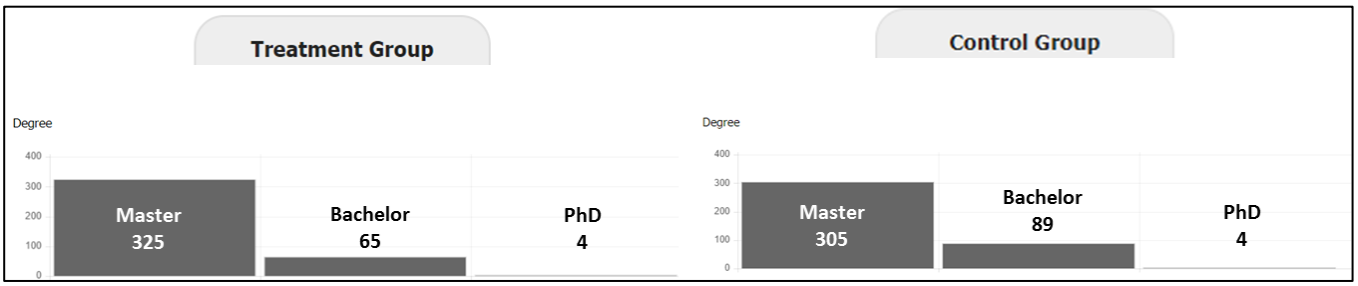


Figure 16: Study degree distribution in treatment and control group

Looking at the age of participants, as illustrated in Figure 17, the data shows in both experimental conditions that the majority of participants was between 22-24 years old (63.96% in the treatment vs. 56.78% in the control group). In the treatment group 111 (28.17%) of participants were older than 25 years while 31 (7.87%) indicated to be between 17-21 years of age. Similarly, 136 (34.17%) and 36 (9.05%) of the participants in the control group were older than 25 years or between 17-21 years old respectively.

Age

Treatment Group

Choices	Percentage	Count
22-24	*63.96%*	252
>25	*28.17%*	111
17-21	*7.87%*	31
	Total	**394**

Age

Control Group

Choices	Percentage	Count
22-24	*56.78%*	226
>25	*34.17%*	136
17-21	*9.05%*	36
	Total	**398**

Figure 17: Age distribution in treatment and control group

Figure 18 displays the conversion rate as well as the average completion time of both the treatment group and control group. The conversion rate of the treatment group lies at 67.70% which is higher than the 63.50% of the control group. Furthermore, the participants of the treatment group took less time to complete the questionnaire with only 3.88min on average compared to the control group with 4.35min on average.

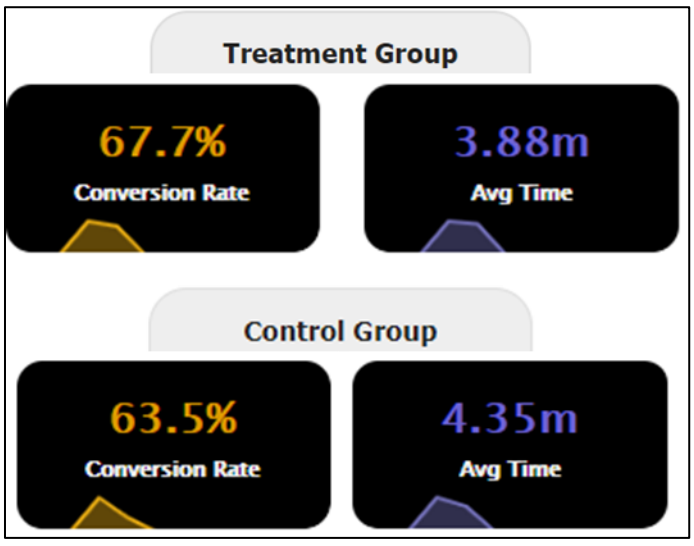


Figure 18: Conversion rate and average completion rate for treatment and control group

Comparing the number of e-mail signups between both groups in Figure 19, one can count 189 signups for the treatment and 175 signups for the control group, marking a difference of 4.00% in total.

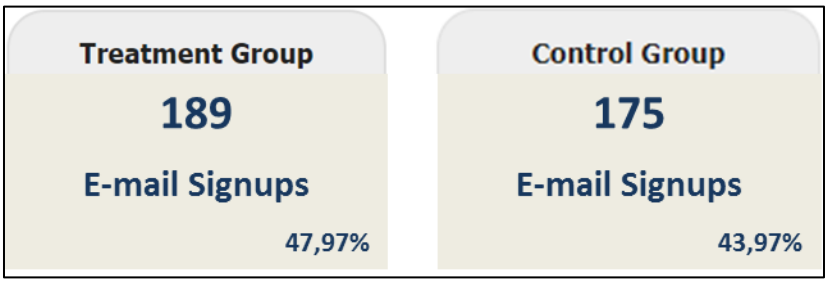


Figure 19: Comparison E-mail signups treatment vs. control group

The respective numbers of e-mail signups can be linked to the maximum level of engagement, level 6, and is visible in the histograms of the control group (Figure 20) and the treatment group (Figure 21). The control group reveals the following frequencies of levels of engagement, ordered from highest to lowest frequency: Level 6 was counted 175 times, level 5 (158), level 1 (45), level 2 (13), level 3 (5), and lastly level 4 occurred 2 times.

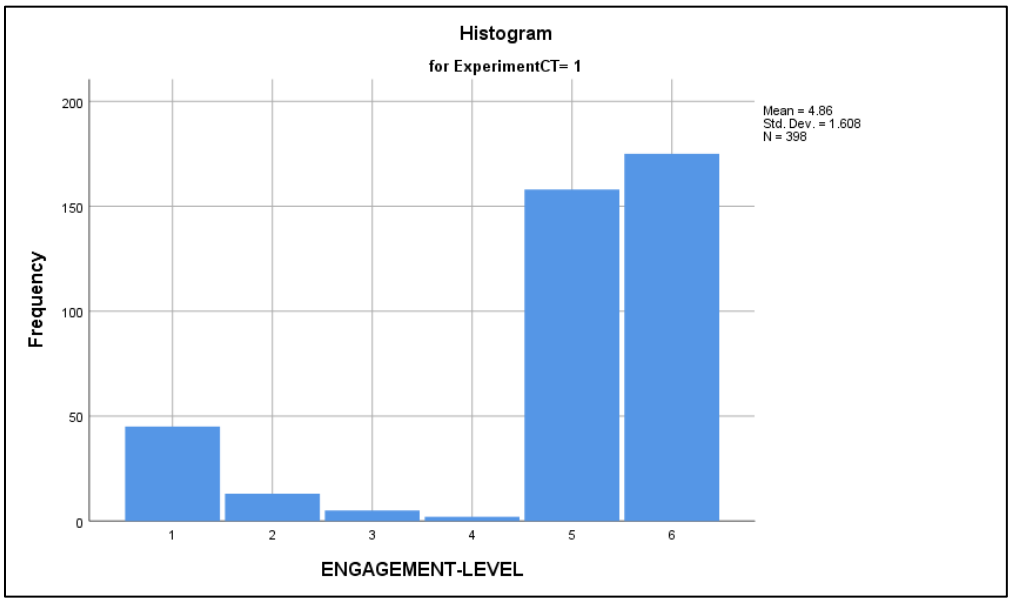


Figure 20: Histogram of level of engagement within the control group

In the treatment group, frequencies of levels of engagement were distributed in the following way, again ordered from highest to lowest frequency: Level 6 was counted 189 times, level 5 (161), level 1 (28), level 2 (13), level 3 (3), and lastly level 4 with no occurrence (Figure 21 below).

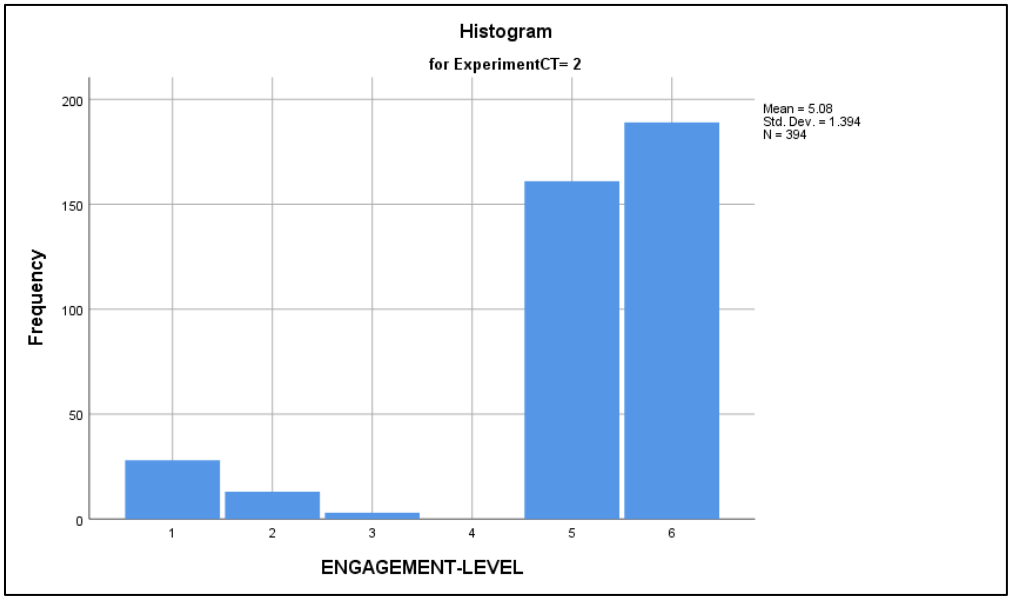


Figure 21: Histogram of level of engagement within the treatment group

All in all, the descriptive statistics of both experimental conditions did not only show an almost identical sample size, they also indicate very similar characteristics of participants in regard to age, gender and study degree. As a consequence, the conducted statistical calculations will yield towards more robust and meaningful results. Differences in the data such as in level of engagement cannot be concluded as significant until inferential statistics are carried out which are presented in the following subchapter.

4.3 Inferential Statistics

The present subchapter provides the reader with the results of the statistical tests on the data for the different hypothesis developed in the 'literature review' section of the thesis. It starts with explaining the selection for the non-parametric Mann-Whitney U-test for testing the statistical significance and Spearman Correlation between the groups relevant for the respective hypotheses. Generally, inferential statistics were obtained using the analytics software IBM SPSS statistics.

In order to fully investigate the impact of the game element countdown-timer on student engagement with consideration of demographic variables such as gender, degree and age; comparisons were drawn at

95% confidence interval ($\alpha = 0.05$). The nature of this study requires comparing the level engagement of participants comparing the control and treatment groups as well as within treatment and control groups.

4.3.1 Selection of Mann-Whitney U-test and Spearman Correlation

Analyzing two independent samples for significant statistical differences is commonly done using the independent samples t-test procedure (Field & Hole, 2002, p.270). Parametric tests like the t-test work on the arithmetic mean and data, therefore, must be measured at an interval or ratio level. Furthermore, the t-test assumes that the dependent variable is normally distributed (Field & Hole, 2002, p.163). The data obtained in the present experiment as in the level of engagement scale, however, have natural, ordered categories from 1 to 6; and the distances between the categories are not known. Hence, the present data is ordinal in nature which by definition excludes a normal distribution. Consequently, the Mann-Whitney U-test is taken as non-parametric equivalent of the independent samples t-test (Field & Hole, 2002, p.234). The Mann-Whitney U-test works on the median of a distribution and compares the latter with particular -but not exclusive- sensitivity to a location shift (Field & Hole, 2002, p.235). It holds the following three assumptions: Firstly, participants of both groups must be randomly sampled and independent from each other. Secondly, the dependent variable needs to be at least ordinal distributed. Lastly, both groups must have homoscedasticity, i.e. the same variance, which can be checked using Levene's test (Nachar, 2008; Ruxton, 2006; p.689). The first and second assumption is covered by the chosen research method and the ordinal nature of the level of engagement measured in the experiment. The third assumption of equal variance will be conducted individually for every tested hypothesis since the group compositions vary between the tests.

Nevertheless, solely relying on calculations on statistical significance can raise the following problem: It does not follow that the effect is important. As Field & Hole put forward, "very small and unimportant effects can turn out to be statistically significant [simply] because [large sample sizes] have been used in the experiment (Field & Hole, 2002, p.150). Consequently, the solution to this problem lies in measuring the size of the effect that is tested (Field & Hole, 2002). For non-parametric data like the present, the Spearman's rank correlation (1904) will be used as "objective and standardized measure of the magnitude of observed effect" (Field & Hole, 2002, p.152). Hereby, the widely accepted suggestions of Cohen (1988) will be taken as guidelines to assess the importance of the experimental effects (in addition to the

significance of the test statistic): r=0.10 constitutes a small effect, r=0.30 a medium effect; and lastly, r=0.50 a large effect (Cohen, 1988).

4.3.2 Comparison on level of engagement within European students

In the following, the level of engagement of all participating European students with exposure or absence of exposure of the game element countdown-timer is compared. Based on the findings of literature, the following hypothesis H_1 was tested:

H_1 = There is a difference in the level of engagement within European students exposed to the game element countdown-timer compared to the control group

Table 1 below provides the particular descriptive statistics of both the control and treatment group, including mean, standard error of the mean, median, variance as well as standard deviation.

Table 1: Descriptive statistics of control and treatment group

Level of Engagement		*N*	M*	SE*	*Mdn*	*VAR*	*SD*
General	Control	398	4.86	.081	5.00	2.585	1.608
	Treatment	394	5.08	.070	5.00	1.942	1.394
	Total:	794					

*included in the table, however, not relevant for statistical calculations

As mentioned before, the first two assumptions of independence between and within samples as well as the (at least) ordinal distributed dependent variable are covered for all Mann-Whitney tests in the present thesis. Nevertheless, in order to enable statistical significance, the assumption of homoscedasticity, i.e. equal variance based on the median needs to be checked. This can be done using Levene's test as shown in Table 2. Grounded on Levene's test, both populations are assumed to have same variance with p = 0.008 based on the median which is significantly greater than p < 0.05. Therefore, as the homoscedasticity-assumption is fulfilled, the p-value of a non-parametric Mann-Whitney statistic can be interpreted and provides meaningful conclusions on the test hypothesis.

Table 2: Test of Homogeneity of Variance using Levene's test

		Levene Statistic	df1	df2	Sig.
Engagement-Level	Based on Median	2.912	1	790	0.088
	Based on Median and with adjusted df	2.912	1	774.669	0.088

Consequently, a Mann-Whitney U test was performed to make comparisons in the level of engagement between control and treatment groups, and therefore test the hypothesis. This test indicated no significant difference in the level of engagement between the control and treatment group of the experiment with p = 0,093. Therefore, the H_1-Hypothesis cannot be confirmed. The difference in the mean of ranks as seen in Table 3 is not large enough to be statistically significant. Consequently, at confidence interval of 95%, there is no significant difference in the level of engagement of European students with exposure or absence of exposure of the game element countdown-timer.

Table 3: Results of the Mann-Whitney U test to compare the level of engagement in control and treatment group

Level of Engagement		Levene's test	*N*	Mean rank	*U*	*Z*	*p*
General	Control	.088	398	384.08	73462.50	-1.679	.093
	Treatment		394	409.05			
			Total: 792				

4.3.3 Comparison on level of engagement within gender of European students

In the following, the level of engagement of within gender of European students with exposure or absence of exposure of the game element countdown-timer is measured and compared. The comparisons have been drawn based on gender between control and treatment group as well as within treatment and control group in order to make significant statements about potential differences. As the game element countdown-timer fulfills the perceived psychological need of competence; which has been shown to be more effective for the male gender, male participants are assumed to show a higher level of engagement in comparison to female participants. Hence, following hypothesis will be investigated:

H_{2a} = There is a difference in the level of engagement within the male and female gender exposed to the game element countdown-timer

H_{2b} = There is a difference in the level of engagement between the male and female gender exposed to the game element countdown-timer

Table 4 provides the particular descriptive statistics of males and females in the control and treatment group, including mean, standard error of the mean, median, variance as well as standard deviation.

Table 4: Descriptive statistics of gender in control and treatment group

Level of Engagement		*N*	M*	SE*	*Mdn*	*VAR*	*SD*
Female	Control	167	4.92	.127	5.00	2.692	1.641
	Treatment	139	5.17	.107	5.00	1.603	1.266
	Total:	306					
Male	Control	231	4.82	.104	5.00	2.515	2.129
	Treatment	255	5.04	.091	5.00	1.586	1.459
	Total:	486					

*included in the table, however, not relevant for statistical calculations

Calculating the equality of variance for all four statistical calculations (indicated in Table 5), Levene's test shows no significance with a p-value greater than 0.05 across all tests. As the homoscedasticity-assumption is fulfilled, the p-value of a non-parametric Mann-Whitney statistic can be interpreted and provides meaningful conclusions on the test hypothesis. The Mann-Whitney test in Table 5 reveals that no significant difference was found between the control and treatment groups of both males with p = 0,057; as well as females with p = 0,702. Furthermore, the statistical results of within the treatment and the control group did not demonstrate significance with p = 0,809 for the treatment group and p = 0,152 for the control group. Therefore, both the $H2_a$ and $H2_b$-Hypotheses cannot be confirmed. Consequently, at confidence interval of 95%, there is no significant difference in the level of engagement within and between gender of European students with exposure or absence of exposure of the game element countdown-timer.

Table 5: Results of the Mann-Whitney U test to compare the level of engagement in control and treatment groups in regard to gender

Level of Engagement		Levene's test	*N*	Mean rank	*U*	*Z*	*p*
Male	Control	.593	231	231.80	26750.50	-1.904	.057
	Treatment		255	254.10			
			Total: 486				
Female	Control	.333	167	151.90	11339.00	-.382	.702
	Treatment		139	155.42			
			Total: 306				
Both	Treatment F	.239	139	199.20	17486.00	-.242	.809
	Treatment M		255	196.57			
			Total: 394				
Both	Control F	.652	167	208.46	17793.00	-1.431	.152
	Control M		231	193.03			
			Total: 398				

4.3.4 Comparison on level of engagement within age groups of European students

In the following, the level of engagement of within the age groups of European students is measured and compared being exposed or not being exposed to the game element countdown-timer. The two age groups are divided into 17-24 years and older than 25 years which represents the Generation Z and Generation Y respectively. The comparisons have been drawn based on age groups between control and treatment group as well as within treatment and control group in order to make significant statements about potential differences. As research has indicated that age influences the effectiveness of gamification on participants, the following two hypotheses will be investigated:

$H3_a$ = There is a difference in the level of engagement within the two age groups of European students exposed to the game element countdown-timer

$H3_b$ = There is a difference in the level of engagement between the two age groups of European students exposed to the game element countdown-timer

Table 6 provides the particular descriptive statistics of the two generational age groups in the control and treatment group, including mean, standard error of the mean, median, variance as well as standard deviation.

Table 6: Descriptive statistics of age in control and treatment group

Level of Engagement		*N*	M*	SE*	*Mdn*	*VAR*	*SD*
≤ 24 years	Control	262	4.77	.102	5.00	2.716	1.648
	Treatment	283	5.12	.081	5.00	1.846	1.359
	Total:	545					
> 25 years	Control	136	5.04	.130	6.00	2.302	1.517
	Treatment	111	4.97	.208	5.00	2.190	1.480
	Total:	147					

* included in the table, however, not relevant for statistical calculations

The assumption of homoscedasticity based on the median is checked by using Levene's test as shown in Table 7 and reveals no significance across all tests with p-values greater than 0.05. Therefore, the variances of the tested populations can be assumed as equal and the non-parametric Mann-Whitney test can be carried out. The Mann-Whitney test in Table 7 reveals that a significant difference was found between the control and treatment groups of participants younger than 25 years old with p = 0,007. Consequently, the p-value is even lower than 0.01, i.e. the probability that such a result could have occurred by chance alone only is less than one in a thousand, thus making this finding more robust. On the other side, no significant difference can be found between the control and treatment groups of participants older than 25 years with p = 0,392. Furthermore, the statistical results of within the treatment group did not demonstrate significance with p = 0,441. However, the Mann-Whitney test within the control groups of both ages reveals a significant difference as shown in Table 7 with p = 0,018.

Table 7: Results of the Mann-Whitney U test to compare the level of engagement in control and treatment groups in regard to age

Level of Engagement		Levene's test	*N*	Mean rank	*U*	*Z*	*p*
≤ 25 years	Control	.102	262	255.61	32518.00	-2.712	.007**
	Treatment		283	289.10			
			Total: 545				
> 25 years	Control	.921	136	127.21	7112.00	-.856	.392
	Treatment		111	120.07			
			Total: 147				
Age	Treatment Y	.438	283	200.01	56603.00	-.771	.441
	Treatment O		111	191.10			
			Total: 394				
Age	Control Y	.588	262	190.46	15446.50	-2.360	.018*
	Control O		136	216.92			
			Total: 398				

*p< 0.05, **p < 0.01

The Spearman's rank order correlation revealed a small, positive correlation between the participants in the age between 17-24 years old with r=0.116, as shown below in Table 8.

Table 8: Spearman's rank correlation between treatment and control group of age 17-24 years

			Engagement-Level
Spearman's rho	17-24 Years old (Generation Z)	Correlation Coefficient	.116**
		Sig. (2-tailed)	.007
		N	545

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9 reveals the Spearman's rank order correlation within participants of the control group, showing a small, positive correlation years old with r=0.118. Both calculated effect sizes (Table 8 and Table 9), therefore, account for a small effect of the countdown-timers on the participants.

Table 9: Spearman's rank correlation within in control group

			Engagement-Level
Spearman's rho	Within Control Group	Correlation Coefficient	.118*
		Sig. (2-tailed)	.018
		N	398

*. Correlation is significant at the 0.05 level (2-tailed).

All in all, the statistical tests on age groups reveal statistical significance between treatment and control group of participants younger than 25 years of age as well as within the control group including both age groups. Consequently, the two hypotheses $H3_a$ and $H3_b$ of the present comparison can partly be confirmed. To be more specific, the $H3_a$-Hypothesis can only be confirmed for participants younger than 25 years old, but not for the older age group. Similarly, the $H3_b$-Hypothesis can be confirmed for the control group, but not for the treatment group. The absence of countdown-timers seems to decrease the level of engagement within participants who belong to Generation Z even though the effect size is considered to be small.

4.3.5 Comparison on level of engagement within study degrees of European students

In the following, the level of engagement of within study degrees of European students with exposure or absence of exposure of the game element countdown-timer is compared. Since the sample size of PhD students in both experimental groups (4 in the control and 0 in the treatment group respectively) was too small, these participants could not be included in the analysis. This present comparison requires investigating the level engagement of participants across control and treatment group as well as within both treatment and control group in order to make significant statements about potential differences. As study degree and age are similarly distributed (visualized in Figure 16 and Figure 17), the assumption is

made that differences in study degrees can moderate the influence of countdown-timers on the level of engagement of participants in a similar way. The assumption can be confirmed with a small, positive correlation of r=0.182 under a statistical significance of p=0.000, as visualized in Table 10.

Table 10: Spearman's rank correlation between study degree and age

			Age
Spearman's rho	Degree	Correlation Coefficient	.182**
		Sig. (2-tailed)	.000
		N	790

**. Correlation is significant at the 0.01 level (2-tailed).

Hence, the following two hypotheses $H4_a$ and $H4_b$ were tested:

$H4_a$ = *There is a difference in the level of engagement within study degrees of European students exposed to the game element countdown-timer*

$H4_b$ = *There is a difference in the level of engagement between European bachelor and master students exposed to the game element countdown-timer*

Table 11 provides the particular descriptive statistics of males and females in the control and treatment group, including mean, standard error of the mean, median, variance as well as standard deviation.

Table 11: Descriptive statistics of study degrees in control and treatment group

Level of Engagement		*N*	M*	SE*	*Mdn*	*VAR*	*SD*
Bachelor	Control	89	4.64	.171	5.00	2.619	1.618
	Treatment	65	5.17	.162	5.00	1.705	1.306
	Total:	154					
Master	Control	305	4.93	.092	5.00	2.567	1.602
	Treatment	325	5.10	.140	5.00	1.885	1.373
	Total:	630					

*included in the table, however, not relevant for statistical calculations

Levene's test in Table 12 indicates that the variances between level of engagement distributions of bachelor and master students in both the treatment and control group as well as within the two groups are assumed to be equal with p-values greater than $p < 0.05$. As a consequence, the assumption of homoscedasticity is fulfilled and a Mann-Whitney U test can be carried out.

Illustrated in Table 12, the Mann-Whitney test reveals a statistically significant difference in the level of engagement within the participating bachelor students groups with $p = 0{,}010 < 0.05$. As a consequence, the $H4_a$-Hypothesis for bachelor students can be confirmed with $p < 0.05$. The bachelor students who participated in the experiment, therefore, differed significantly in their level of engagement when being exposed to the game element countdown-timer.

Table 12: Results of the Mann-Whitney U test to compare the level of engagement in control and treatment groups in regard to study degrees

Level of Engagement		Levene's test	*N*	Mean rank	*U*	*Z*	*p*
Bachelor	Control	.433	89	70.35	2256.00	-2.559	.010*
	Treatment		65	87.29			
			Total: 154				
Master	Control	.064	305	311.21	48254.50	-.629	.529
	Treatment		325	319.52			
			Total: 630				
Both	Treatment	.536	65	196.42	10503.00	-.079	.937
	Treatment		325	195.32			
			Total: 390				
Both	Control	.512	89	170.92	11206.50	-2.714	.007**
	Control		305	205.26			
			Total: 394				

*p < .05, **p < .01

On the other side, no significant difference can be derived between the control and treatment groups of participating master students with p = 0,529. Furthermore, the statistical results of within the treatment group did not demonstrate significance with p = 0,937. In the contrary, both bachelor and master students within the treatment group had an almost identical level of engagement based on the p-value of 0.937. However, the Mann-Whitney test within the control groups of both bachelor and master students reveals a significant difference as shown in Table 7 with p = 0,007. A p-value lower than 0.01, as indicated here, presents the probability that such a result could have occurred by chance alone only is less than one in a thousand. It is a very unambiguous and clear result showing the difference of engagement for bachelor students due to the absence of the game element countdown-timer.

The Spearman's rank order correlation revealed a moderate, positive correlation between the participants treatment and control group of bachelor and master students with r=0.207, as shown below in Table 13.

Table 13: Spearman's rank correlation between

			Engagement-Level
Spearman's rho	Within Bachelor students	Correlation Coefficient	.207*
		Sig. (2-tailed)	.010
		N	154

*. Correlation is significant at the 0.05 level (2-tailed).

Table 14 reveals the Spearman's rank order correlation within bachelor and master students in the control group, showing a small, positive correlation years old with r=0.137. Both calculated effect sizes (Table 13 and Table 14), therefore, account for a small to moderate effect of the countdown-timers on the participants.

Table 14: Spearman's rank correlation within in the group of bachelor and master students

			Engagement-Level
Spearman's rho	Within Control Group	Correlation Coefficient	.137*
		Sig. (2-tailed)	.006

	N	394

**. Correlation is significant at the 0.01 level (2-tailed).

All in all, the statistical tests on age groups reveal significant differences between treatment and control group of participating bachelor students as well as within the control group including both study degrees. As assumed, the results are similar to the findings with age as moderator variable. Consequently, the two hypotheses of the present comparison can partly be confirmed. To be more specific, the $H4_a$-Hypothesis can only be confirmed for participants pursuing a bachelor degree, but not for the master students. Similarly, the $H4_b$-Hypothesis can be confirmed for the control group but for the treatment group. The presence of countdown-timers seems to increase the level of engagement within participating bachelor students group significantly although the importance of the effect is small to moderate.

4.4 Summary of the findings

First of all, no significant difference was found in the level of engagement of European students with exposure or absence of exposure of the game element countdown-timer.

Secondly, there was no significant difference in the level of engagement within gender of European students with exposure or absence of exposure of the game element countdown-timer.

Thirdly, taking age as moderator variable into consideration revealed statistical significance between treatment and control group of participants younger than 25 years of age as well as within the control group including both age groups. The presence of countdown-timers seems to increase the level of engagement within the younger age group even though the effect size is indicated as small.

Finally, similarly to the comparison on age, statistical significant differences could be found between treatment and control group of participating bachelor students as well as within the control group including both study degrees. Hereby, the presence of countdown-timers seems to increase the level of engagement within participating bachelor students group significantly with small to moderate effect sizes.

All in all, the 'results' chapter aimed at revealing what results have been found in analyzing the data of the experiment. It leaves the interpretation of the results for the following section.

5. DISCUSSION

The present chapter aims at explaining the meaning of the results in relation to the developed hypotheses in the literature review as well as in relation to relevant gamification theory and previous research on the topic. As described in the previous section no significant differences in the level of engagement between and within the treatment and control group could be found in conducting the experiment. Furthermore, no significant differences in in the level of engagement between and within gender. On the other hand, significant differences could be found between and within different age groups and study degrees. In the following, these findings will be reviewed as to whether they are consistent or contradicting with theoretical accounts on gamification.

5.1 Differences in the level of engagement between and within treatment and control group

The results of the statistical calculations on significant differences in the level of engagement of European students with exposure or absence of exposure of the game element countdown-timer did not reveal significance between treatment and control group even though it showed a favorable statistical trend.

Applying the framework of self-determination theory (SDT), for participants to be engaged in a task or activity the three basic psychological needs of autonomy, competence and relatedness are required to be satisfied to a sufficient level. Fulfilling at least one of the needs can effectively increase intrinsic motivation to complete a task (Forde et al., 2016). The majority of gamification research to date, however, combined several game elements in gamification studies in various different ways, mostly addressing all three psychological needs (Koivisto & Hamari, 2014). As a consequence, findings could not be generalized since the studies were lacking the ability to be replicated, did not use control groups and did not isolate game elements in order to measure the individual effects (Landers et al., 2018). The present experiment, on the other hand, isolated the game element countdown-timer to measure its effectiveness on the engagement level of participants.

The countdown-timer incorporated in the experiment mainly addresses the psychological need of competence since completing the questionnaire before the countdown-timer ends is intended to provide the participants with a feeling of accomplishment. Generally, satisfaction of the psychological need of competence is achieved when users feel like they are progressing towards a meaningful goal (Richards et al., 2014; Sailer et al, 2017). Thus, the perceived level of competence is dependent on the degree of

meaningfulness that is associated with completing the questionnaire, i.e. the goal in the perspective of participants. There was no additional goal formulated that provided meaningfulness nor was it measured. Therefore, it is not possible to evaluate the degree of perceived competence within participants which in turn could have explained the obtained results. The findings suggest, however, that the psychological need of competence was not satisfied (enough) for participants to show a significant level of engagement.

The feeling of competence can also be connected to the work of Nakamura & Csikszentmihalyi (2014) where the balance of challenge and skill provides the optimal condition to experience flow and therefore the satisfaction of competence need in gamified contexts (Chan et. al., 2019; Hamari & Koivisto, 2014). The present experiment challenged participants in the treatment group to complete the questionnaire within the timeframe of five minutes. The emphasis within the framework of flow lies in an optimal balance, i.e. "just-manageable" challenges (Nakamura & Csikszentmihalyi, 2014, p.90). Applied to the present experiment, the optimal challenge is supposed to show an average completion time within the treatment group of about five minutes or shortly below. In contrast to that, the findings revealed a time below four minutes. Consequently, it cannot be assumed that participants were challenged optimally in order to experience flow to a higher degree and thus be more engaged in the experiment as such. Critically reflecting on the set up of the experiment, the researcher could have lowered the completion time and countdown-timer to four minutes in hindsight, based on the average completion time of 4:07mins revealed by functionality and usability testing before carrying out the actual experiment. By doing so, the completion of the questionnaire could have be turned into a "just-manageable" challenge and potentially revealing a higher level of engagement within the treatment group and, hence resulting in significant differences between the two experimental conditions.

Countdown-timers can be positioned within the loss and avoidance dimension of game element categories where users have to play against time to meet the requirements of a particular scenario (Economou et al., 2015). Thus, countdown-timers seem to be effective in enhancing engagement by exposing users or participants to a potential loss that they try to avoid such as negative marking and loosing points or badges that have been gained in the process already (Economou et al., 2015). In contrast to that, the present experiment was not designed in a way that exposes participants with potential losses which in turn could have made the experimental scenario more meaningful.

In addition to that, attitudes and behavior can be influenced by differences in how the game element is perceived. Countdown-timers and progress bars are elements that provide users with feedback on time left or progress respectively. Hereby, it can make a difference on intrinsic motivation when the implementation of countdown-timers is perceived as an informational and not controlling feedback, as shown with leaderboards and points by Forde et al. (Forde et al., 2016). When perceived as informational, the particular feedback mechanism can magnify the felt satisfaction of competence as the user gets a sense of getting closer to a desired outcome (Sailer et al., 2017). Given that the average time spent in both treatment and control condition indicated less than five minutes of participation, being exposed or not exposed to the countdown-timer might not have triggered the psychological need for competence as much as it could have done in order to produce a significant difference. Additionally, in relation to the former, the setting of the countdown-timer might not have generated a sense of urgency within participants which according to Kim et al. (2020) can change the participants' engagement (Kim et al., 2020). The emotion of urgency has not been measured throughout the experiment; consequently it is assumed that based on the findings no significant degree of urgency was created, which in turn, would have affected the participants more.

According to Cohen et al. (2003), situational moderator variables such as environmental conditions or economic downturns can influence the outcome of interest, too (Cohen et al., 2003). During the conducted experiment, participants experienced a global crisis triggered by a global virus outbreak, which in turn could have shaped the level of engagement. Nevertheless, the effect is assumed to have been cancelled out due to random allocation of participants and the large sample size employed in the experiment.

5.2 Differences in the level of engagement between and within gender

The results of the experiment revealed no significant difference in the level of engagement within gender of European students with exposure or absence of exposure of the game element countdown-timer. The hypothesis that there are no significant differences between male and female European students cannot be rejected for both treatment and control group. Literature on this topic, however, reveals the tendency towards different psychological and behavioral changes in men and women depending on the specific characteristics of the game element(s) in use. The findings of the present experiment in regard to gender differences are not congruent with literature on the topic.

In a more granular perspective, gamification is supposed to motivate and engage genders differently based on research on gender roles. As the commonly applied theory of motivation in the field, self-determination theory (SDT) proposes that motivation requires the satisfaction of the three basic psychological needs: autonomy, competence and relatedness (Ryan & Deci, 2000). Within gamification research of motivation and engagement, the fulfillment of the psychological need of competence has been shown to influence the behavior of men more than women while women are more influenced by the satisfaction of the psychological need for relatedness (Koivisto & Hamari, 2014; Oyibo, 2017; Yee, 2006). The countdown-timer incorporated in the experiment mainly touches the psychological need of competence since completing the questionnaire before the countdown-timer ends is intended to provide the participants with a feeling of accomplishment and competence, thus engaging them more in the experiment. These competitive characteristics are shown to influence male behavior more than female (Koivisto & Hamari, 2014). Furthermore, the countdown-timer has no social extensions attached to it; the experiment does not contain any element or dimension that elevates the satisfaction of the need for relatedness, which would potentially engage women more than men.

Consequently, based on this proposition male participants should have shown a higher engagement than female participants in the experiment. However, the findings reveal the proposition not to be true – there was no significant difference between male and female level of engagement. On the contrary, men and women showed a very similar level of engagement on average in the experiment. The reasons for this finding can be derived by two possible explanations:

First of all, the incorporation of the game element countdown-timer might not have fulfilled the psychological need for competence sufficiently. The experiment did not investigate if and, more specifically, to which degree the incorporation of the countdown-timer actualized in a change of attitude. The author assumes, therefore, that the countdown-timer in the experiment did not trigger nor fulfill the need for competition and achievement such as that it could have resulted in a psychological and behavioral change of the participants. As a result, psychological needs were equally fulfilled across genders, thus no difference in the level of engagement could be found.

Secondly, based on the research of Sailer et al. (2017), the targeted engaging effect of game design elements depends on how well the design was implemented in terms of aesthetics and quality (Sailer et al., 2017). Potentially, the reason for a similar gender behavior throughout the experiment could be

explained through an ineffective implementation of the game element in aesthetic and qualitative terms. However, the execution of functionality and usability testing before carrying out the actual experiment revealed positive responses in regard to aesthetic and quality of the overall experiment, including the questionnaire as well as the countdown-timer for the treatment condition. Based on the positive feedback, a potential explanation of poor aesthetic and qualitative design can be ruled out.

As a result, the current theory - suggesting gender differences in gamification effectiveness do exist - cannot be rejected or confirmed through the current experiment.

5.3 Differences in the level of engagement within and between age groups

The results of the experiment revealed significant statistical differences between treatment and control group of participants younger than 25 years of age as well as within the control group including both age groups, with small to moderate effect sizes. More precisely, the absence of the game element countdown-timers seems to decrease the level of engagement within the younger age group significantly more compared to participants older than 25 years of age.

First of all, the significant differences found confirm the hypotheses made based on the generational differences identified in the literature: Members of Generation Z, referring to participants in the experiment younger than 25 years old (born after 1995), are native to digital technologies and games (McGonigal, 2012; Seemiller & Grace, 2016; Twenge, 2017). This generation is assumed to have more prior gaming experience and exposure compared to any other generation (McGonigal, 2012). Prior gaming experience, in turn, seems to have a positive effect on the level of engagement as well as on performance within a gamified activity or experience (Bittner, 2013). Hereby, age discrepancies in the context of gamification are assumed to be similar to those in other digital gaming contexts (Koivisto & Hamari, 2014). Secondly, younger age groups are likely to perceive more flow and enjoyment as well as usefulness with gamified experiences compared to older age groups (Bittner & Shipper, 2014). Furthermore, the findings of Bittner & Shipper (2014) showed that more experience with games can lead to higher perceived control of gamified services, most likely based on the familiarity with specific game mechanics such as countdown-timers (Bittner & Shipper, 2014). A higher degree of perceived control is related to the psychological need of autonomy since the latter is found by giving users a sense of control and freedom over one's actions (Ryan & Deci, 2000). In relation to the conducted experiment, it was assumed that participants within in Generation Z, i.e. younger than 25 years, would perceived a higher degree of control

over the experience, and thus in the level of engagement in the present experience. Although research is backing the claim that age can make a difference in the perceived control over the experience, and thus in the level of engagement and immersion (Bittner & Shipper, 2014), the present experiment cannot assume it to be the case in the present findings since neither prior gaming experience nor the degree of fulfillment of the psychological need of autonomy were measured in the study.

All in all, contrarily to both theoretical explanations, the engagement-level within Generation Z participants was not higher compared to the one of Generation Y participants in general and in comparison of the treatment groups, rather they were found at a surprisingly identical level. As theory suggested, younger age groups were supposed to be more sensitive towards and affected by the incorporation of gamification (Brauner et al., 2013). The present experiment, however, revealed the opposite as in younger age groups being more sensitive towards the absence of gamification, not the presence of it. This phenomenon could be observed in a lower level of engagement when members of the Generation Z were not exposed to the gamified version. More specifically, as the findings revealed younger participants (<25 years) were significantly less engaged in the control group.

The reason for the findings of the experiment can be explained by one particular characteristic that is associated with members of Generation Z: They face difficulties staying focused and keeping attention (Berkup, 2014; Bíró, 2014; Eck, 2006; Seemiller & Grace, 2016). Related to the present experiment, it is assumed that the absence of the game element countdown-timer decreased the focus and attention of younger participants in the experiment under which their level of engagement suffered. All in all, it is characteristic for this generation to switch tasks, especially when being online during the experiment while having endless possibilities to be distracted (Twenge, 2017).

As a consequence, current theory can be enhanced based on the finding that the incorporation of one game element or gamification as such seems like to be taken for granted by the younger participants. The implication hereby is that without gamification, participants younger than 25 years old will be more likely to be disengaged, while not necessarily being more engaged than other age groups through the introduction of a single game element. It comes as no surprise that gamification is considered to be useful (even necessary) for Generation Z, especially for enhancing motivation as well as in increasing attention (Biro, 2014; Geck, 2007; Sanmugam et al., 2016; Skinner, 2018). Nevertheless, it raises a critical issue at the same time: Are members of Generation Z generally “spoiled” by games and gamification? Are they

more externally controlled and less intrinsically motivated than other generations such as Generation Y? The latter seemed to be quite insensitive towards the countdown-timer in the present experiment. How can we go about it as in educational settings and in broader society? How can gamification be part of the solution and not of the problem? Is it even a problem? These questions cannot be answered at this point, but they pose serious practical implications.

5.4 Differences in the level of engagement within and between study degrees

The results of the experiment revealed significant statistical differences in the level of engagement between treatment and control group of participants who pursue a bachelor degree. Moreover, a significant difference could be found within bachelor and master students of the control group. More precisely, the absence of the game element countdown-timers seemed to increase the level of disengagement within bachelor students significantly more compared to participants who pursued a master degree at the time the experiment was conducted.

Based on the almost identical statistical results, the presented findings between and within different study degrees can be linked to the findings explained and qualified in the previous subchapter on age differences. This assumption was confirmed by testing the correlation between age groups and study degrees. Nevertheless, the largest effect size throughout all conducted statistical calculations related to the present experiment was found in the moderately strong correlation in the level of engagement between the treatment and control group of bachelor students. Consequently, bachelor students seem to be disengaged the most by the absence of gamification or more specifically, by being exposed to a “plain” questionnaire landing page, compared to any other group or moderator variable tested in the experiment.

Based on empirical results, it is assumed that the difference in intrinsic motivation between bachelor and master students might be larger compared to the generational differences found. Therefore, master students seem to be engaged with and without the presence of the countdown-timer. Bachelor students, potentially being less motivated by the content itself but rather extrinsic incentives, revealed to be significantly more sensitive to the game element. To explain these findings in more depth and in relation the theory, however, complementary qualitative research and psychometric measurements are necessary and will therefore be proposed in the following limitation and future research sections.

6. LIMITATIONS

The following chapter provides insights into the limitations of the present study. The limitations cover a critical reflection based on the chosen research philosophy and research strategy as well as the execution of the experiment.

6.1 Limitations of post-positivistic ontology and epistemology

First of all, the present study contains a lack of understanding. Gamification as a social phenomenon is not considered in its subjective, individual aspects and with empathy through the lens of post-positivism (Saunders, 2009). Human behavior therein entails a vast complexity of feelings, emotions, thoughts and motives that ultimately shape the outcome. Investigating such a vastly complex phenomenon by only relying on isolation of few variables (albeit accounting for the context) therefore, is tenuous.

Secondly, post-positivism assumes that the researcher can formulate objective inferences and conclusions as long as the experiment is objective and the researcher disregards own emotions. As human behavior naturally comes with and is driven by emotional responses, the research cannot guarantee to completely disregard human behavior. The manipulation of variables, as executed in the present thesis, is a practice that cannot be seen as completely objective. As Marsh & Smith have formulated, "i*n the social sciences […] subjective ontological and epistemological positions should not be treated like a pullover that can be ''put on'' when we are addressing such philosophical issues and 'taken off' when we are doing research"* (Marsh & Smith, 2005: p.531).

6.2 Limitations to deductive reasoning in the present thesis

Firstly, the limitation to deductive reasoning generally lies in the assumption of all statements or theories to be true. Ultimately, the complete study derives from the initial literature on gamification that one has to assume to be true. For example, if the initial statements are based on theoretical frameworks or are inferences based on observations, it means they have arisen from inductive reasoning which becomes problematic when the set premises are not true or the terms not clear. Conclusions of deductive logic can only be true when the former stands on firm ground. Relating it to the present context of gamification, research on the topic is still very young and scattered. As described in the literature review section, the field of gamification still lacks a coherent research practice and sufficient empirical studies with interpretable and generalizable research findings (Koivisto & Hamari, 2019; Landers et al., 2018). Since

most of the empirical research on the topic was conducted without control groups and without identical or validated measurement instruments for particular variables that are likely to account for the effectiveness of gamification (Koivisto & Hamari, 2018), the theory on gamification might not stand on firm ground. Consequently, the present thesis attempts to address the described lack within research by conducting research based on the construct presented by Landers et al. (2018) and by doing so, it focuses on a quite contemporary and evolving research problem as research within gamification did not concern itself with investigating the effects of single game elements before. Therefore, leaning on the theory of gamification in general to prove its effectiveness can pose a risk when following a deductive approach in the study.

Secondly, as demonstrated through cognitive experiments, both inductive and deductive reasoning contain cognitive errors or biases. Most relevant for the present thesis is the so called "confirmation bias", i.e. the tendency to look for evidence and interpret data in a way that fit to assumptions, expectations or, simply put, the hypothesis in hand (Nickerson, 1998). As a consequence, the researcher might miss the occurrence of certain phenomena or the other relationships between variables due to the focus on theory or hypothesis testing rather than being focused on the theory of hypothesis generation, too (Nickerson, 1998). Even though the researcher has attempted to rule out any errors in the experiment as well as in the collection and analysis, one has to admit that cognitive biases might still occur without being aware of it. As a result, it is important to start a conversation with fellow gamification researchers and work on advancing the field together by complementing each other's blind spots.

6.3 Limitations to the mono-quantitative method

When conducting quantitative studies like the present, researchers encounter difficulties when it comes to the interpretation of the collected data. The quantitative method, consequently, does not consider the meaning behind a social phenomenon like gamification. In addition to that, it does not allow participants to explain e.g. their level of engagement and its meaning it had in the present experiment (Carr, 1994). Following the quantitative research strategy allows to find answers to specific questions in order to prove or disprove a particular hypothesis while not being concerned about the motives of individuals that participate in the study.

While gamification research aims at finding causal interlinks between the incorporation of single or multiple game elements and targeted behavioral changes, it is also concerned about altering attitudes and other psychological states with incorporated game elements (Landers et al., 2018). According to Hamari et al. (2014), both dimensions should be at the core of investigating any gamification intervention (Hamari et al., 2014). To understand these changes in psychological states and attitudes, however, the execution of quantitative research only is not sufficient and therefore poses a limitation to the present study. Herein, there are no answers to the question of "why" participants engaged in the specific level that was measured in the experiment. Therefore, explanations of statistically calculated relationships cannot claim accuracy and must remain assumptions, especially when investigating a specific and contemporary phenomenon such as the presence of countdown-timers that was not studied isolated in prior gamification research. Still, dimensions such as the satisfaction of the psychological need satisfaction of autonomy, competence and control could have been asked in the experiment; however, they have been left out to avoid potential biases in participants.

Nevertheless, even the presence of insights and data on the degree of fulfillment of particular psychological needs would still provide only a limited insight into the meaning behind the phenomenon when solely employing a quantitative method, based simply on the lacking opportunity to ask for clarity and due to the anonymous nature of the collected data. All in all, it can be said that the quantitative method in the present context does raise many questions about the reasons and meanings behind certain causal relationships of the implemented gamification that cannot be answered and consequently pose a limit to the present study.

6.4 Limitations to the experimental set up

The first limitation in regard to the setup of the experiment can be found in the nature of the questionnaire. All fields to be filled out were set up as 'required' by the researcher, i.e. only by giving an answer to every field, participants were able to proceed to the next page. There was only one field that was not required to be filled out: The field where participants could type in their e-mail on the last page. The assumption is that many participants were so used to filling out every blank ("maturation-effect") which could have led to them inserting their e-mail address while not being very engaged in the experiment, thus skewing the results of the study. Since there is no certainty as in whether or not this effect played a role in the experiment, it poses a limitation to be presented.

Secondly, prior research examining the effect of particular game elements on participants and found that general attitude towards these elements was a moderator of the effectiveness of gamification (Mekler et al., 2017). More precisely, participants who have been over-exposed to the game element leaderboard might have (developed) negative attitudes toward the concept of leaderboards, thus decreasing the users' engagement and altering the strength of the effect (Landers et al., 2018). As the example shows, design-relevant moderator variables such as the general attitude towards a particular game element influence the effect of the latter on targeted psychological states. Applied to the present experiment, a limitation can be identified such as that the general attitude of participants towards countdown-timers has not been tested in the either the usability testing or in the actual experiment. Thereby, a potential positive or negative moderator variable that might have influenced the level of engagement of participants has not been considered in the experiment. Additionally, the same statement can be made about prior gaming experience of participants which has been shown to influence their behavior and could consequently serve as explanation of the results (Bittner & Shipper, 2014). Not setting up a measurement for both the general attitude towards the game element countdown-timer as well as for prior gaming experience contributes to a lower resolution of understanding the results and thereby poses a limitation invoked by the setup of the experiment.

6.5 Limitations to the generalizability of the results

First of all, the target population of participants, i.e. European students could potentially be improperly represented in the experiment. Despite applying a randomized sampling strategy to achieve proportional equality of participants, the latter cannot be viewed as given. Unequal prevalence of different study areas such as information technology and engineering on the analytical spectrum and psychology or sociology on the social science spectrum could potentially skew the results since gamification effects individuals differently based on their personality (Buckley & Doyle, 2017). In the present experience, participants were recruited through the professionals network LinkedIn only which presents several potential biases such as in collecting data from a more career oriented target population that possibly differs from the general European student population. In consideration of a having limited resources in regard to time, it was not feasible for the present experiment to contact students of specific study degrees equally in order to generate a proportional representation in the sample for both treatment and control group. Even carrying out the described approach contains flaws since (a) personality in study degrees is subject to

certain variation, too; and (b) response and acceptance rate cannot be controlled by the experimenter as well. The assumption lies in the cancelling out effect of a larger sample. Nevertheless, the present experiment lacks at least the measurement of the respective study degrees and personality types. Therefore, the potential falsity in proportion poses a limit to the generalizability of the presented findings.
Furthermore, there is a limit as to how many participants can be contacted on the LinkedIn platform daily. During the current experiment, the researcher came to a limit where contacting more participants was not possible and had to be stopped. Therefore, one needs to plan the time for conducting such a participant recruiting strategy ahead and over a certain period of time.

Secondly, the researcher faces the problem of not being able to control the environment of the experiment (Baxter, 2008). The investigated level of engagement of participants can possibly depend on the specific conditions occurring before or during the particular time of visiting the landing page of the experiment. Based on random allocation of participants and the large sample size, however, the impact of specific conditions is assumed to have cancelled out. Nevertheless, not having at least a certain degree of insight and understanding of the particular circumstances the participants where in while participating in the experiment does limit its generalizability.

Thirdly, as Field & Hole (2002) put forward, the best measure for generalizable results lies in empirical testing such as through replications of the experiment on other populations and other contexts (Field & Hole, 2002). Even though it was scheduled to be done, it was not possible to replicate the present experiment in the timeframe of the thesis project. As a consequence, the present study lacks in strength of generalizability since it has not been tested in a different context than the present in order to find out if the similar effects and findings will be obtained under other environmental conditions. In addition to that, the experiment was conducted in the midst of a global pandemic affecting the lives of every participant which in turn could have influenced the level of engagement in both groups. As the present circumstances are characterized by a high degree of volatility and ambiguity, it might be difficult to replicate the same experiment under similar circumstances. The timing of the study, therefore, might represent a limitation for the present study on gamification.

7. FUTURE RESEARCH

7.1 Psychometric measurement

Future research on individual game elements or on combinations of many should extend the study by adding another dimension: psychometric measurement. One the one hand, psychometric measurements can be integrated into the study to better understand the causal relationships between the game element and the targeted outcome while still following a post-positivist philosophy. On the other hand, using psychometrically reliable and valid measurement tools increase scientific rigor and the trustworthiness of the findings. More precisely, the former can be operationalized by measuring and accounting for the coefficient alpha as an estimate of reliability (Cortina, 1993). Overall, it increases the test validity, one of the key goals of any proper gamification study, and thus the quality of conclusions that can be made based on the experiment (Landers et al., 2018).

As mentioned above, psychometric measurement enables a better understanding of individuals that participate in a gamification study, too. Since gamification is concerned with altering attitudes and other psychological states such as motivation and engagement which in turn lead to (desired) behavioral changes, psychometric measurement offers the opportunity to understand these psychological changes with a higher resolution and should therefore be at the core of any gamification intervention conducted by future researchers (Hamari et al., 2014, Landers et al., 2018). The big-five personality test could be an example of a psychometric measurement to investigate personality structures and relate it to the results of a gamification intervention (Codish & Ravid, 2014; DeYoung et al., 2007; McCrae & John, 1992).

In addition to that, measurement scales can be created by future researchers on their own in order to measure the general attitudes of participants towards the game elements used in an experiment. Nevertheless, it is important hereby to measure the reliability of the construct to provide more scientific rigor and increase the quality of conclusions drawn upon it. Within gamification, limited valid psychometric instruments are found to describe individuals regarding their gaming preferences. One approach is the Bartle Player Type that origins from the analysis of massive multiplayer online games and that contains a scale with the three main dimensions immersion, achievement and social for different gaming motives (Brauner et al., 2013; Hamari & Tuunanen, 2014; Yee, 2006).

Furthermore, the satisfaction of the psychological needs autonomy, mastery and relatedness could be measured and accounted for as the basis of self-determination theory which is commonly used as framework to explain the effectiveness of gamification on motivation; however, it is rarely tested and included in empirical research on gamification. All in all, the core statement is that future research needs to concern itself with measuring gamification interventions psychometrically in a reliable and valid way to increase the quality and the trustworthiness of the findings.

7.2 Mixed-method research strategy

For future gamification research, a mixed-method research strategy can provide a more nuanced and deeper understanding of the human behavior involved. Both quantitative and qualitative research strategies have their strengths and weaknesses. In combination, however, they can be extremely effective with one another. Applied to research philosophy, future social scientists on gamification can attempt to fuse the two halves of social phenomenal experience, (post-)positivism on the one side and the subjective, interpretivism on the other (Lincoln et al., 2011). As a consequence, based on systematic integration of the two worlds, they do not necessarily have to tolerate less rigor in their methodological approach while still being able to understand the psychological outcomes that gamification produces. Qualitative research in gamification can give study participants a voice and ensures that the findings of a study are grounded in the experience of participants. Furthermore, it provides future researchers with in-depth perspectives and insights that might enable meaningful reflections on and explanations for the quantitative findings, especially in the case of obtaining contradictions between quantitative results and qualitative findings. While taking into consideration that future researchers integrating both strategies require greater resources and time, the premise lies in a more complete and synergistic utilization of data, and thus in greater strength, richness and, most importantly, meaningfulness of gamification research.

7.3 Competition and Incentivization within the experiment

First of all, future researchers aiming to replicate the present experiment could adjust either the content or the length of the experiment in order to make it more competitive. As seen in the descriptive statistics, users of both control and treatment group completed the questionnaire on average faster than the indicated five minutes. The assumption is that by extending the questions to be answered or by shortening the time of the countdown, participants might be more engaged in the experiment. As a

possibility, future researcher could test the assumption by building two treatment and control groups that have two different time ranges for the completion of the questionnaire. An adequate measurement hereby, lies in the perceived degree of satisfaction of the psychological needs of autonomy and competence in both groups and between genders. Furthermore, it can be of interest for future research to measure the individual completion times as an additional measurement of engagement and the influence of the game element countdown-timer in general.

Secondly, future studies could incentivize the participants for the completion of the questionnaire in order to increase the level of engagement in the presence or absence of countdown-timers or other game elements. This can be achieved by integrating and internalizing extrinsic motivation that is congruent with the users' intentions based on self-determination theory such as being faster than other participants or getting exclusive information about the study when finishing in less than the indicated time. Mechanisms like these need to be carefully controlled for in the experiment, nevertheless, they offer valuable insights into possible shifts of behavior within the participants.

Lastly, future research on gamification should build on the present study and continue testing the individual effects of various different game elements with scientific rigor. In addition to that, future studies could even consider testing the effects of different game elements in combination, such as countdown-timers and progress bars together. Combinations like these, when carried out in a properly designed experimental setting, allow future researchers to investigate their correlation and account for the results and thus advance the understanding and practical implications of gamification.

8. PRACTICAL IMPLICATIONS

The study presents two main practical implications for researchers and practitioners of gamification: Dispensing with and waiving gamification could potentially disengage bachelor students or more generally students that are younger than 25 years. On the other hand, spoiling a generation with too much of gamification might harm their intrinsic motivation to be engaged from within.

First of all, the present study has shown that there are significant differences in the level of engagement within bachelor students and students younger than 25 years old when not being exposed to the game element countdown-timer. More precisely, it revealed that these two groups are significantly more disengaged without exposure to gamification. Consequently, the implication lies in integrating

gamification across education, marketing, as well as other industries and contexts. By doing so, practitioners at least do not lose the Generation Z. However, it also poses the practical question: Do practitioners want to play to not lose or do they want to play to win (the Generation Z)? Grounded in the findings of the present study, practitioners receive a step-by-step plan of how to carefully integrate a single, well-designed game element into a specific context, and thus to at least not lose, i.e. disengage the bachelor students and students younger than 25 years of age. The presented countdown-timers can be used in various settings in practice: For time-limited tasks such as (online) exams or lectures in academic settings, (virtual) meetings in organizational contexts, and advertising material in the internet, wherein the latter already is often seen to apply gamification and more specifically countdown-timers.

The second practical implication lies in addressing an important question to society as a whole: Do we possibly harm the motivation of Generation Z students by exposing them too much to gamification? Generation Z consists of "Digital Natives" only and has therefore been exposed to digital technologies, games and gamification the most in their upbringing. The study shows that they achieve a similar level of engagement compared to older students only when being exposed to the game element countdown-timer. Consequently, younger students' intrinsic interest in the content of the study suffered significantly. This finding implies that motivation or the general level of engagement of a steadily growing proportion in higher education and in organizational contexts could rely on the incorporation of extrinsic factors such as the game elements countdown-timer that holds engagement at a normal level. In practice, it might not be sustainable and beneficial for the present and future of society. Public decision-makers and educators, therefore, can raise issues like these and open a discussion of how to go about technology use and exposure to extrinsically motivating game elements at young age. A practical solution could be to incorporate more game elements in educational and organizational settings that touch upon intrinsic motivation for the greater benefit of society.

9. CONCLUSION

Taking the critical reflections and problems within current gamification research into account, the present thesis aimed at applying a coherent gamification research practice that measures the effect of one single game element accurately, and thus leads to replicable and generalizable results. Consequently, in order to answer the research question, the researcher adopted a post-positivist perspective and conducted a true experiment by investigating the cause-effect relationship between the game element countdown-timer and the level of engagement of over 800 participating European students on a gamified online questionnaire. While the research question was developed carefully by addressing the gap in current gamification literature and scientific rigor was applied in the execution and analysis of the experiment, the present thesis is limited by a lack in understanding of the participants' behavior based on the chosen research approach as well as by a lack of replication of the experiment in another context.

Nevertheless, the present thesis on gamification still revealed interesting findings in regard to the research question: While the incorporation of a countdown-timer generally did not achieve statistical significance in the level of engagement within European students, the absence of the game element did decrease engagement particularly within participants younger than 24 years old. Members of Generation Z, therefore, are significantly more disengaged when not being exposed to the game element countdown-timer in the context of a landing page questionnaire. The practical implication lies in gamification and games, as the initial quote of the thesis suggested, therefore, seeming to be "the new normal" particularly for those individuals born after 1995.

Therefore, the present thesis has contributed to empirical research on gamification by applying a coherent research practice that isolated the effect of a single game element; and has advanced the understanding of demographic differences such as gender and age in the effectiveness of gamification.

With the still existing novelty of the gamification phenomenon, many directions for further research should be considered, such as the suggested relationship to psychometrics and personality traits. The emphasis and proposition hereby, lies in conducting research that leads to valid and generalizable findings. It is the author's hope that the present thesis and its critical reflection present a little step towards changing the current trajectory of gamification research.

In closing both this thesis and chapter of life:

Let us make proper and coherent gamification research the new normal.

REFERENCES

Aaker, D. A., Kumar, V., & Day, G. S. (2008). Marketing research. John Wiley & Sons.

Adamou, B. (2018). Games and Gamification in Market Research: Increasing Consumer Engagement in Research for Business Success. Kogan Page Publishers.

Alsawaier, R. S. (2018). The effect of gamification on motivation and engagement. The International Journal of Information and Learning Technology.

Armstrong, M. B., Ferrell, J. Z., Collmus, A. B., & Landers, R. N. (2016). Correcting misconceptions about gamification of assessment: More than SJTs and badges. Industrial and Organizational Psychology, 9(3), 671-677.

Bai, S., Hew, K. F., & Huang, B. (2020). Is gamification “bullshit”? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. Educational Research Review, 100322.

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid?. American Political Science Review, 104(2), 226-242.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

Bassiouni, D. H., & Hackley, C. (2014). 'Generation Z'children's adaptation to digital consumer culture: A critical literature review. Journal of Customer Behaviour, 13(2), 113-133.

Berkling, K., & Thomas, C. (2013, September). Gamification of a Software Engineering course and a detailed analysis of the factors that lead to it's failure. In 2013 International Conference on Interactive Collaborative Learning (ICL) (pp. 525-530). IEEE.

Berkup, S. B. (2014). Working with generations X and Y in generation Z period: Management of different generations in business life. Mediterranean Journal of Social Sciences, 5(19), 218-218.

Bernard, H. R., & Bernard, H. R. (2013). Social research methods: Qualitative and quantitative approaches. Sage.

Bíró, G. I. (2014). Didactics 2.0: A pedagogical analysis of gamification theory from a comparative perspective with a special view to the components of learning. Procedia-Social and Behavioral Sciences, 141, 148-151.

Bista, S. K., Nepal, S., Paris, C., & Colineau, N. (2014). Gamification for online communities: A case study for delivering government services. International Journal of Cooperative Information Systems, 23(2), 1-25.

Bittner, J. V., & Shipper, J. (2014). Motivational effects and age differences of gamification in product advertising. Journal of consumer marketing.

Blanche, M. T., Blanche, M. J. T., Durrheim, K., & Painter, D. (Eds.). (2006). Research in practice: Applied methods for the social sciences. Juta and Company Ltd.

Blohm, I., & Leimeister, J. M. (2013). Gamification. Wirtschaftsinformatik, 55(4), 275-278.

Bogost, I. (2011). Persuasive games: exploitationware. Gamasutra. Online:< http://www. gamasutra. com/view/feature/134735/persuasive_games_exploitationware. php.

Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. American educational research journal, 5(4), 437-474.

Brauner, P., Valdez, A. C., Schroeder, U., & Ziefle, M. (2013, July). Increase physical fitness and create health awareness through exergames and gamification. In International Conference on Human Factors in Computing and Informatics (pp. 349-362). Springer, Berlin, Heidelberg.

Bryman, A. (2016). Social research methods. Oxford university press.

Buckley, P., & Doyle, E. (2017). Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. Computers & Education, 106, 43-55.

Burke, B. (2016). Gamify: How gamification motivates people to do extraordinary things. Routledge.

Burrell, G., & Morgan, G. (2017). Sociological paradigms and organisational analysis: Elements of the sociology of corporate life. Routledge.

Butler, C. (2015, August). Reimagining Project Management Applications via Gamification. In International Conference on HCI in Business (pp. 437-445). Springer, Cham.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1983). Beyond external validity. Journal of Consumer Research, 10(1), 112-114.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. Psychological bulletin, 54(4), 297.

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

Carr, L. T. (1994). The strengths and weaknesses of quantitative and qualitative research: what method for nursing?. Journal of advanced nursing, 20(4), 716-721.

Carr, D. (2005). Contexts, gaming pleasures, and gendered preferences. Simulation & Gaming, 36(4), 464-482.

Chan CK., Leung HM., Kung MW. (2019) Understanding the Effect of Gamification of Learning Using Flow Theory. In: Ma W., Chan W., Cheng C. (eds) Shaping the Future of Education, Communication and Technology. Educational Communications and Technology Yearbook. Springer, Singapore

Chou, Y. K. (2019). Actionable gamification: Beyond points, badges, and leaderboards. Packt Publishing Ltd.

Cialdini, R. B., & Cialdini, R. B. (1993). Influence: The psychology of persuasion.

Cirillo, F. (2006). The pomodoro technique (the pomodoro). Agile Processes in Software Engineering and, 54(2), 35.

Codish, D., & Ravid, G. (2014). Personality based gamification-Educational gamification for extroverts and introverts. In Proceedings of the 9th CHAIS Conference for the Study of Innovation and Learning Technologies: Learning in the Technological Era (Vol. 1, pp. 36-44).

Cohen, J. (1988). Statistical power analysis for the behavioral sciences {Ind cd. Hillsdale. NJ: Erihaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.

Conley, K., & Donaldson, C. (2014). Gamification: The measurement of benefits. In Gamification in education and business (pp. 673-688).

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of applied psychology, 78(1), 98.

Cook, T. D., & Campbell, D. T. (1979). The design and conduct of true experiments and quasi-experiments in field settings. In Reproduced in part in Research in Organizations: Issues and Controversies. Goodyear Publishing Company.

Crano, W. D., Brewer, M. B., & Lac, A. (2014). Principles and methods of social research. Routledge.

Cugelman, B. (2013). Gamification: what it is and why it matters to digital health behavior change developers. JMIR serious games, 1(1), e3.

Dichev, C., Dicheva, D., & Irwin, K. (2018, July). Gamification Driven Learning Analytics. In International Conference on E-Learning; Kidmore End (pp. 70-76).

Darina Dicheva, Christo Dichev, Gennady Agre, & Galia Angelova. (2015). Gamification in Education: A Systematic Mapping Study. Journal of Educational Technology & Society, 18(3), 75-88.

da Rocha Seixas, L., Gomes, A. S., & de Melo Filho, I. J. (2016). Effectiveness of gamification in the engagement of students. Computers in Human Behavior, 58, 48-63.

DePoy, E., & Gitlin, L. N. (2019). Introduction to research E-book: Understanding and applying multiple strategies. Mosby.

De Sousa Borges, S., Durelli, V. H., Reis, H. M., & Isotani, S. (2014, March). A systematic mapping on gamification applied to education. In Proceedings of the 29th annual ACM symposium on applied computing (pp. 216-222).

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York: Plenum.

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

Deterding, S. (2015). The lens of intrinsic skill atoms: A method for gameful design. Human–Computer Interaction, 30(3-4), 294-335.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. Journal of personality and social psychology, 93(5), 880.

Ding, D., Guan, C., & Yu, Y. (2017). Game-Based learning in tertiary education: A new learning experience for the Generation Z. International Journal of Information and Education Technology, 7(2), 148.

Dymek, M., & Zackariasson, P. (Eds.). (2016). The business of gamification: a critical analysis. Taylor & Francis.

Economou, D., Doumanis, I., Pedersen, F., Kathrani, P., Mentzelopoulos, M., & Bouki, V. (2015, July). Evaluation of a dynamic role-playing platform for simulations based on Octalysis gamification framework. In Intelligent Environments (Workshops) (pp. 388-395).

Eickhoff, C., Harris, C. G., de Vries, A. P., & Srinivasan, P. (2012, August). Quality through flow and immersion: gamifying crowdsourced relevance assessments. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 871-880). New York, NY: ACM.

Ethridge, D.E. (2004) “Research Methodology in Applied Economics” John Wiley & Sons, p.24

Fairchild, A. J., & McQuillin, S. D. (2010). Evaluating mediation and moderation effects in school psychology: A presentation of methods and review of current practice. Journal of School Psychology, 48, 53-84.

Field, A., & Hole, G. (2002). How to design and report experiments. Sage.

Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2017). Replicability and other features of a high-quality science: Toward a balanced and empirical approach. Journal of Personality and Social Psychology, 113(2), 244.

Fister-Gale, S. (2015). Forget millennials: Are you ready for Generation Z. Chief Learning Officer, 14(7), 38-48.

Forde, S., Mekler, E., & Opwis, K. (2016). Informational, but not Intrinsically Motivating Gamification': Preliminary Findings. Proceedings of the 2016 Annual Symposium on computer-human interaction in play companion extended abstracts (pp. 157–163). ACM.

Fox, N.J. (2008) Post-positivism. In: Given, L.M. (ed.) The SAGE Encyclopaedia of Qualitative Research Methods. London: Sage.

Fox, W. & Bayat, M.S. (2007) "A Guide to Managing Research" Juta Publications, p.45

Garbaya, S., Romano, D. M., & Hattar, G. (2019). Gamification of assembly planning in virtual environment. Assembly Automation.

Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. MIS quarterly, 389-400.

Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. Simulation & Gaming, 41(2), 238-259.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. Journal of personality and social psychology, 85(2), 197.

Griffiths, M. D., Davies, M. N., & Chappell, D. (2004). Demographic factors and playing variables in online computer gaming. CyberPsychology & behavior, 7(4), 479-487.

Goasduff, L., & Pettey, C. (2011). Gartner says by 2015, more than 50 percent of organizations that manage innovation processes will gamify those processes. Retrieved May, 10, 2020.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology?. In Handbook on ontologies (pp. 1-17). Springer, Berlin, Heidelberg.

Hallebone, E. & Priest, J. (2009) "Business and Management Research: Paradigms and Practices" Palgrave Macmillan

Hamari, J., & Tuunanen, J. (2014). Player types: A meta-synthesis.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In 2014 47th Hawaii international conference on system sciences (pp. 3025-3034). Ieee.

Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. Computers in Human Behavior, 71, 469-478.

Hendrikx, M., Meijer, S., Van Der Velden, J., & Iosup, A. (2013). Procedural content generation for games: A survey. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 9(1), 1-22.

Hofacker, C. F., De Ruyter, K., Lurie, N. H., Manchanda, P., & Donaldson, J. (2016). Gamification and mobile marketing effectiveness. Journal of Interactive Marketing, 34, 25-36.

Hoffman, L. W. (1972). Early childhood experiences and women's achievement motives. Journal of Social Issues, 28(2), 129-155.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004, July). MDA: A formal approach to game design and game research. In Proceedings of the AAAI Workshop on Challenges in Game AI (Vol. 4, No. 1, p. 1722).

Huotari, K., & Hamari, J. (2012, October). Defining gamification: a service marketing perspective. In Proceeding of the 16th international academic MindTrek conference (pp. 17-22).

Huotari, K., & Hamari, J. (2017). A definition for gamification: anchoring gamification in the service marketing literature. Electronic Markets, 27(1), 21-31.

Jagušt, T., Botički, I., & So, H. J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. Computers & education, 125, 444-457.

Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. Internet interventions, 6, 89-106.

Jurs, S. G., & Glass, G. V. (1971). The effect of experimental mortality on the internal and external validity of the randomized comparative experiment. The Journal of Experimental Education, 40(1), 62-66.

Kafai, Y. B. (1998). Video game designs by girls and boys: Variability and consistency of gender differences. From Barbie to Mortal Kombat: gender and computer games, 90-114.

Kapp, K. M. (2012). The gamification of learning and instruction: game-based methods and strategies for training and education. John Wiley & Sons.

Kapp, K. (2014). Gamification: Separating fact from fiction. Chief Learning Officer, 13(3), 45-52.

Kasurinen, J. & Knutas, A. (2018): Publication trends in gamification: A systematic mapping study. Computer Science Review 27 (2018), p. 33–44.

Kempthorne, O. (1961). The design and analysis of experiments with some reference to educational research. In Research design and analysis: Second annual Phi Delta Kappa symposium on educational research (pp. 97-126). Bloomington, Ind: Phi Delta Kappa (Eighth St. and Union Ave.).

Kerlinger, F. N. (1973). Foundation of behavioral science. New York: Holt, Renehard and Winston.

Khaleel, F. L., Ashaari, N. S., Wook, T. S. M. T., & Ismail, A. (2017, November). Gamification-based learning framework for a programming course. In 2017 6th International Conference on Electrical Engineering and Informatics (ICEEI) (pp. 1-6). IEEE.

Kim, H., Chung, J. Y., & Lee, M. (2020). Consumer's response to time restrictions: role of construal level. Journal of Marketing Communications, 26(2), 207-228.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. Computers in Human Behavior, 35, 179-188.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. International Journal of Information Management, 45, 191-210.

Kuo, M. S., & Chuang, T. Y. (2016). How gamification motivates visits and engagement for online academic dissemination–An empirical study. Computers in Human Behavior, 55, 16-27.

Krause, M., Mogalle, M., Pohl, H., & Williams, J. J. (2015, March). A playful game changer: Fostering student retention in online education with social gamification. In Proceedings of the Second (2015) ACM conference on Learning@ Scale (pp. 95-102).

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. Simulation & gaming, 45(6), 752-768.

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. Simulation & Gaming, 49(3), 315-337.

Larson, K. (2020). Serious Games and Gamification in the Corporate Training Environment: a Literature Review. *TechTrends*, *64*(2), 319-328.

Limanond, T., Chookerd, S., & Roubtonglang, N. (2009). Effects of countdown timers on queue discharge characteristics of through movement at a signalized intersection. Transportation research part C: emerging technologies, 17(6), 662-671.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. The Sage handbook of qualitative research, 4, 97-128.

Long, K., Han, L. D., & Yang, Q. (2011). Effects of countdown timers on driver behavior after the yellow onset at Chinese intersections. Traffic injury prevention, 12(5), 538-544.

Marczewski, A. (2013, 03 11). What's the difference between Gamification and Serious Games? Retrieved on March 15th 2020 from Gamasutra: http://www.gamasutra.com/blogs/AndrzejMarczewski/20130311/188218/Whats_the_difference_between_Gamification_and_Serious_Games.php

Marsh, D., & Furlong, P. (2002). A skin not a sweater: Ontology and epistemology in political science. Theory and methods in political science, 2, 17-41.

Matallaoui, A., Koivisto, J., Hamari, J., & Zarnekow, R. (2017, January). How effective is "exergamification"? A systematic review on the effectiveness of gamification features in exergames. In Proceedings of the 50th Hawaii International Conference on System Sciences.

Matenga, A., Crow, T., Walmsley, M., Luxton-Reilly, A., & Wuensche, B. (2018, April). Evaluation of the Implementation of a Timer in Gamified Programming Exercises. In 2018 International Conference on Learning and Teaching in Computing and Engineering (LaTICE) (pp. 36-41).

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of personality, 60(2), 175-215.

McGonigal, J. (2011). Reality is broken: Why games make us better and how they can change the world. Penguin.

Mekler, E.D., Brühlmann, F., Tuch, A.N. & Opwis, K. (2017): Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Computers in Human Behavior 71 (2017) p. 525-534.

Minton, H. L., & Schneider, F. W. (1980). Differential Psychology Waveland Press. Prospect Heights, IL.

Mitchell, R., Schuster, L., & Jin, H. S. (2018). Gamification and the impact of extrinsic motivation on needs satisfaction: Making work fun?. Journal of Business Research.

Morschheuser, B., Hamari, J., & Koivisto, J. (2016, January). Gamification in crowdsourcing: a review. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 4375-4384). IEEE.

Myers, L., & Sirois, M. J. (2004). Spearman correlation coefficients, differences between. Encyclopedia of statistical sciences, 12.

Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. Tutorials in quantitative Methods for Psychology, 4(1), 13-20.

Nah F.FH., Zeng Q., Telaprolu V.R., Ayyappa A.P., Eschenbrenner B. (2014) Gamification of Education: A Review of Literature. In: Nah F.FH. (eds) HCI in Business. HCIB 2014. Lecture Notes in Computer Science, vol 8527. Springer, Cham

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In Flow and the foundations of positive psychology (pp. 239-263). Springer, Dordrecht.

Nelson, M. J. (2012, October). Soviet and American precursors to the gamification of work. In Proceeding of the 16th international academic MindTrek conference (pp. 23-26).

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology, 2(2), 175-220.

O'Donovan, S., Gain, J., & Marais, P. (2013, October). A case study in the gamification of a university-level games development course. In Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference (pp. 242-251).

Olsson, M., Mozelius, P., & Collin, J. (2015). Visualisation and Gamification of e-Learning and Programming Education. Electronic journal of e-learning, 13(6), 441-454.

Oyibo, K., Orji, R., & Vassileva, J. (2017, April). Investigation of the Persuasiveness of Social Influence in Persuasive Technology and the Effect of Age and Gender. In PPT@ PERSUASIVE (pp. 32-44).

Pearl, J. (2000). Causality: Models, reasoning and inference. Cambridge, UK: Cambridge University

Pettey, C., & van der Meulen, R. (2012). Gartner says by 2014, 80 percent of current gamified applications will fail to meet business objectives primarily due to poor design [Press Release]. Retrieved May 10, 2020.

Pettit, R. K., McCoy, L., Kinney, M., & Schwartz, F. N. (2015). Student perceptions of gamified audience response system interactions in large group lectures and via lecture capture technology. BMC medical education, 15(1), 92.

Phillips, B. S. (1966). Social research: Strategy and tactics (Vol. 966). New York: Macmillan.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior research methods, 40(3), 879-891.

Popper, K. R. (1963). Science as falsification. Conjectures and refutations, 1, 33-39.

Pyky, R., Koivumaa-Honkanen, H., Leinonen, A.-M., Ahola, R. Hirvonen, N., Ikaheimo, T. (2017). Effect of tailored, gamified, mobile physical activity intervention on life satisfaction and self-rated health in young adolescent men: A population-based, randomized controlled trial (MOPO study). Computers in Human Behavior, 72, 13-22.

Raddon, A. (2010). Early stage research training: Epistemology & ontology in social science research. Generic Skills Training for Research Students, 1-14.

Raymer, R. (2011). Gamification: using game mechanics to enhance eLearning. ELearn, 2011(9).

Rehn, A. (2016). Gamification as Ideological Praxis: On Play, Games and the Modding of Management. In The Business of Gamification (pp. 62-78). Routledge.

Reeves, B., & Read, J. L. (2009). Total engagement: How games and virtual worlds are changing the way people work and businesses compete. Harvard Business Press.

Richards, C., Thompson, C., & Graham, N. (2014). Beyond designing for motivation: the importance of context in gamification. Proceedings of the first ACM SIGCHI annual symposium on computer-human interaction in play (pp. 217–226). ACM.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. Behavioral Ecology, 17(4), 688-690.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68-78.

Sailer, M., Hense J-U., Mayr S.K. & Mandl, H. (2017): How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior 69 (2017) p. 371-380.

Sanmugam, M., Zaid, N. M., Abdullah, Z., Aris, B., Mohamed, H., & van der Meijden, H. (2016, December). The impacts of infusing game elements and gamification in learning. In 2016 IEEE 8th international conference on engineering education (ICEED) (pp. 131-136). IEEE.

Saunders, M., Lewis, P., & Thornhill, A. (2009). Understanding research philosophies and approaches. Research methods for business students, 4(106-135).

Saunders, M., Lewis, P. & Thornhill, A. (2012) "Research Methods for Business Students" 6th edition, Pearson Education LimitedVan den Bergh, J., & Behrer, M. (2016). How cool brands stay hot: Branding to Generations Y and Z. Kogan Page Publishers.

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

Seemiller, C., & Grace, M. (2016). Generation Z goes to college. John Wiley & Sons.

Skinner, H., Sarpong, D., & White, G. R. (2018). Meeting the needs of the Millennials and Generation Z: gamification in tourism through geocaching. Journal of Tourism Futures.

Škuta, P., & Kostolányová, K. (2016). The Inclusion of Gamification Elements in the Educational Process. DIVAI 2016.

Slack, M. K., & Draugalis Jr, J. R. (2001). Establishing the internal and external validity of experimental studies. American Journal of Health-System Pharmacy, 58(22), 2173-2181.

Smith, B. (2012). Ontology. In The furniture of the world (pp. 47-68). Brill Rodopi.

Spearman, C. (1904). Measurement of association, Part II. Correction of 'systematic deviations'. Am J Psychol, 15, 88-101.

Spence, J. T., & Helmreich, R. L. (1980). Masculine instrumentality and feminine expressiveness: Their relationships with sex role attitudes and behaviors. Psychology of Women Quarterly, 5(2), 147-163.

Stanculescu, L. C., Bozzon, A., Sips, R.-J., & Houben, G.-J. (2016). Work and play: An experiment in enterprise gamification. In CSCW '16 (pp. 346-358). New York, NY: ACM.

Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. Journal of Computer Assisted Learning, 31(3), 268-286.

Sun, H., & Zhang, P. (2006). The role of moderating factors in user technology acceptance. International journal of human-computer studies, 64(2), 53-78.

Sun, J. C. Y., & Hsieh, P. H. (2018). Application of a gamified interactive response system to enhance the intrinsic and extrinsic motivation, student engagement, and attention of English learners. Journal of Educational Technology & Society, 21(3), 104-116.

Thom, J., Millen, D., & DiMicco, J. (2012, February). Removing gamification from an enterprise SNS. In Proceedings of the acm 2012 conference on computer supported cooperative work (pp. 1067-1070).

Thyer, B. (2009). The handbook of social work research methods. Sage Publications.

Tsay, C. H. H., Kofinas, A., & Luo, J. (2018). Enhancing student learning experience with technology-mediated gamification: An empirical study. Computers & Education, 121, 1-17.

Twenge, J. M. (2017). iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy--and completely unprepared for adulthood--and what that means for the rest of us. Simon and Schuster.

Van den Bergh, J., & Behrer, M. (2016). How cool brands stay hot: Branding to Generations Y and Z. Kogan Page Publishers.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. EDUCAUSE review, 41(2), 16.

Van Wyk, B. (2012). Research design and methods Part I. University of Western Cape.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS quarterly, 115-139.

Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. Journal of computer-mediated communication, 13(4), 993-1018.

Williams, D., Consalvo, M., Caplan, S., & Yee, N. (2009). Looking for gender: Gender roles and behaviors among online gamers. Journal of communication, 59(4), 700-725.

Werbach, K., & Hunter, D. (2012). For the win: How game thinking can revolutionize your business. Wharton Digital Press.

Xi, N. & Hamari, J. (2019): Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. International Journal of Information Management 46 (2019) p. 210 – 221.

Yee, N. (2006). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. Presence: Teleoperators and virtual environments, 15(3), 309-329.

Zhang, M. W., Ying, J. B., Song, G., & Ho, R. (2018). A review of gamification approaches in commercial cognitive bias modification gaming applications. Technology and Health Care, 26(6), 933-944.

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

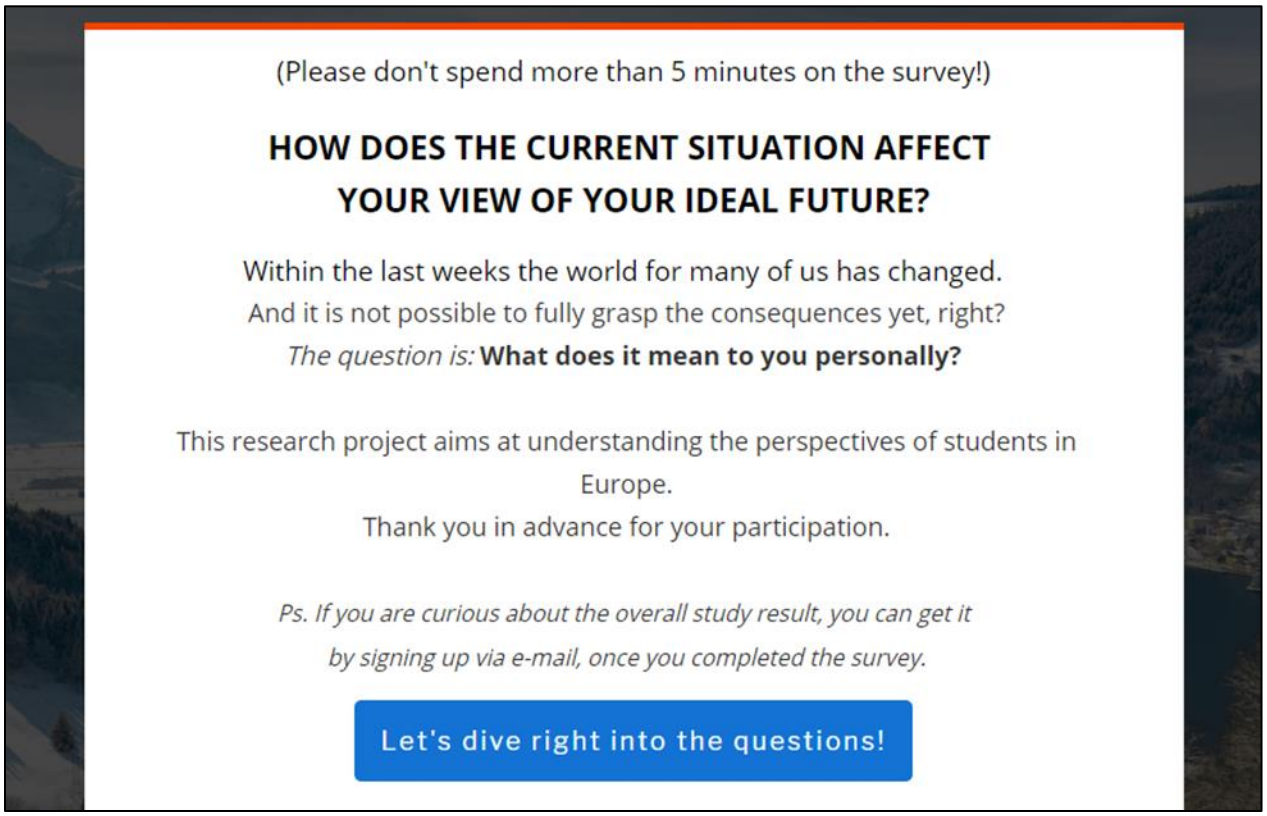
Zikmund, W.G., Babin, J., Carr, J. & Griffin, M. (2012) “Business Research Methods: with Qualtrics Printed Access Card” Cengage Learning

APPENDICES

APPENDIX A

Landing Page:

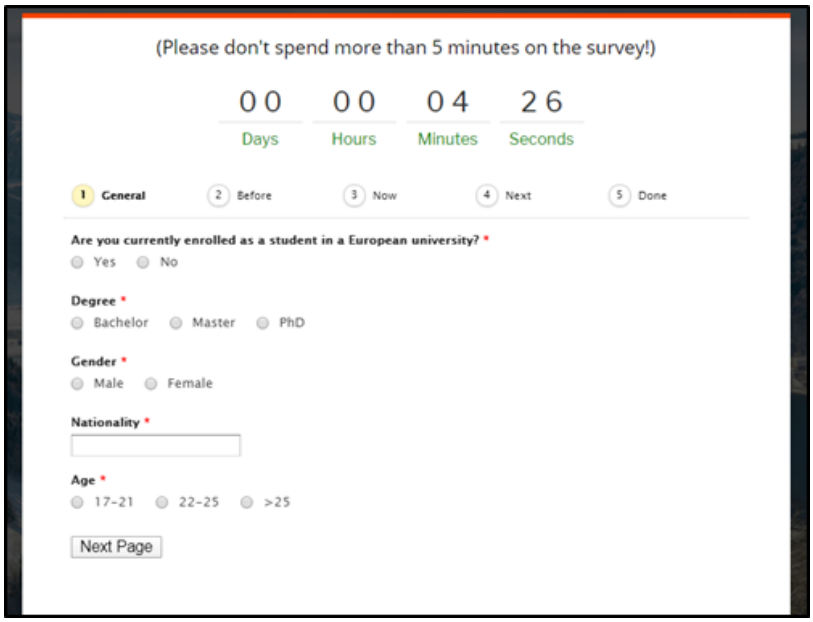
- same for treatment (website ending “lp”) and control group (website ending “gen”)
- to start the experiment, participants had to click on the blue button

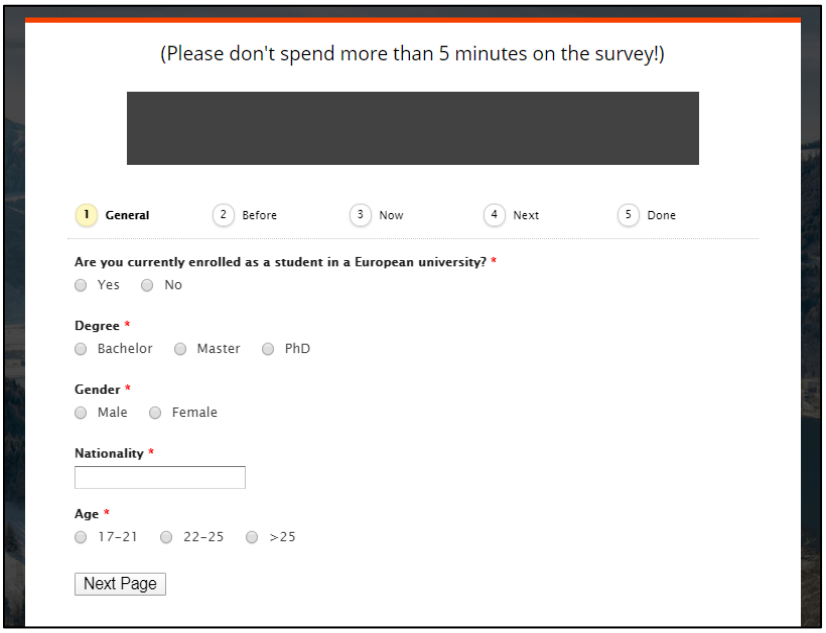


APPENDIX B

First page of the questionnaire:

- treatment group with countdown-timer (top figure), control group with black box instead (lower figure)
- general information; completion = level 1 of engagement

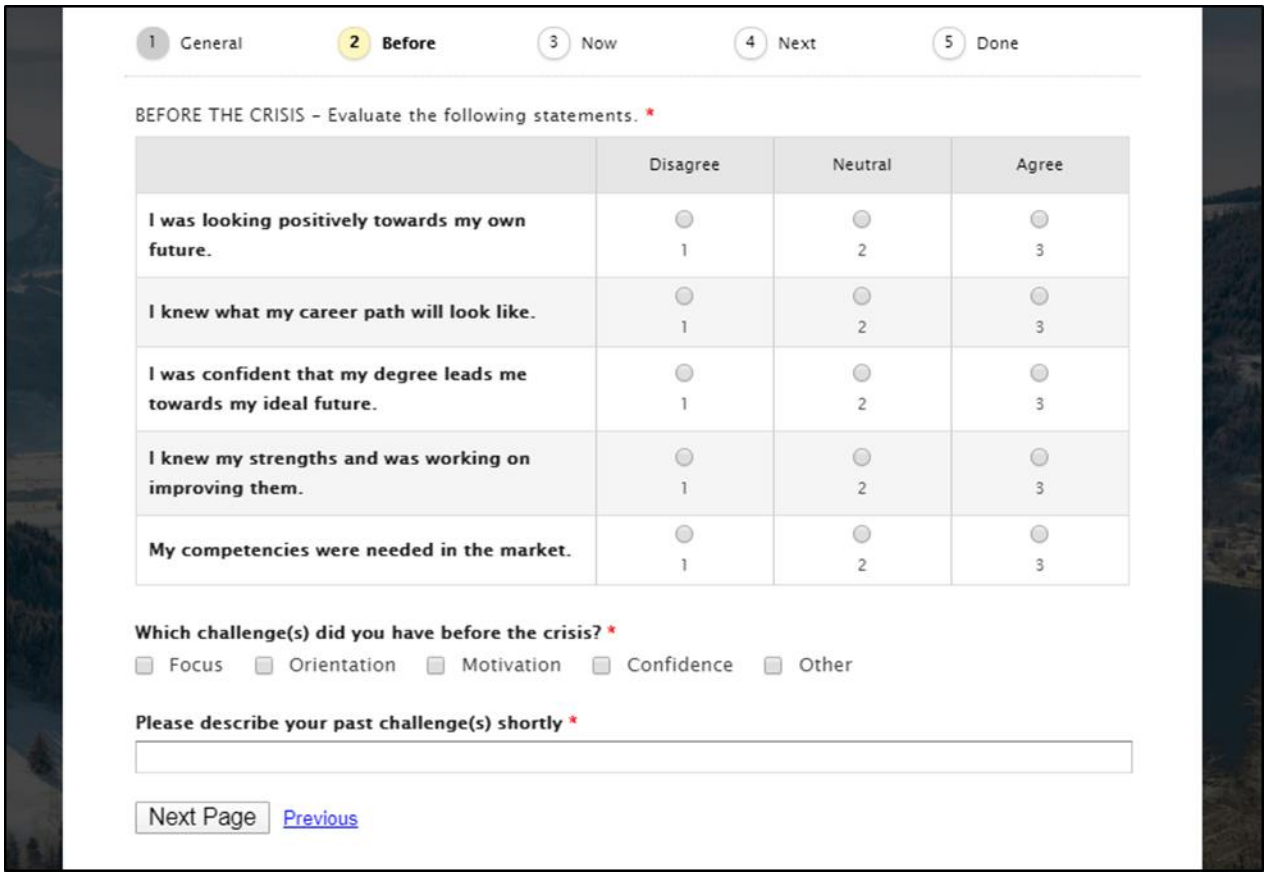




APPENDIX C

Second page of the questionnaire:

- 3-item Likert-scale, checkboxes, text field (evaluating ideal future before current crisis, not relevant for measurement of engagement)
- Completion = level 2 of engagement



1 General 2 **Before** 3 Now 4 Next 5 Done

BEFORE THE CRISIS – Evaluate the following statements. *

	Disagree	Neutral	Agree
I was looking positively towards my own future.	1	2	3
I knew what my career path will look like.	1	2	3
I was confident that my degree leads me towards my ideal future.	1	2	3
I knew my strengths and was working on improving them.	1	2	3
My competencies were needed in the market.	1	2	3

Which challenge(s) did you have before the crisis? *

Focus Orientation Motivation Confidence Other

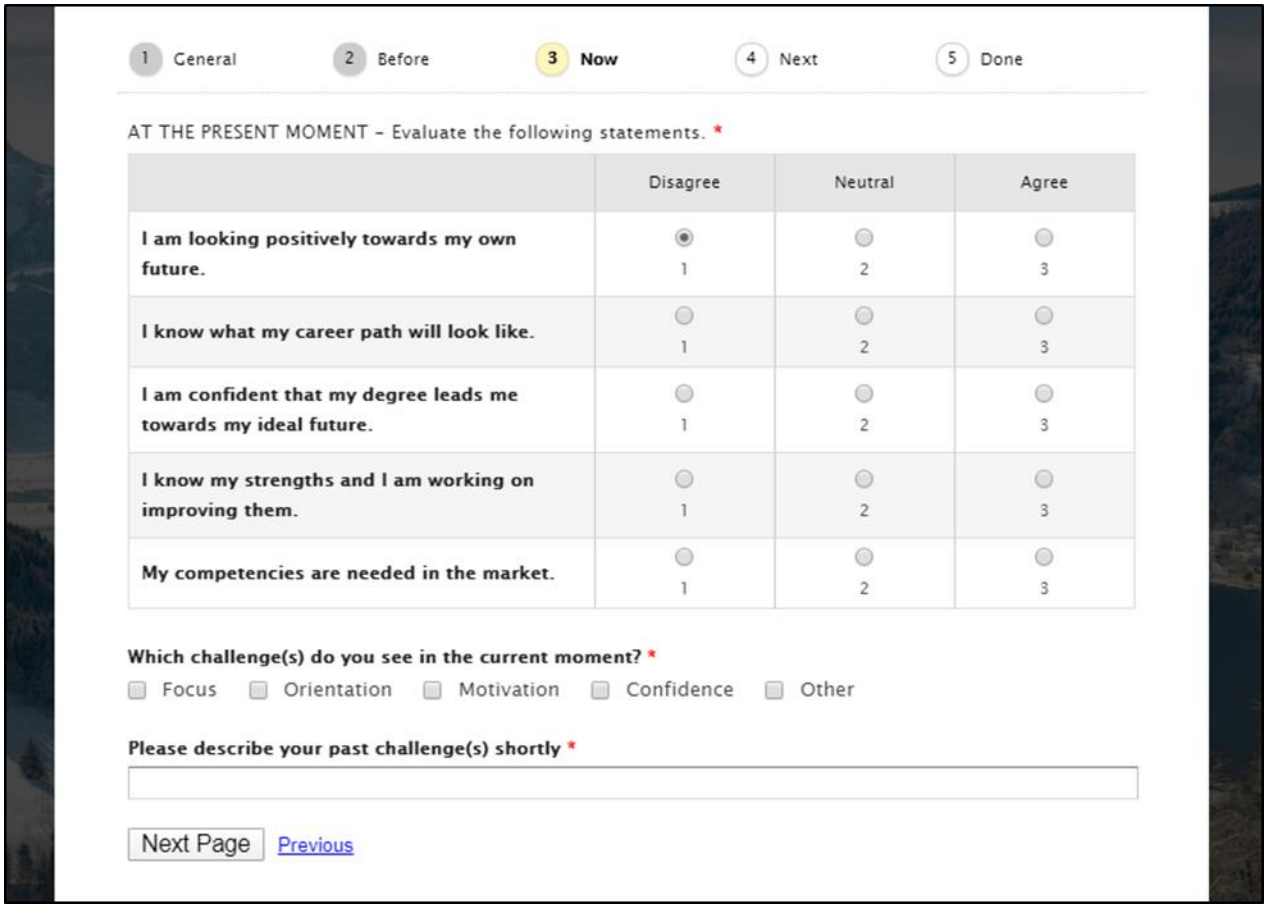
Please describe your past challenge(s) shortly *

Next Page Previous

APPENDIX D

Third page of the questionnaire:

- Identical 3-item Likert-scale, checkboxes, text field (now evaluating the current situation)
- Completion = level 3 of engagement



1 General 2 Before 3 **Now** 4 Next 5 Done

AT THE PRESENT MOMENT – Evaluate the following statements. *

	Disagree	Neutral	Agree
I am looking positively towards my own future.	1	2	3
I know what my career path will look like.	1	2	3
I am confident that my degree leads me towards my ideal future.	1	2	3
I know my strengths and I am working on improving them.	1	2	3
My competencies are needed in the market.	1	2	3

Which challenge(s) do you see in the current moment? *

Focus Orientation Motivation Confidence Other

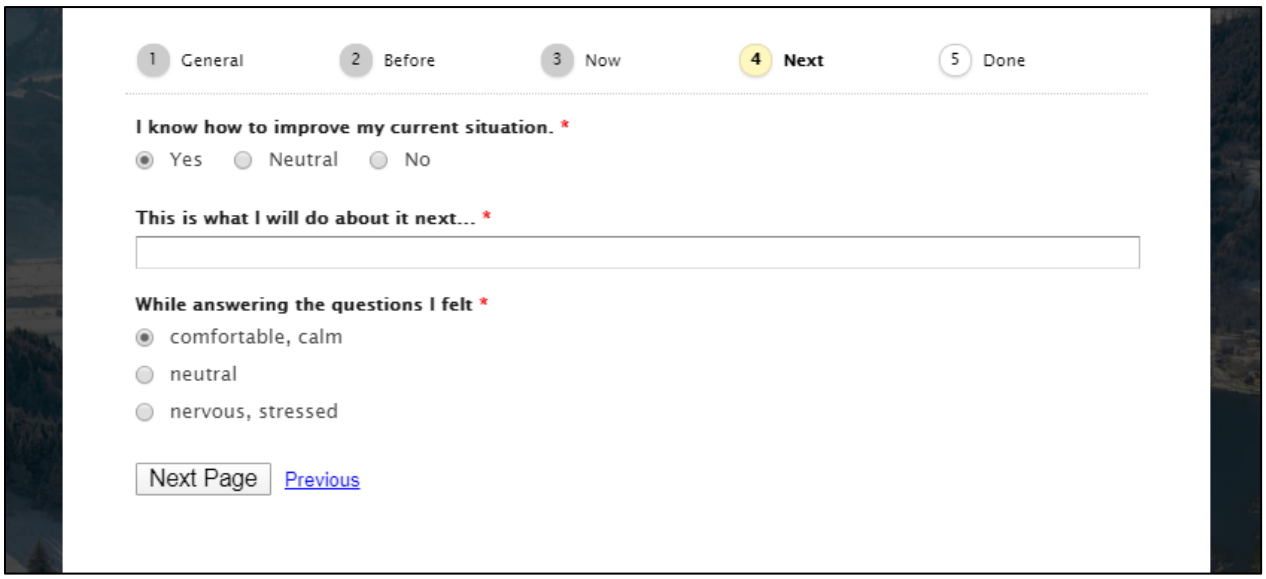
Please describe your past challenge(s) shortly *

Next Page Previous

APPENDIX E

Fourth page of the questionnaire:

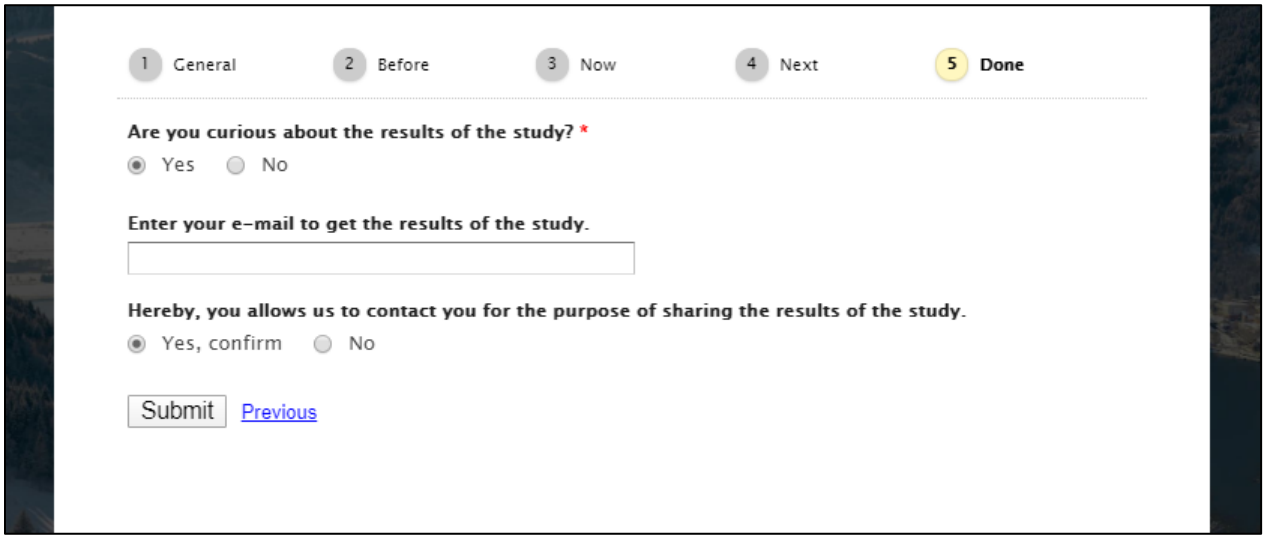
- Multiple choice, checkboxes, text field
- Completion = level 4 of engagement



APPENDIX F:

Fifth and last page of the questionnaire:

- Multiple choice, text field
- Completion = level 5 of engagement
- Signup with valid e-mail address = level 6 of engagement

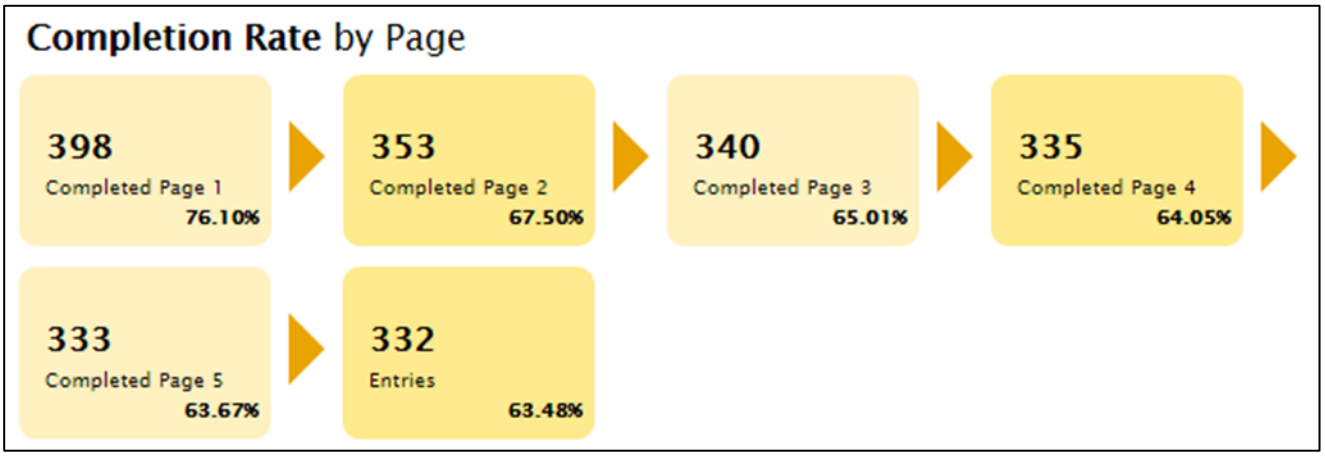


APPENDIX G:

Message of appreciation after completing the survey or after countdown-timer run out

APPENDIX H

Completion Rate of Control Group



Completion Rate of Treatment Group