

Article

Leaderboard Positions and Stress—Experimental Investigations into an Element of Gamification

Marc Schlömmner , Teresa Spieß * and Stephan Schlögl 

MCI—The Entrepreneurial School, 6020 Innsbruck, Austria; marc.schloemmer@outlook.com (M.S.); stephan.schloegl@mci.edu (S.S.)

* Correspondence: teresa.spiess@mci.edu

Abstract: Gamification, i.e., the use of game elements in non-game contexts, aims to increase peoples' motivation and productivity in professional settings. While previous work has shown both positive as well as negative effects of gamification, there have been barely any studies so far that investigate the impact different gamification elements may have on perceived stress. The aim of the experimental study presented in this paper was thus to explore the relationship between (1) leaderboards, a gamification element which exchanges and compares results, (2) heart rate variability (HRV), used as a relatively objective measure for stress, and (3) task performance. We used a coordinative smartphone game, a manipulated web-based leaderboard, and a heart rate monitor (chest strap) to investigate respective effects. A total of $n = 34$ test subjects participated in the experiment. They were split into two equally sized groups so as to measure the effect of the manipulated leaderboard positions. Results show no significant relationship between the measured HRV and leaderboard positions. Neither did we find a significant link between the measured HRV and subjects' task performance. We may thus argue that our experiment did not yield sufficient evidence to support the assumption that leaderboard positions increase perceived stress and that such may negatively influence task performance.

Keywords: gamification; leaderboard positions; heart rate variability; stress; experimental study



Citation: Schlömmner, M.; Spieß, T.; Schlögl, S. Leaderboard Positions and Stress—Experimental Investigations into an Element of Gamification. *Sustainability* **2021**, *13*, 6608. <https://doi.org/10.3390/su13126608>

Academic Editors: Muhamed Turkanović and Marjan Heričko

Received: 9 April 2021
Accepted: 5 June 2021
Published: 10 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gamification, the use of game elements outside a game context ([1], p. 2), enjoys an evergrowing interest in the workplace as well as in educational contexts. The main goal of gamification is to improve a process through playful experiences in such a way that the personal added value of those carrying it out increases ([2], p. 19). Furthermore, it has shown to have positive effects on motivation and productivity [3]. From a negative perspective, however, aspects such as off-task behavior [4,5], negative feedback, as well as physical and psychological damage [6] have also been linked to gamification.

Game elements are different components that are used in games ([1], p. 3). Reeves and Read [7] define the ten most common game elements, highlighting examples such as *time pressure*, *feedback*, *stories*, reputation in the form of *levels*, *awards*, *rankings* and *competition* according to set rules. Dale uses the term game mechanics for game elements and also lists *luck* and the *exchange of results* within a community as the most frequently used game mechanics ([3], p. 82). The effects of gamification are highly dependent on both the environment used and the users of the system ([8], p. 3027), where it is important to have a group of users who are pursuing the same goal ([9], p. 8). Yet, if gamification is applied correctly, it offers a captivating experience accompanied by a noticeable learning effect ([3], p. 85).

1.1. Recent Work on Gamification

In a meta-analysis of gamification in learning settings, Sailer & Homner found a small effect of gamification on motivational learning outcomes [10]. A more recent study

($n = 205$) further showed a connection between the use of gamification and the subsequent application of knowledge, which was moderated by students' learning process performance [11]. Zainuddin and colleagues, on the other hand, found gamification to be a promising and effective alternative to formative assessments [12]. To this end, however, it was also shown that the effects of gamification often do not last and that lower performing students are less likely to benefit from its use [13]. Yet, it seems to enhance a learner's user experience and thus may improve more traditional educational contexts [14]. Finally, Groening & Binnewies' work showed that achievements as an element of gamification may foster motivation and subsequently impact positively on performance [15].

1.2. The Leaderboard Element

One specific element of gamification are so-called *leaderboards*. Leaderboards are tools which can be used to exchange and compare the results within a community [3,16]. They serve as a feedback mechanism for social competition and may promote the engagement of participants. They show an overall ranking and are regularly updated so that players can compete for higher placements ([17], pp. 94–95). In contrast to other game elements such as badges or levels, leaderboards allow immediate and direct comparisons [18]. However, they can be harmful for users with poor performance ([4], p. 179). That is, if users feel forced into a competitive situation, this may have negative effects on skills and attitudes, or as Lopez puts it: "*Being consistently at the bottom of the ranking list is negative feedback.*" (online: <https://www.latimes.com/health/la-xpm-2011-oct-19-la-me-1019-lopez-disney-20111018-story.html> (accessed on 26 March 2021)). Users who feel that they will never be at the top are inclined to let go of their ambition ([19], p. 565). Hence, leaderboard-triggered transparency may not always be the ideal solution, as users can easily feel weakened compared to better placed persons ([6], p. 166), and consequently may encounter higher levels of stress.

1.3. Gamification and Stress

Jamal ([20], p. 728) defines stress as the reaction to elements of the environment which appear threatening. Stress factors can for example result from the tasks to be carried out, or the environment itself ([21], p. 5). Abbe's [21] model assumes that subjective stress can lead to conditions such as anxiety, hostility, and depression. These conditions have a negative effect on task performance. It is also assumed that subjective stress is caused by events occurring during task performance. The more frequent and more intense the events are for the individual, the greater the subjectively perceived level of stress. The environment determines, in part, the frequency with which these events occur. Individual characteristics such as prior experience, Type-A behavioral patterns and fear of negative evaluation determine the frequency and intensity of the stress burden on individuals ([22], pp. 618–620). While stress can also be measured by self-report questionnaires [23,24], a person's heart rate variability (HRV) is a physical measure of stress, which may be used to more objectively measure the level of stress experienced ([25], p. 236). Also Jobbágy et al. recommend HRV to assess the current stress level ([26], p. 238).

Although there is a significant body of work on the positive effects of gamification, little is known about the impact gamification elements may have on perceived stress. A recent study by Paniagua and colleagues e.g., found that gamification reduces stress levels and improves the academic performance of chemical engineering students [27]. Teenakoon & Wanninayake [28], on the other hand, highlighted a moderating effect on the relationship between work stress and work performance of non-managerial bank employees in Sri Lanka. Yet, to our knowledge, there have not been any investigations into the effects of leaderboards.

1.4. Research Question

The fact that (1) little is known regarding the effect of leaderboards on stress experience, and (2) that most previous studies concerning stress in gamification focused on self-report scales of stress, inspired the following research question guiding our investigations:

How is the leaderboard position related to (physical measures of) stress and changes in task performance?

Following we start our report with Section 2 discussing the *Physical Measures of Stress*. Next, Section 3 outlines the *Hypotheses, Materials and Procedure* we used for our study. Section 4 summarizes the gained *Results* and Section 5 provides a *Discussion* of the respective conclusions which may be drawn. Finally, Section 6 highlights the work's *Limitations* and Section 7 provides an overall *Summary and Future Recommendations*.

2. Physical Measures of Stress

In order to investigate the above outlined research question we designed an experimental setting. We focused on factors influencing task performance while controlling for other variables potentially influencing stress as well as task performance. The following outlines the used measurement variables and subsequently explains the research model.

2.1. Heart Rate Variability and Its Measurement

HRV was first studied by Hon and Lee [29] when they found that embryos were preceded by a change in the intervals between heartbeats before any appreciable change in heart rate occurred. For respective studies, Castaldo et al. ([30], pp. 376–377) propose the following:

- Define the length of the HRV measurements and the anticipated stressors to the best of available knowledge
- If at all possible, avoid physical activity
- Perform the analysis of the HRV values according to standardized guidelines
- Carry out the stress measurement according to the aim of the study, i.e., before, during or after the session

HRV indicators are determined by various differences between a series of ordered heartbeats ([25], p. 231). In order to record a series of heartbeats, a continuous measurement of the heart rate is necessary (ibid.). After the acquisition, the row must be corrected for abnormal impacts, so-called artifacts. An artifact is given if the interval between two heartbeats exceeds a certain limit of the measured median. The removal of artifacts must be carried out carefully, as even individual artifacts can have a significant influence on the analysis results [31]. Artifacts are of technical or physiological origin and can be caused by poorly attached measuring devices or by movement of the test subjects ([32], p. 1). An artifact can either be carried out manually by removing individual heartbeats or automatically using a software program ([33], p. 7). When the raw data is cleaned up by a software program, a limit value for the removal of artifacts must be determined ([31], p. 4). In the next step, the interbeat intervals of the heartbeats are determined ([25], p. 231). The interbeat intervals are defined as the time interval between the R-peaks, which reflects the contraction of the heart chambers. Since the intermediate beat intervals are determined by successive “normal” heartbeats, these intervals are often referred to as “normal to normal interval” (NN interval) or “R to R interval” (RR interval) (ibid.) ([34], p. 3). To evaluate measurement data recorded with a chest strap, Baumgartner et al. ([34], p. 3) recommend the Kubios HRV (online: <https://www.kubios.com/> (accessed on 26 March 2021)) software. It allows for threshold-based artifact corrections ([31], p. 4).

2.2. HRV Analysis Methods

The RR intervals can be analyzed using various methods [25,33,35]. Below we discuss two common analysis methods, i.e., the time domain analysis and the frequency analysis.

- **Time domain analysis**—The time domain analysis describes fluctuations in the RR intervals ([33], p. 2). In the time domain analysis, among other things, the square root of the mean value of consecutive RR interval differences (*RMSSD*), the number of pairs of adjacent NN intervals that differ by more than 50 ms (*NN50*), and that from the division percentage (*pNN50*) resulting from the *NN50* by the total number of NN intervals, are calculated [25,26]. Castaldo et al. carried out a systematic review of acute psychological stress assessment through HRV analysis. The reviewed studies showed that *pNN50* and *RMSSD* decreased during the stress measurement compared to the rest of the measurement [30].
- **Frequency analysis**—The frequency analysis calculates the distribution of the absolute and relative performance of the heart in different frequency bands ([33], p. 2). The intervals are divided into “very low frequency” (*VLF*), “low frequency” (*LF*) and “high frequency” (*HF*) [25,35]. The division of the relative performance values (i.e., $\frac{LF}{HF}$), allows for a direct comparison between people despite large differences in the absolute performance ([33], p. 2). There is an inverse correlation between the *HR* value and stress [30,33], where an increase in the *LF* values points to less stress ([30], p. 373). *HF* values are strongly influenced by artifacts, since R-peaks occurring in the data increase the power at higher frequencies ([31], p. 9).

2.3. Duration of HRV Measurements

The duration of *HRV* measurements are divided into three categories: 24-h, short-term and ultra-short-term measurements. Short-term measurements have a duration of approx. five minutes, with ultra-short-term measurements comprising all measurements shorter than five minutes ([33], p. 2). Although *HRV* values are traditionally calculated from five-minute to 24-h recordings, also ultra-short-term recordings can determine cardiac activity ([35], p. 355). Shaffer and colleagues recorded five-minute rest measurements from 38 students and correlated ultra-short-term measurements with the complete measurement. They found that measurements take 60 s for *RMSSD* and *pNN50*, 90 s for *LF* and 180 s for *HF* and $\frac{LF}{HF}$ ([36], p. 231). Here Nussinovitch et al. [37] compared ten seconds and one-minute rest recordings with five-minute recordings from 70 healthy test subjects, and found that ultra-short-term *RMSSD* measurements achieved acceptable correlations. When examining 467 volunteers, Baek, Cho, Cho & Woo recorded five-minute resting measurements. They found that *HF* 20 s, *RMSSD* 30 s, *pNN50* 60 s and *LF* 90 s measurements are needed in order to be able to make valid statements ([38], p. 413). In order to investigate ethnic and gender-specific differences in *HRV* measurements, Li et al. conducted three times 30 s measurements to infer *RMSSD* and *HF* [39]. To validate short-term measurements, Salahuddin, Cho, Jeong & Kim divided a 24-h measurement into 30-min intervals. These half-hour intervals were divided into three ten-minute measurements, from which 10 to 150 s measurements were taken at random. They found that with 10 s *RMSSD* and with 20 s *pNN50*, *HF* and $\frac{LF}{HF}$ conclusions about acute psychological stress can be drawn ([40], p. 4658).

2.4. AI-Driven Physiological Measures

While the above describes *HRV* as a previously used and commonly accepted measure of stress, recent work in artificial intelligence and Big Data has led to alternative approaches for measuring this and other types physiological data streams. Massaro et al., for example, presented an AI-driven diagnostics platform connected to a wearable device to predict individual physiological data [41,42]. Kim et al., on the other hand, used facial image thresholding to recognize emotions, from which one may also deduce individual stress levels [43]. Finally, Gonzalez-Viejo and colleagues demonstrated non-contact heart rate and blood pressure measures using the photoplethysmography technique and machine learning [44].

3. Hypotheses, Measures and Procedure

Following we outline the assumed hypotheses for our investigations (subdivided into core and control hypotheses), present the used measures and explain our procedure.

3.1. Core Hypotheses: Leaderboard, Stress and Task Performance

Lopez (online: <https://www.latimes.com/health/la-xpm-2011-oct-19-la-me-1019-lopez-disney-20111018-story.html> (accessed on 26 March 2021)) examined the use of digital leaderboards in Disneyland hotels in Anaheim, California. There digital leaderboards were implemented to compare the domestic staff with one another. In some cases, this created such fear and shame among workers that some even skipped toilet breaks because of the fear of losing their jobs. Lazarus (1990, p. 3) explains that stress—besides individual triggers—needs environmental triggers in order to occur [45]. As shown by Lopez (ibid.) the mere introduction of a leaderboard can actually be interpreted as an environmental stressor. Thus, used in an educational context, students may even find the mere idea of a leaderboard introduction stressful—while especially being placed in lower ranks of such a leaderboard might also have an impact on academic performance [46]. However, leaderboards can also encourage people to compete for higher placements, and thus have a positive effect on task performance ([17], p. 95). Consequently we may deduce the following hypotheses:

Hypothesis 1 (H1). *The use of a leaderboard leads to increased levels of stress.*

Hypothesis 2 (H2). *Individuals experiencing higher levels of stress show lower task performance.*

Hypothesis 3 (H3). *The use of a leaderboard leads to higher task performance.*

In order to enhance gaming experience, Bowey et al. manipulated the leaderboard ranking. They found that an emphasis of the best and worst placements by color intensified the felt experience ([16], p. 116). However, forcing users into such competitive situations can also have negative effects on skills and attitudes. For example, employees who feel that they will never be at the top may be inclined to abandon their work ([19], p. 565). Hence, it is actually sometimes recommended to avoid displaying the lowest leaderboard ranks ([47], p. 1958) as such may be harmful for users with poor performance ([4], p. 179). Furthermore, Lazarus et al. found that subjects who perform well on the first test tend not to improve, whereas subjects who perform bad tend to improve on a second test ([48], pp. 299–300). Hjortskov et al. on the other hand claim that psychological stress reduces performance ([49], pp. 87–88). Hence, we may argue that:

Hypothesis 4 (H4). *Individuals ranked among the ‘Bottom 3’ experience more stress than individuals ranked among the ‘Top 3’.*

Hypothesis 5 (H5). *Individuals ranked among the ‘Bottom 3’ show a higher increase in task performance than individuals ranked among the ‘Top 3’.*

3.2. Control Hypotheses: Individual Behavior and Personality

One area of stress research focuses on the understanding of stress-related behavior, which can also be referred to as the Type-A behavior pattern ([20], p. 728). Perlman and colleagues ([50], p. 6) define the Type-A behavior pattern as an intense pursuit of performance, easily provoked hostility and impatience in combination with excessive competitive pressure and a permanent feeling of time pressure. Individuals not showing these characteristics, are likely to possess a Type-B behavior pattern ([20], pp. 728–729). Type-A people act in a way that causes more stressful events for them and they also experience these events as more stressful ([22], p. 620). Using a sample of 215 nurses, Jamal examined whether there is a connection between workplace stress, workplace stressors

and Type-A behavioral patterns with job satisfaction, commitment and fluctuation. It was found that Type-A nurses experience significantly more stress and strain at work than Type-B nurses [20]. Hence, we may argue:

Hypothesis 6 (H6). *Type-A behavior increases stress.*

Individuals who are afraid of negative evaluations are more likely to experience higher levels of stress [22,51]. Fear of negative evaluation is one of the most widely used scales to measure social phobias ([52], p. 982). Carleton, McCreary, Norton & Asmundson define fear of negative evaluation as a scale for measuring fears and anxieties that arise from the fear of being judged negatively by others ([53], p. 297). Motowidlo et al. [22] and Packard & Motowidlo [51] found that nurses with a high fear of negative evaluation are more likely to experience high levels of stress. Consequently, we may argue:

Hypothesis 7 (H7). *Fear of negative evaluation increases stress.*

The Big Five model is used to describe personality ([54], p. 7). The model distinguishes between five dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness [54,55]. A higher value for extraversion stands for characteristics such as sociability, talkability and assertiveness, whereas a lower value stands for characteristics such as being quiet and withdrawn. Furthermore, people with a low score in the dimension of agreeableness can be described as cool, critical and suspicious. If it is high, there is talk of high interpersonal trust, cooperativity and compliance. Conscientiousness, on the other hand, distinguishes determined, disciplined and reliable people from negligent, inconsistent and indifferent people. Neuroticism describes a person's emotional behavior. If a low value is achieved, this speaks for serenity and relaxation. A high value means uncertainty, nervousness and fear. People with a high degree of openness are considered imaginative, inquisitive and intellectual. A low score stands for firm views, conservatism and traditionalism (ibid.).

Ebstrup et al. examined the relationship between personality types and perceived stress on the basis of 3471 individuals. They found a significant negative relationship between perceived stress and extraversion, conscientiousness, as well as agreeableness. A significant positive relationship was found between perceived stress and neuroticism. Neuroticism had the greatest impact on perceived stress, followed by conscientiousness, extraversion, and agreeableness. Openness did not show a significant relationship to perceived stress ([56], pp. 414–416). Thus, we hypothesise:

Hypothesis 8 (H8). *High levels of extraversion reduce stress.*

Hypothesis 9 (H9). *High levels of conscientiousness reduce stress.*

Hypothesis 10 (H10). *High levels of agreeableness reduce stress.*

Hypothesis 11 (H11). *High levels of neuroticism increase stress.*

A positive, significant correlation with stress was demonstrated for Negative Affect (NA) [57–59]. The affect is the experience of feelings. By dividing it into positive and negative affect, it is used to investigate sensations and feelings ([60], p. 6). Both factors can be measured either as traits, i.e., general persistent properties, or as current states. Positive Affect (PA) reflects the extent to which a person feels active, excited, and in tune with the environment. High PA can be described with terms related to high energy, mental alertness, high concentration and determination. Lower PA stands for sadness and indolence [57–59]. Negative Affect (NA) is a factor of subjective distress and describes a wide range of aversive mood states, including despair, fear, guilt, and nervousness. Low NA indicates a state of calmness and serenity ([59], p. 1063).

Watson examined the relationship between affect and health complaints, perceived stress and daily activities. A positive significant correlation between *NA* and stress could be demonstrated in 75% of the test subjects. The relationship between stress and *NA* was stronger than the relationship between stress and *PA*. Consequently Watson concluded that the *PA* scale is related to social activities and the *NA* scale is significantly related to perceived stress ([58], p. 1024). This is also in line with the results of a study by Kanner et al., who found that stress levels are significantly related to affect, especially the *NA* scale ([57], p. 21). Hence, finally we may argue that:

Hypothesis 12 (H12). *Negative Affect (NA) increases stress.*

An overview of the different variables and their respective connection is depicted by Figure 1. Note: Our investigations did not aim to confirm or disprove this model based on Motowidlo et al. ([22], p. 619). We merely show it because we believe for it to nicely outline the connection between construct variables and thus help in understanding potential influences.

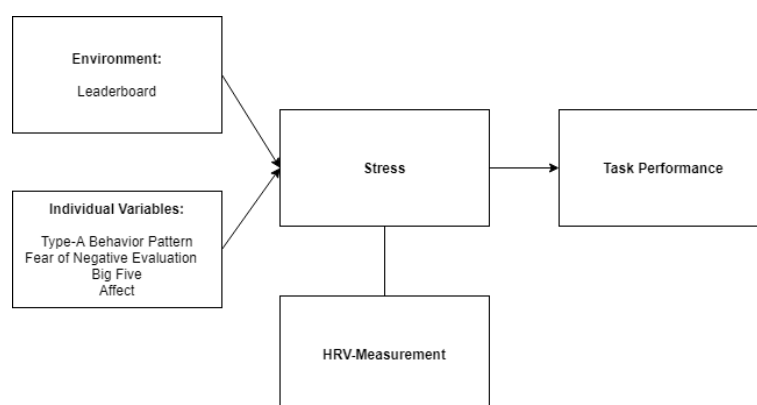


Figure 1. Research Model based on Motowidlo et al. ([22], p. 619).

3.3. Measures

Leaderboard

We designed a leaderboard using common Internet technologies (i.e., HTML, CSS and JavaScript). Based on the findings of Bowey et al. ([16], p. 116), the best and worst rankings on the leaderboard were coloured in order to reinforce participants' experienced feelings. Depending on participants' achieved score, the other scores displayed were calculated randomly. Furthermore, the exact placement, which was dependent on the assigned group, was randomly displayed in the respective colour range. With this participant-dependent presentation of the scores, an attempt was made to present the manipulated results as realistically as possible, which according to Lazarus et al. is a key point to consider when manipulating results [48]. Figure A1 in Appendix A shows the manipulated leaderboard for a top ranking. In this example, the second place was randomly assigned. Based on the achieved score of 134, which was entered in a pop-up window before the leaderboard was displayed (cf. Figure A2 in Appendix A). The scores for the remaining placements were randomly calculated.

Type-A Behavior Pattern

The Framingham Type-A Scale was used to measure the Type-A behaviour pattern [61]. The scale contains ten questions about personal characteristics and feelings experienced at the end of a productive day. The questions cover topics such as competitive pressure and feelings of time urgency. The first five questions use a Likert scale containing the following four answer options: *Very Good* (1.00), *Quite Good* (0.67), *Somewhat* (0.33) and *Not at all* (0.00). Questions six to ten use a binary scale, i.e., *Yes* (1.00) and *No* (0.00) (please

refer to Figure A6 in Appendix B for a copy of the complete scale in German). All answers were added up so as to calculate a total score (referred to as *TypeA – Sum*). Higher scores indicate a Type-A behaviour pattern. According to Perlman et al. ([50], p. 16) the scale shows an internal consistency of Cronbach's $\alpha = 0.71$ for men and $\alpha = 0.70$ for women ($n = 3000$).

Fear of Negative Evaluation

The *Brief Fear of Negative Evaluation—Revised Scale (BFNE-R)* [53] translated into German by Reichenberger et al. [62] and validated on the basis of four studies was used to measure fear of negative evaluation. The *BFNE – R* consists of 12 positively phrased questions evaluated on a five-point Likert scale ranging from 1 = *not at all a characteristic of me* to 5 = *absolutely a characteristic of me* (please refer to Figure A4 in Appendix B for a copy of the complete *BFNE – R* in German). Its cumulative value may take on numbers between 12 and 60 (referred to as *Fear – Sum*). Higher values correspond to a higher fear of negative evaluation. Reichenberger et al. ([62], p. 173) found an internal consistency of Cronbach's $\alpha = 0.94$ ($n = 414$) for the *BFNE – R*.

Big Five

Goldberg developed the *International Personality Item Pool-Five Factor Model (IPIP-FFM)*, which consists of 50 items and is used to examine the five dimensions of personality [63]. In order to keep the questionnaire short, we used the *Mini – IPIP Scale* by Donnellan et al. ([64], p. 193). It uses 20 questions to be answered on a five-point Likert scale. Previous studies showed an internal consistency of >0.60 for Cronbach's α (ibid.).

There are four questions per personality dimension. Except for the dimension *Openness*, which has three negatively and one positively directed question, all dimensions of the model have two positively and two negatively directed questions. The positive questions are rated from 1 = *very inaccurate* to 5 = *very accurate*. Negatively directed questions use the reverse order. For interpretation, the values for each dimension are added up (referred to as *Extra – Sum* for extraversion, *Conscient – Sum* for conscientiousness, *Neuro – Sum* for neuroticism and *Agree – Sum* for agreeableness). The German translation for this study was taken from the official IPIP website (online: <https://ipip.ori.org/newItemTranslations.htm> (accessed on 27 March 2021)) and a copy of it can be found in Figure A3 of Appendix B.

Affect

In order to measure affect we used the German translation of the short form of the *Positive Affect Negative Affect Scale (PANAS)* [59,60]. The scale uses 20 adjectives to describe feelings and sensations, which are to be rated on a five-point Likert scale ranging from 1 = *not at all* to 5 = *absolutely*. Half of the adjectives concern *PA* the other half *NA* (please refer to Figure A5 in Appendix B for a copy of the *PANAS* in German). Evaluations use the respective means. Previous studies have shown high internal consistency for the scale, with Cronbach's α values ranging from 0.86 to 0.90 for *PA* and from 0.84 to 0.87 for *NA* [59,60].

Task performance

In our study we were interested in whether stress has an effect on task performance. In order to measure task performance we used *Dots: The coordination game* smartphone app (online: <https://www.dots.co/> (accessed on 30 March 2021)). The number of points achieved in this game increase by vertically or horizontally connecting points of the same color. Extra points are given when similar colored squares are connected. Our test participants played the one-minute game mode *time play* twice after they had completed the game introduction. Their first round was played under normal conditions. After displaying the leaderboard, which indicated their rank being either at the top or at the bottom of the leaderboard, the second round was played under stress conditions. The difference in

the number of points achieved in both rounds was used to interpret the change in task performance (referred to as *points – DIF*).

Stress

We used *RMSSD* to objectively measure stress levels of all participants throughout the 60 s they were playing. The relative change from the second to the first measurement was used to interpret the changed stress level (referred to as *RMSSD – DIF*). As in Baumgartner et al. [34], a chest strap equipped with a heart rate sensor was used for recording. The measurement data was recorded with the *Suunto Ambit3 Run* heart rate monitor. The standard version of Kubios HRV was used to evaluate the data. Measurement quality was ensured via the artefact correction integrated in Kubios HRV. Since the recording of the *HRV* measurement had to be stopped manually, all measurement data in Kubios HRV was shortened to the 60 s relevant for playing the *Dots: The coordination game*. For the measurement of *HRV* values, the points recommended by Castaldo et al. ([30], pp. 376–377) were considered as follows: The length of the *HRV* measurements was set to 60 s, i.e., the duration of the game mode *Time play*. The leaderboard position was set as the stressor. Test subjects played the smartphone game in a sitting position, which avoided additional physical activity.

3.4. Procedure

The analysis of different effects of the leaderboard position on stress as well as task performance was investigated using a between-subject experimental design. We used the measures outlined in Section 3.3 to form two homogeneous groups. Questionnaires containing the 62 questions were provided online and completed by participants prior to starting the experiment. In addition, an ID was generated with each questionnaire, which was then used to link the questionnaire with the *HRV*-measurement during the experiment.

Participants were offered a twenty minutes time window for carrying out the experiment in a prepared room at our institution. To assure a standardised procedure, the experimental procedure was pre-tested and validated. At the start of the experiment, participants were told which smartphone game they will play. They were instructed that a one-minute timed mode will be played twice and that the goal was to score as many points as possible. Before playing the intro, they were asked to put on the chest strap (note: they were shown how to correctly wear such a chest strap). The start of the *HRV* measurement was triggered by our experiment facilitator via the *Suunto Ambit3 Run* heart rate monitor. The mobile game was played by all subjects on a provided *Honor 9 Lite* smartphone.

After the first round of the game, the *HRV* measurement was stopped. It was explained to the participants that their performance will now be displayed on a leaderboard, and compared to 19 other players. The achieved score was entered into the pop-up window (cf. Figure A2) and the (manipulated) placement subsequently displayed on the leaderboard. Afterwards, the game was played for a second time, before participants were asked to take off the chest strap and eventually were debriefed (i.e., told about the manipulation) and asked not to inform other participants about the experiment procedure and conditions. The experiment was conducted in spring 2019 and was approved by the school's Research Ethics group in terms of ethical considerations regarding research with human participation.

4. Results

From March 2019 to April 2019, a total of 48 people completed the online questionnaire and were consequently asked to participate in the experimental study. We received 40 responses for which we then used the above outlined measurements (cf. Section 3.3) to evenly split them into two groups á 20 participants. The experiment was conducted in May 2019 at our institution.

4.1. Data Cleansing

80 heart measurements, two from each individual, were exported from the *Suunto Ambit3 Run* and analysed in the *Kubios HRV Standard* software. Measurements were cropped to a duration of 60 s and artefact correction was carried out. For the artefact correction the respective threshold was set to *Medium*. The data of five participants had to be removed due to unrealistic *RMSSD* values.

In order to remove *RMSSD – DIF* outliers we applied Hoaglin et al.'s outlier labelling rule [65], which led to the exclusion of one more dataset, leaving us with a final sample size of $n = 34$ (21 female). Nine of the participants were born before 1995, all the other were younger. Normal distribution for the *RMSSD – DIF* was evaluated using the Kolmogorov-Smirnov test (K-S test), whereas for the internal consistency we used Cronbach's α (cf. Table 1).

Table 1. Internal Consistency of Survey Constructs.

Scale Name	Number of Items	Cronbach's α
Framingham Type-A Scale	10	0.415
BFNE – R	12	0.920
Mini – IPIP Extraversion	4	0.702
Mini – IPIP Agreeableness	4	0.780
Mini – IPIP Conscientiousness	4	0.723
Mini – IPIP Neuroticism	4	0.720
PANAS – NA	12	0.672

4.2. Hypothesis Evaluation

Tables 2 and 3 summarize our experiment results with respect to the earlier stated hypotheses. The following Section 4.3 will discuss some additional subsequent analyses.

Table 2. Evaluation Core Hypotheses: Leaderboard, Stress and Task Performance.

Hypothesis	Statistics	Variable	Result
H1: The use of a leaderboard leads to increased levels of stress.	Paired-samples <i>t</i> -test	Mean <i>RMSSD</i>	38.923 ms (Round 1) 35.342 ms (Round 2) $p = 0.052$
H2: Individuals experiencing higher levels of stress show lower task performance	Correlation according to Bravais and Pearson	<i>RMSSD</i> (Round 1)	$r = -0.317$ $p = 0.068$
H3: The use of a leaderboard leads to higher task performance.	Paired-samples <i>t</i> -test	mean score of points in dots	101.441 (without leaderboard) 110.735 (with leaderboard) $p = 0.075$
H4: Individuals ranked among the 'Bottom 3' experience more stress than individuals ranked among the 'Top 3'.	<i>t</i> -test for independent samples	Independent variable: 'Top 3' or 'Bottom 3' Dependent variable: <i>RMSSD – DIF</i>	Normal distribution of <i>RMSSD – DIF</i> K-S test: $p = 0.200$; Levene's test of equality of variance $p = 0.530$; <i>t</i> -test $p = 0.592$; mean <i>RMSSD – DIF</i> 'Top 3' -0.023 ; mean <i>RMSSD – DIF</i> 'Bottom 3' -0.071
H5: Individuals ranked among the 'Bottom 3' show a higher increase in task performance than individuals ranked among the 'Top 3'.	<i>t</i> -test for independent samples	Independent variable: 'Top 3' or 'Bottom 3'; Dependent variable: <i>points – DIF</i>	Normal distribution of <i>points – DIF</i> K-S test: $p = 0.200$; Levene's test of equality of variance $p = 0.312$; <i>t</i> -test $p = 0.238$; mean <i>points – DIF</i> 'Top 3' 15.688; mean <i>points – DIF</i> 'Bottom 3' 3.611

Table 3. Evaluation Control Hypotheses: Individual Behavior and Personality.

Hypothesis	Method	Variable	Result
H6: <i>Type-A behavior increases stress.</i>	Correlation according to Bravais and Pearson	RMSSD (Round 1); <i>TypeA – Sum</i>	$r = -0.286$ $p = 0.101$
H7: <i>Fear of negative evaluation increases stress.</i>	Correlation according to Bravais and Pearson	RMSSD (Round 1); <i>Fear – Sum</i>	$r = -0.058$ $p = 0.744$
H8: <i>High levels of extraversion reduce stress.</i>	Correlation according to Bravais and Pearson	RMSSD (Round 1); <i>Extra – Sum</i>	$r = 0.128$ $p = 0.470$
H9: <i>High levels of conscientiousness reduce stress.</i>	RMSSD (Round 1)	<i>Conscient – Sum</i>	$r = -0.106$ $p = 0.552$
H10: <i>High levels of agreeableness reduces stress.</i>	Correlation according to Bravais and Pearson	RMSSD (Round 1) <i>Agree – Sum</i>	$r = -0.270$ $p = 0.123$
H11: <i>High levels of neuroticism increases stress.</i>	Correlation according to Bravais and Pearson	RMSSD (Round 1) <i>Neuro – Sum</i>	$r = 0.282$ $p = 0.106$
H12: <i>Negative Affect (NA) increases stress.</i>	Correlation according to Bravais and Pearson	RMSSD (Round 1) <i>NA – Sum</i>	$r = -0.080$ $p = 0.654$

4.3. Additional Analyses

Additionally, we conducted some exploratory analyses using variance-based partial least squares structural equation modelling (PLS-SEM) with *SmartPLS* (v. 3.3.2) [66,67] (online: <https://www.smartpls.com/> (accessed on 7 April 2021)). The exploratory character of this additional analysis led to the decision to use variance-based partial least squares SEM (PLS-SEM) instead of co-variance-based SEM [66,67]. PLS-SEM has been applied in a variety of different research settings [68], and is especially suited to predict complex relationships on an exploratory level due to its block-wise estimation process [69]. Hence, it is considered an appropriate choice for analyzing the proposed hypotheses despite its lack of global model fit indices such as the Root Mean Square Error of Approximation (RMSEA) in co-variance based SEM. Although Tenenhaus, Vinzi, Chatelin, and Lauro [70] developed a Goodness of Fit index for PLS, this has been widely criticized and is also only applicable when comparing models in limited contexts, such as multigroup-analysis [67]. When compared to the advantages of using PLS for the respective study, this shortcoming seems acceptable, as the intention of this additional analysis lies within testing relationships on an exploratory level.

We were especially interested in finding out whether the manipulated leaderboard position accounts for different outcomes of the stress-performance relationship above and beyond the other indicators. Since the setting of our study did not really put participants in a competitive situation, we decided to especially focus on fear of negative evaluation as a predictor of task performance besides stress for these exploratory analyses. In the first step of the analysis the reliability of indicators and constructs, as well as construct and discriminant validity were examined more closely following the common recommendations for reliability and validity in *SmartPLS* [67,71]. The majority of the indicators load above the threshold of 0.700. Two indicators are above or exactly 0.600. The remaining two rank under 0.600 (more precisely: 0.575 and 0.543). In line with Hulland [72] these indicators are kept in the model, as only values of 0.400 or lower should definitely be excluded. This means that indicator reliability can be assumed for all scales. Composite reliability can also be assumed, as it is above the threshold of 0.700 for all constructs. In addition, all constructs show Average Variance Extracted (AVE) values above 0.500. In order to look at discriminant validity, item level cross-loadings were analyzed [67]. As no high cross-loadings were found on an item level, discriminant validity may be given at the item level.

After analyzing the reliability and validity of the model, we evaluated the structural model based on 5000 bootstraps while controlling for age and gender. The respective path model is depicted in Figure 2.

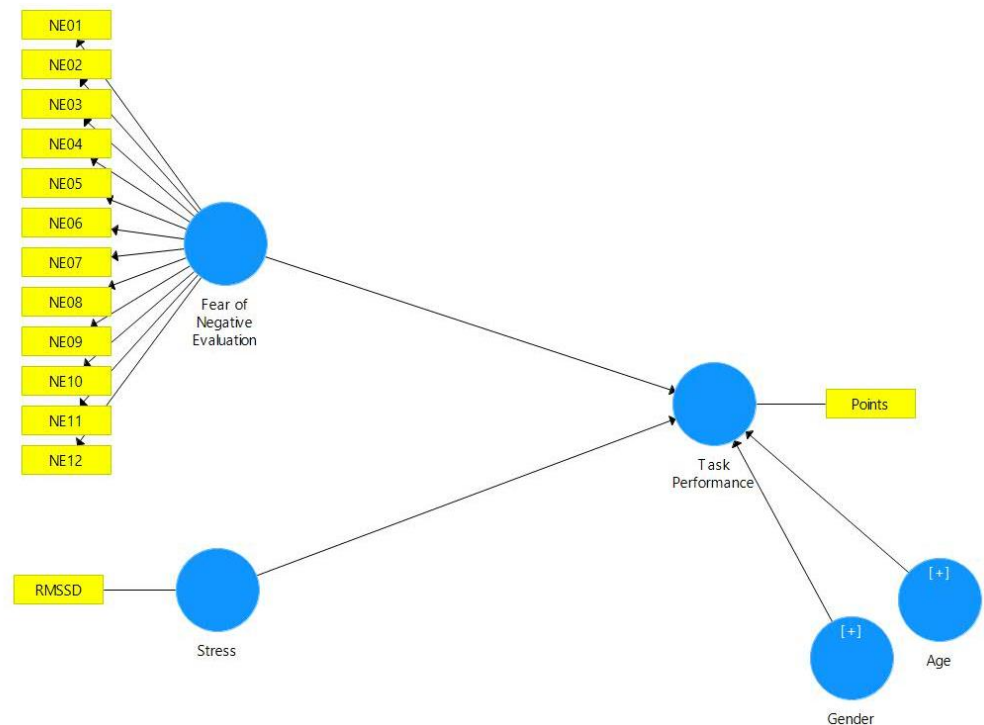


Figure 2. Model.

In order to identify potential group differences in this exploratory approach, we calculated the structural model for the different settings. For Round 1 we calculated the relationships for the entire sample, for Round 2 we added two groups indicating the random selection for the leaderboard manipulation. Group 0 included the subjects randomly assigned to the 'Top 3', while Group 1 included the subjects randomly assigned to the 'Bottom 3'. This led to the following results:

- In Round 1, considering the full sample, stress is an indicator of task performance (-0.360^{**}). Fear of Negative Evaluation does not show a significant relationship with task performance.
- In Round 2, Fear of Negative Evaluation is an indicator of performance (-0.479^{***}) for both groups.
 - In Group 0, stress also as a negative indicator of performance (-0.111^{***})
 - In Group 1, however, stress as a positive indicator of performance (0.184^{***})

In summary this means that prior to the introduction of a leaderboard, the experience of stress seems to play a role in terms of task performance. That is, those test participants who experienced higher levels of stress achieved a lower number of points during the first round. After the introduction of a leaderboard, however, stress as well as *Fear of Negative Evaluation* led to lower task performance during Round 2 for Group 0 (those participants who were randomly assigned to the 'Top 3'). For Group 1 (those participants who were randomly assigned to the 'Bottom 3'), on the other hand, only *Fear of Negative Evaluation* led to lower task performance, whereas higher levels of stress even led to a higher number of points achieved by members of this group.

5. Discussion

A manipulated leaderboard was used in order to analyze the respective effects for stress and task performance. During the experiment participants played the smartphone

game *Dots: The coordination game* in a supervised setting. By using the one-minute time mode, the game element time pressure, which can increase task-related stress, was used ([48], p. 298). In order to compare the test participants' task performance under stress, they were, as recommended by Lazarus et al. ([48], p. 299), tested twice. All participants carried out the first measurement under the same conditions. The second measurement was carried out after the respective manipulated leaderboard positions were announced. During the experiment the guidelines for HRV experiments according to Castaldo et al. ([30], pp. 376–377) were applied and implemented.

To examine the effect of using a leaderboard and its impact on stress, the mean values of the first measurement were compared to the mean values of the second measurement. There was no leaderboard in use for the first measurement. The second measurement was carried out after the leaderboard position was announced. However, the results do not indicate any significant differences in stress levels between Rounds 1 and 2, which means that H1 cannot be confirmed. Also, individuals experiencing higher levels of stress did not show lower task performance, thus showing no support for H2. In a next step the mean value scores of the two game rounds were compared. In the second round of the game, after the leaderboard position was displayed, a significant increase in task performance was expected. The results indicate an increase in the average mean, but no significant difference between the two rounds could be determined, contradicting the assumption put forward by H3.

For H4, we wanted to find out whether the leaderboard position leads to different levels of stress in Round 2. In order to make the different *RMSSD* values comparable, we used *RMSSD – DIF*. The leaderboard position was interpreted as a stress factor. The non-significant result may thus be interpreted in light of the statement expressed by Lazarus et al. ([48], pp. 299–300) that some people are stressed by the danger of failure, while others seem to not be effected.

In contrast to Lazarus et al. ([48], pp. 299–300), however, no significant difference between the two populations could be identified (cf. H5). That is, it could not be determined with sufficient significance that test persons' results, depending on the leaderboard position shown to them, improved or deteriorated. Yet, in contrast to the hypothesis, an increase in the mean value was found in both groups, where the increase in the 'Top 3', at 15.688 points, was significantly higher than the increase in the 'Bottom 3' (3.611).

And finally, contrary to H6–H12, also none of the individual variables accounted for significant differences in stress during our experiment.

Thus, in summary, based on the 12 hypotheses evaluated during this experiment, we were unable to find a significant relationship between stress and the leaderboard position. In addition, no significant relationship between stress and task performance could be demonstrated.

Being confronted with these somewhat unsatisfactory results, we decided to perform some additional exploratory analyses using PLS-SEM. Specifically analyzing the parallel effects of *Fear of Negative Evaluation* and stress on task performance, we could find that stress was indeed a significant indicator of task performance in Round 1. Although, *Fear of Negative Evaluation* did not have an effect on the task performance without the use of a leaderboard.

The effect size for the negative relationship between *Fear of Negative Evaluation* and task performance was even higher in Round 2—after the introduction of the leaderboard. On top, we could find a difference in the direction of the relationship between stress and task performance for the 'Top 3' and 'Bottom 3' groups. For individuals ranked among the 'Top 3' stress led to lower task performance, while for those individuals ranked among the 'Bottom 3' higher levels of stress led to even higher levels of task performance.

These results are in line with the considerations of King [17], who describe that individuals might be encouraged to compete for higher placements by a leaderboard, which can have a positive effect on task performance. Also the Transactional Theory of Stress [73] includes the concept of positive stressors. Thus a stressful event—like being

ranked among the ‘Bottom 3’ of a leaderboard—might spur ambition and therefore even lead to increased performance for those individuals specifically, who do not experience high levels of *Fear of Negative Evaluation*.

6. Limitations

Our investigations into the effects leaderboard positions may have on peoples’ heart rate variability and respective task performance were driven by Motowidlo et al.’s work on occupational stress and its causes and consequences for job performance [22]. To this end, our results show no significant connection between how people were told to be ranked on a leaderboard and their perceived stress level (expressed by a change in heart rate). Lack of significance, however, may primarily be owed to our rather small sample size (i.e., $n = 40$), which should thus be considered a significant limitation of our study.

Also, the sample of participants was not only small, but did consist of students and colleagues. While the small size was subject to the given limitations in time and available resources, the inclusion criteria were intended, since we wanted to impede any effects related to age and/or previous experience with the used stimulus (i.e., *Dots: The coordination game*). Such, however, heavily impacts on the generalizability of our results.

Furthermore, it should be noted that not all individual variables affecting stress had been taken into account. For example, we focused solely on measuring stress via heart rate variability measured through a chest strap, where recordings were cropped to frames of 60 s. No subjective assessment of stress were collected, neither did we consider medical reasons and their potential effects on stress, which both could have significantly improved validity with respect to participants’ perceived stress.

Also, it has to be highlighted that the experimental setting was rather artificial and did not incorporate any consequences with respect to participants’ performance. That is, they had nothing to gain nor anything to lose. Our goal was to create stress by introducing a manipulated leaderboard. When introducing these types of false results, test participants need to be convinced that the presented information is plausible ([48], p. 297). Yet, during the experiment, four of our participants asked whether the results on the leaderboard were correct, which suggests that at least some of them questioned the sequence shown and suspected manipulation. Hence, although we aimed at creating a realistic setting, the artificial nature of our experiment may have significantly limited the validity of results.

Finally, personal motivation influences how much a person aims to prevent failure ([48], p. 296), which in turn might influence perceived levels of stress. Here it may be that participants did not perceive a low leaderboard position to be a failure and therefore did not experience failure-related stress symptoms. Also, it might be that playing a game was not perceived as being such a stressful task. As Lazarus et al. ([48], p. 297) point out, the motivation of test subjects in the case of task-related stress depends on how the requirements are interpreted and whether there is ultimately a noticeable increase in one’s performance level. The fact that the number of achieved points was not publicly visible may have added to this feeling of indifference. Also, according to Hamari and Koivisto ([9], p. 8), it is important with gamification that users pursue the same goal. In the context of our study, participants’ own goals were, however, not taken into account.

7. Summary and Future Recommendations

We reported on the results of an experimental study investigating the effects leaderboard positions would have on peoples’ heart rate variability and respective task performance. We used a smartphone game, a manipulated leaderboard and a chest strap heart rate monitor to investigate potential connections. Although our experiment with $n = 40$ participants did not yield significant statistical results, findings hint towards a link between the use of a leaderboard and increased levels of stress ($p = 0.052$). Connections between concrete leaderboard positions and perceived stress, consequent effects concerning task performance, or individual characteristics such as Type-A behavior or affect, however, could not be identified. Although a lack of connection as such may not be considered an in-

significant finding, we have discussed a number of limitations that could have affected the validity of our experiment (e.g., sample size, measurement instruments etc.; cf. Section 6 for a respective discussion). Hence, for future studies we would recommend a number of alterations.

First, we recommend that they should focus on a larger and more stratified sample, and investigate potential differences in target populations. Here, it is further advised to search for more stable personality-type measures, as in our case the internal consistency of both the *Framingham Type-A Scale* and the *PANAS* were not satisfactory. To this end, future studies may also control for individual motivation as a predictor for task performance.

Second, we suggest the use of additional/other potentially less intrusive technologies to collect physiological parameters such heart rate or blood pressure as, e.g., demonstrated by Gonzalez-Viejo and colleagues [44] or Massaro et al. [42].

Finally, future studies should expand on the ‘realness’ of the experimental condition and thus control for potential contextual influence factors.

Author Contributions: The article was a collaborative effort by all three authors. M.S. was the lead researcher in the presented study, which was supervised by T.S. T.S. furthermore wrote the original draft of the article, supported by S.S. who helped writing, reviewing and editing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the MCI Ethics Board on 19 March 2019.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to informed consent restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

#	ID	Points
1	LMH1	136
2	RAFE5	134
3	RPTA7	129
4	RDDI2	125
5	RATI8	124
6	REGO7	119
7	IPNU6	114
8	IBKU6	111
9	TCMI4	110
10	IUMA5	109
11	RDPA6	103
12	TPGA6	100
13	ILTO1	95
14	TLKE5	91
15	RICE4	90
16	RURU8	89
17	ZMDU8	86
18	RAFI2	80
19	RPSI5	74
20	TGWE4	73

Figure A1. Manipulated leaderboard displaying a rank in the top position.

Auf dieser Seite wird Folgendes angezeigt

Score eingeben

Ok Abbrechen

Figure A2. Pop-up window for entering the score.

Appendix B

2. Bitte beantworten Sie die folgenden Fragen möglichst wahrheitsgetreu

Ich...	sehr unzutreffend	sehr zutreffend
bringe eine Party in Schwung	<input type="radio"/>	<input type="radio"/>
fühle mit anderen Menschen mit	<input type="radio"/>	<input type="radio"/>
erledige Hausarbeiten unmittelbar	<input type="radio"/>	<input type="radio"/>
habe häufig Stimmungsschwankungen	<input type="radio"/>	<input type="radio"/>
habe eine lebhafte Vorstellungskraft	<input type="radio"/>	<input type="radio"/>
spreche nicht viel	<input type="radio"/>	<input type="radio"/>
bin nicht interessiert an den Problemen anderer Leute	<input type="radio"/>	<input type="radio"/>
vergesse oft, Dinge an ihren Platz zurückzulegen	<input type="radio"/>	<input type="radio"/>
bin die meiste Zeit entspannt	<input type="radio"/>	<input type="radio"/>
bin an abstrakten Ideen nicht interessiert	<input type="radio"/>	<input type="radio"/>
spreche mit vielen verschiedenen Leuten auf Partys	<input type="radio"/>	<input type="radio"/>
fühle die Gefühle anderer	<input type="radio"/>	<input type="radio"/>
mag Ordnung	<input type="radio"/>	<input type="radio"/>
rege mich leicht auf	<input type="radio"/>	<input type="radio"/>
habe Schwierigkeiten abstrakte Ideen zu verstehen	<input type="radio"/>	<input type="radio"/>
halte mich im Hintergrund	<input type="radio"/>	<input type="radio"/>
bin nicht wirklich interessiert an Anderen	<input type="radio"/>	<input type="radio"/>
verursache großes Durcheinander	<input type="radio"/>	<input type="radio"/>
fühle mich selten deprimiert (bedrückt)	<input type="radio"/>	<input type="radio"/>
habe kein gutes Vorstellungsvermögen	<input type="radio"/>	<input type="radio"/>

Figure A3. German translation of the *Mini – IPIP* Scale by Donnellan et al. ([64], p. 193).

3. Bitte beantworten Sie die folgenden Fragen möglichst wahrheitsgetreu

	überhaupt nicht charakteristisch für mich	äußerst charakteristisch für mich
Ich mache mir Gedanken darüber, was andere Leute von mir denken, auch wenn ich weiß, dass es egal ist	<input type="radio"/>	<input type="radio"/>
Es bekümmert mich, wenn ich merke, dass andere Leute einen schlechten Eindruck von mir bekommen	<input type="radio"/>	<input type="radio"/>
Ich habe oft Angst, dass andere Leute meine Fehler bemerken	<input type="radio"/>	<input type="radio"/>
Ich mache mir öfters Gedanken darüber, welchen Eindruck ich auf jemand anderes mache	<input type="radio"/>	<input type="radio"/>
Ich habe Angst, dass andere sich nicht positiv über mich äußern	<input type="radio"/>	<input type="radio"/>
Ich habe Angst, dass andere Leute etwas an mir auszusetzen haben	<input type="radio"/>	<input type="radio"/>
Die Meinung anderer Leute über mich ist mir wichtig	<input type="radio"/>	<input type="radio"/>
Wenn ich mit jemandem spreche, mache ich mir Gedanken darüber, was der andere über mich denken könnte	<input type="radio"/>	<input type="radio"/>
Normalerweise mache ich mir Gedanken darüber, wie ich auf andere wirke	<input type="radio"/>	<input type="radio"/>
Es macht mir etwas aus, wenn ich weiß, dass mich jemand beurteilt	<input type="radio"/>	<input type="radio"/>
Manchmal glaube ich, ich beschäftige mich viel zu sehr damit, was andere Leute von mir denken	<input type="radio"/>	<input type="radio"/>
Ich habe oft Angst, dass ich etwas Falsches sagen oder tun würde	<input type="radio"/>	<input type="radio"/>

Figure A4. German translation of the *BFNE – R* [53,62].

4. Die folgenden Wörter beschreiben unterschiedliche Gefühle und Empfindungen. Lesen Sie jedes Wort und tragen Sie dann in die Skala die Intensität ein. Geben Sie bitte an, wie Sie sich im Allgemeinen fühlen.

	gar nicht	äußerst
aktiv	<input type="radio"/>	<input type="radio"/>
bekümmert	<input type="radio"/>	<input type="radio"/>
interessiert	<input type="radio"/>	<input type="radio"/>
freudig erregt	<input type="radio"/>	<input type="radio"/>
verärgert	<input type="radio"/>	<input type="radio"/>
stark	<input type="radio"/>	<input type="radio"/>
schuldig	<input type="radio"/>	<input type="radio"/>
erschrocken	<input type="radio"/>	<input type="radio"/>
feindselig	<input type="radio"/>	<input type="radio"/>
angeregt	<input type="radio"/>	<input type="radio"/>
stolz	<input type="radio"/>	<input type="radio"/>
gereizt	<input type="radio"/>	<input type="radio"/>
begeistert	<input type="radio"/>	<input type="radio"/>
beschämt	<input type="radio"/>	<input type="radio"/>
wach	<input type="radio"/>	<input type="radio"/>
nervös	<input type="radio"/>	<input type="radio"/>
entschlossen	<input type="radio"/>	<input type="radio"/>
aufmerksam	<input type="radio"/>	<input type="radio"/>
durcheinander	<input type="radio"/>	<input type="radio"/>
ängstlich	<input type="radio"/>	<input type="radio"/>

Figure A5. German translation of the Positive Affect Negative Affect Scale (PANAS) [59,60].

5. Eigenschaften und Qualitäten, die Sie beschreiben

	überhaupt nicht	einigermaßen	ziemlich gut	sehr gut
ehrgeizig und kompetitiv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
normalerweise unter Zeitdruck	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
rechthaberisch oder dominierend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
das Bedürfnis haben, in den meisten Dingen zu brillieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
zu schnell essen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Gefühl am Ende eines produktiven Tages in der Hochschule

	nein	ja
Ich fühlte mich oft unter Zeitdruck	<input type="radio"/>	<input type="radio"/>
Die Aufgaben bleiben mir im Gedanken, deswegen denke ich auch am Ende des Tages darüber nach	<input type="radio"/>	<input type="radio"/>
Die Arbeit hat mich oft an meine Grenzen gebracht	<input type="radio"/>	<input type="radio"/>
Ich fühlte mich oft unsicher, unwohl oder unzufrieden damit, wie gut ich mit meinen Aufgaben zurecht komme	<input type="radio"/>	<input type="radio"/>
Ich rege mich auf, wenn ich auf etwas oder jemanden warten musste	<input type="radio"/>	<input type="radio"/>

Figure A6. German translation of the Type-A Scale according to Framingham [61].

References

- Deterding, S.; Dixon, D.; Khaled, R.; Nacke, L. From game design elements to gamefulness: Defining “gamification”. In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 28–30 September 2011; pp. 9–15.
- Huotari, K.; Hamari, J. Defining gamification: A service marketing perspective. In Proceedings of the 16th International Academic MindTrek Conference, Tampere, Finland, 3–5 October 2012; pp. 17–22.
- Dale, S. Gamification: Making work fun, or making fun of work? *Bus. Inf. Rev.* **2014**, *31*, 82–90. [\[CrossRef\]](#)
- Andrade, F.R.; Mizoguchi, R.; Isotani, S. The bright and dark sides of gamification. In Proceedings of the International Conference on Intelligent Tutoring Systems, Zagreb, Croatia, 6–10 June 2016; Springer: Cham, Switzerland, 2016; pp. 176–186.
- Farzan, R.; DiMicco, J.M.; Millen, D.R.; Dugan, C.; Geyer, W.; Brownholtz, E.A. Results from deploying a participation incentive mechanism within the enterprise. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, 5–10 April 2008; pp. 563–572.
- Kim, T.W. Gamification of labor and the charge of exploitation. *J. Bus. Ethics* **2018**, *152*, 27–39. [\[CrossRef\]](#)
- Reeves, B.; Read, L. *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete*; Harvard Business School Publishing: Boston, MA, USA, 2009.
- Hamari, J.; Koivisto, J.; Sarsa, H. Does gamification work?—A literature review of empirical studies on gamification. In Proceedings of the 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 3025–3034.
- Hamari, J.; Koivisto, J. Social Motivations To Use Gamification: An Empirical Study Of Gamifying Exercise. In Proceedings of the 21st European Conference on Information Systems, Utrecht, The Netherlands, 5–8 June 2013.
- Sailer, M.; Homner, L. The gamification of learning: A meta-analysis. *Educ. Psychol. Rev.* **2020**, *32*, 77–112. [\[CrossRef\]](#)
- Sailer, M.; Sailer, M. Gamification of in-class activities in flipped classroom lectures. *Br. J. Educ. Technol.* **2021**, *52*, 75–90. [\[CrossRef\]](#)
- Zainuddin, Z.; Shujahat, M.; Haruna, H.; Chu, S.K.W. The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system. *Comput. Educ.* **2020**, *145*, 103729. [\[CrossRef\]](#)
- Sanchez, D.R.; Langer, M.; Kaur, R. Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Comput. Educ.* **2020**, *144*, 103666. [\[CrossRef\]](#)
- Zainuddin, Z.; Chu, S.K.W.; Shujahat, M.; Perera, C.J. The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educ. Res. Rev.* **2020**, *30*, 100326. [\[CrossRef\]](#)
- Groening, C.; Binnewies, C. “Achievement unlocked!”—The impact of digital achievements as a gamification element on motivation and performance. *Comput. Hum. Behav.* **2019**, *97*, 151–166. [\[CrossRef\]](#)
- Bowey, J.T.; Birk, M.V.; Mandryk, R.L. Manipulating leaderboards to induce player experience. In Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, London, UK, 5–7 October 2015; pp. 115–120.
- King, D.; Delfabbro, P.; Griffiths, M. Video game structural characteristics: A new psychological taxonomy. *Int. J. Ment. Health Addict.* **2010**, *8*, 90–106. [\[CrossRef\]](#)
- Thiebes, S.; Lins, S.; Basten, D. Gamifying Information Systems—A Synthesis of Gamification Mechanics and Dynamics. In Proceedings of the ECIS 2014, 22nd European Conference on Information Systems, Tel Aviv, Israel, 9–11 June 2014.

19. Callan, R.C.; Bauer, K.N.; Landers, R.N. How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. In *Gamification in Education and Business*; Springer: Cham, Switzerland, 2015; pp. 553–568.
20. Jamal, M. Relationship of job stress and Type-A behavior to employees' job satisfaction, organizational commitment, psychosomatic health problems, and turnover motivation. *Hum. Relat.* **1990**, *43*, 727–738. [\[CrossRef\]](#)
21. Abbe, O.O.; Harvey, C.M.; Ikuma, L.H.; Aghazadeh, F. Modeling the relationship between occupational stressors, psychosocial/physical symptoms and injuries in the construction industry. *Int. J. Ind. Ergon.* **2011**, *41*, 106–117. [\[CrossRef\]](#)
22. Motowidlo, S.J.; Packard, J.S.; Manning, M.R. Occupational stress: Its causes and consequences for job performance. *J. Appl. Psychol.* **1986**, *71*, 618. [\[CrossRef\]](#)
23. Spector, P.E. Using self-report questionnaires in OB research: A comment on the use of a controversial method. *J. Organ. Behav.* **1994**, *15*, 385–392. [\[CrossRef\]](#)
24. Cohen, S.; Kamarck, T.; Mermelstein, R. A global measure of perceived stress. *J. Health Soc. Behav.* **1983**, *24*, 385–396. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Appelhans, B.M.; Luecken, L.J. Heart rate variability as an index of regulated emotional responding. *Rev. Gen. Psychol.* **2006**, *10*, 229–240. [\[CrossRef\]](#)
26. Jobbágy, Á.; Majnár, M.; Tóth, L.K.; Nagy, P. HRV-based stress level assessment using very short recordings. *Period. Polytech. Electr. Eng. Comput. Sci.* **2017**, *61*, 238–245. [\[CrossRef\]](#)
27. Paniagua, S.; Herrero, R.; García-Pérez, A.I.; Calvo, L.F. Study of Binqui. An application for smartphones based on the problems without data methodology to reduce stress levels and improve academic performance of chemical engineering students. *Educ. Chem. Eng.* **2019**, *27*, 61–70. [\[CrossRef\]](#)
28. Tennakoon, W.; Wanninayake, W. Where play become effective: The moderating effect of gamification on the relationship between work stress and employee performance. *Econ. Res.* **2020**, *7*, 63–86. [\[CrossRef\]](#)
29. Horn, E.; Lee, S. Electronic evaluation of the fetal heart rate patterns preceding fetal death. *Am. J. Obstet. Gynecol.* **1963**, *87*, 814–826.
30. Castaldo, R.; Melillo, P.; Bracale, U.; Caserta, M.; Triassi, M.; Pecchia, L. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomed. Signal Process. Control* **2015**, *18*, 370–377. [\[CrossRef\]](#)
31. Tarvainen, M.P.; Niskanen, J.P.; Lipponen, J.A.; Ranta-Aho, P.O.; Karjalainen, P.A. Kubios HRV—heart rate variability analysis software. *Comput. Methods Programs Biomed.* **2014**, *113*, 210–220. [\[CrossRef\]](#)
32. Peltola, M. Role of editing of RR intervals in the analysis of heart rate variability. *Front. Physiol.* **2012**, *3*, 148. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Shaffer, F.; Ginsberg, J. An overview of heart rate variability metrics and norms. *Front. Public Health* **2017**, *5*, 258. [\[CrossRef\]](#)
34. Baumgartner, D.; Fischer, T.; Riedl, R.; Dreiseitl, S. Analysis of heart rate variability (HRV) feature robustness for measuring technostress. In *Information Systems and Neuroscience*; Springer: Cham, Switzerland, 2019; pp. 221–228.
35. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation* **1996**, *93*, 1043–1065. [\[CrossRef\]](#)
36. Shaffer, F.; Shearman, S.; Meehan, Z.M. The promise of ultra-short-term (UST) heart rate variability measurements. *Biofeedback* **2016**, *44*, 229–233. [\[CrossRef\]](#)
37. Nussinovitch, U.; Elishkevitz, K.P.; Katz, K.; Nussinovitch, M.; Segev, S.; Volovitz, B.; Nussinovitch, N. Reliability of ultra-short ECG indices for heart rate variability. *Ann. Noninvasive Electrocardiol.* **2011**, *16*, 117–122. [\[CrossRef\]](#)
38. Baek, H.J.; Cho, C.H.; Cho, J.; Woo, J.M. Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemed. E-Health* **2015**, *21*, 404–414. [\[CrossRef\]](#)
39. Li, Z.; Snieder, H.; Su, S.; Ding, X.; Thayer, J.F.; Treiber, F.A.; Wang, X. A longitudinal study in youth of heart rate variability at rest and in response to stress. *Int. J. Psychophysiol.* **2009**, *73*, 212–217. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Salahuddin, L.; Cho, J.; Jeong, M.G.; Kim, D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007; pp. 4656–4659.
41. Massaro, A.; Maritati, V.; Savino, N.; Galiano, A. Neural networks for automated smart health platforms oriented on heart predictive diagnostic big data systems. In Proceedings of the 2018 AEIT International Annual Conference, Bari, Italy, 3–5 October 2018; pp. 1–5.
42. Massaro, A.; Ricci, G.; Selicato, S.; Raminelli, S.; Galiano, A. Decisional Support System with Artificial Intelligence oriented on Health Prediction using a Wearable Device and Big Data. In Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 3–5 June 2020; pp. 718–723.
43. Kim, J.H.; Poulou, A.; Han, D.S. The Extensive Usage of the Facial Image Threshing Machine for Facial Emotion Recognition Performance. *Sensors* **2021**, *21*, 2026. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Gonzalez Viejo, C.; Fuentes, S.; Torrico, D.D.; Dunshea, F.R. Non-contact heart rate and blood pressure estimations from video analysis and machine learning modelling applied to food sensory responses: A case study for chocolate. *Sensors* **2018**, *18*, 1802. [\[CrossRef\]](#)
45. Lazarus, R.S. Theory-based stress measurement. *Psychol. Inq.* **1990**, *1*, 3–13. [\[CrossRef\]](#)
46. Christy, K.R.; Fox, J. Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Comput. Educ.* **2014**, *78*, 66–77. [\[CrossRef\]](#)

47. Jia, Y.; Liu, Y.; Yu, X.; Vaida, S. Designing leaderboards for gamification: Perceived differences based on user ranking, application domain, and personality traits. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 1949–1960.
48. Lazarus, R.S.; Deese, J.; Osler, S.F. The effects of psychological stress upon performance. *Psychol. Bull.* **1952**, *49*, 293–317. [\[CrossRef\]](#)
49. Hjortskov, N.; Rissén, D.; Blangsted, A.K.; Fallentin, N.; Lundberg, U.; Søgaard, K. The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* **2004**, *92*, 84–89. [\[CrossRef\]](#)
50. Perlman, B.; Hartman, E.A.; Lenahan, B. Validating of a Six Item Questionnaire For Assessing Type A Behavior. In proceedings of the Annual Meeting of the Midwestern Psychological Association, Chicago, IL, USA, 3–5 May 1984.
51. Packard, J.S.; Motowidlo, S.J. Subjective stress, job satisfaction, and job performance of hospital nurses. *Res. Nurs. Health* **1987**, *10*, 253–261. [\[CrossRef\]](#)
52. Pitarch, M.J.G. Brief version of the fear of negative evaluation scale—straightforward items (BFNE-S): Psychometric properties in a spanish population. *Span. J. Psychol.* **2010**, *13*, 981–989. [\[CrossRef\]](#)
53. Carleton, R.N.; McCreary, D.R.; Norton, P.J.; Asmundson, G.J. Brief fear of negative evaluation scale—Revised. *Depress. Anxiety* **2006**, *23*, 297–303. [\[CrossRef\]](#)
54. Rammstedt, B.; Koch, K.; Borg, I.; Reitz, T. Entwicklung und validierung einer kurzska für die messung der big-five-persönlichkeitsdimensionen in umfragen. *Zuma Nachrichten* **2004**, *28*, 5–28.
55. Lee-Baggley, D.; Preece, M.; DeLongis, A. Coping with interpersonal stress: Role of Big Five traits. *J. Personal.* **2005**, *73*, 1141–1180. [\[CrossRef\]](#)
56. Ebstrup, J.F.; Eplov, L.F.; Pisinger, C.; Jørgensen, T. Association between the Five Factor personality traits and perceived stress: Is the effect mediated by general self-efficacy? *Anxiety Stress Coping* **2011**, *24*, 407–419. [\[CrossRef\]](#)
57. Kanner, A.D.; Coyne, J.C.; Schaefer, C.; Lazarus, R.S. Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *J. Behav. Med.* **1981**, *4*, 1–39. [\[CrossRef\]](#)
58. Watson, D. Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. *J. Personal. Soc. Psychol.* **1988**, *54*, 1020–1030. [\[CrossRef\]](#)
59. Watson, D.; Clark, L.A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1988**, *54*, 1063–1070. [\[CrossRef\]](#)
60. Breyer, B.; Bluemke, M. *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)*; Technical Report; GESIS—Leibniz-Institute for the Social Sciences: Mannheim, Germany, 2016. [\[CrossRef\]](#)
61. Friedman, M.; Rosenman, R.H. Association of specific overt behavior pattern with blood and cardiovascular findings: Blood cholesterol level, blood clotting time, incidence of arcus senilis, and clinical coronary artery disease. *J. Am. Med. Assoc.* **1959**, *169*, 1286–1296. [\[CrossRef\]](#)
62. Reichenberger, J.; Schwarz, M.; König, D.; Wilhelm, F.H.; Voderholzer, U.; Hillert, A.; Blechert, J. Angst vor negativer sozialer Bewertung: Übersetzung und Validierung der Furcht vor negativer Evaluation–Kurzska (FNE-K). *Diagnostica* **2015**, *62*, 169–181. [\[CrossRef\]](#)
63. Goldberg, L.R. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personal. Psychol. Eur.* **1999**, *7*, 7–28.
64. Donnellan, M.B.; Oswald, F.L.; Baird, B.M.; Lucas, R.E. The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychol. Assess.* **2006**, *18*, 192. [\[CrossRef\]](#)
65. Hoaglin, D.C.; Iglewicz, B.; Tukey, J.W. Performance of some resistant rules for outlier labeling. *J. Am. Stat. Assoc.* **1986**, *81*, 991–999. [\[CrossRef\]](#)
66. Hair, J.F.; Ringle, C.M.; Sarstedt, M. PLS-SEM: Indeed a silver bullet. *J. Mark. Theory Pract.* **2011**, *19*, 139–152. [\[CrossRef\]](#)
67. Hair, J.F., Jr.; Hult, G.T.M.; Ringle, C.; Sarstedt, M. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*; Sage Publications: Thousand Oaks, CA, USA, 2016.
68. Henseler, J.; Dijkstra, T.K.; Sarstedt, M.; Ringle, C.M.; Diamantopoulos, A.; Straub, D.W.; Ketchen, D.J., Jr.; Hair, J.F.; Hult, G.T.M.; Calantone, R.J. Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organ. Res. Methods* **2014**, *17*, 182–209. [\[CrossRef\]](#)
69. Haenlein, M.; Kaplan, A.M. A beginner’s guide to partial least squares analysis. *Underst. Stat.* **2004**, *3*, 283–297. [\[CrossRef\]](#)
70. Tenenhaus, M.; Vinzi, V.E.; Chatelin, Y.M.; Lauro, C. PLS path modeling. *Comput. Stat. Data Anal.* **2005**, *48*, 159–205. [\[CrossRef\]](#)
71. Henseler, J.; Ringle, C.M.; Sinkovics, R.R. The use of partial least squares path modeling in international marketing. In *New Challenges to International Marketing*; Emerald Group Publishing Limited: Bingley, UK, 2009.
72. Hulland, J. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strateg. Manag. J.* **1999**, *20*, 195–204. [\[CrossRef\]](#)
73. Lazarus, R.S.; Folkman, S. Transactional theory and research on emotions and coping. *Eur. J. Personal.* **1987**, *1*, 141–169. [\[CrossRef\]](#)