

To compare or not to compare oneself? Studying the effect of motivation, affect, and comparison direction

Fleur Leenders^[5204402]

Universiteit Utrecht
Bachelor Information Sciences
Utrecht, the Netherlands
f.leenders@uu.nl

Abstract. Social comparison is a widely used tool to foster motivation among learners. Recent research has drawn attention to individual differences in preferences for and responses to social comparison in learning analytics. However, little is known about the behaviour of learners when they are presented with the choice to compare themselves. This research aimed to shed a light on the topic of choices in social comparison by investigating the predictive effects of affective state, achievement goals, and social comparison direction on these choices. Furthermore, the effects of achievement goals and comparison direction on affective response after such a self-chosen comparisons were investigated. 96 participants took part in an online experiment. They played a word game, during which they got the opportunity to compare themselves. The majority of participants compared themselves all three times, mostly out of curiosity. Furthermore social comparison direction did significantly affect feelings of pleasantness and arousal, both being higher after viewing a downwards comparison. The study had limitations, and no strong predictors for comparison choice were found. However, the results do provide several opportunities for future research.

Keywords: Social comparison · Learning analytics · Achievement goals · Affect · Comparison direction · Motivation.

Preface

Before you lies the bachelor thesis *“To compare or not to compare oneself? Studying the effect of motivation, affect, and comparison direction on the choice to compare”*. I wrote this thesis to complete my Bachelor’s degree in Information Sciences at Utrecht university.

This thesis started with an interest in what motivates students, and why. Despite being intrinsically motivated to learn new things, I noticed I can get too caught up in grades and performance, to the point where this holds me back. My curiosity for the effect of grades on motivation eventually led me to the topic of social comparison and its effects on learners.

First, I would like to thank my supervisor Jeroen Ooge. It was a joy to exchange ideas and learn from his feedback. His enthusiasm for this thesis never wavered, even when I doubted everything.

One other thing this thesis taught me, is that I am surrounded by many lovely people. From technical support and proofreads, to hugs and pep talks, these people were there for me. There are too many to name them all, but I hope they know that I am extremely grateful for all the support they gave me during this process.

I hope you enjoy reading.

Fleur Leenders

Utrecht, July 1, 2025

1 Introduction

Motivation is a topic of interest in many research contexts, including education. Researchers try to understand what motivates learners, and why [?]. A commonly used tool to foster motivation is social comparison [13, 18], the phenomenon where we compare our own abilities and opinions to those of others [16]. Ever since Festinger proposed his theory of social comparison [16], the motivations for and consequences of such comparisons have been widely researched in educational settings, showing that social comparison can affect emotions, motivation, and in turn, behaviour of learners [13].

While social comparison was already present in traditional classroom settings [13], technological advancements provide new opportunities for such comparisons. One such example is learning analytics, the collection and reporting of learner data [29, 33]. Learning analytics dashboards often include social comparison features [23].

Recent studies have investigated learners' responses to and interaction with social comparison in online learning environments [2, 3, 25]. Though providing mixed results, these studies all highlighted the same point: the need to understand and design for individual preferences when implementing social comparison in learning analytics. Individual preferences are thus important factors for designing social comparisons.

However, existing literature barely touched upon the following question: what do people do when they are given the choice to compare, and why? Research shows that some learners perceive social comparison as a useful tool to measure their performance or progress, while others prefer other reference frames to do so [1, 19, 28]. In addition, various studies have recorded participants not using comparisons [5, 8, 18, 22], yet little is known about what could affect this choice to compare.

This research aims to add to the existing body of work on social comparison in learning environments by exploring people's behaviour when they get the choice to compare themselves. It aims to do so by answering the following research questions:

- RQ1.** How do comparison direction, affective state, and achievement motivation affect the choice to compare oneself?
- RQ2.** How do comparison direction and achievement motivation affect affective state after self-chosen comparison?

First, I present existing work on social comparison, affect and achievement goals that inspired the research questions. To answer these research questions, I designed an experiment in which participants got the choice to compare their performance in a word game. Next, I present the results, with which this study adds to the growing body of knowledge on social comparison. It shows that often, the choice is made to compare oneself, and provides a starting point for understanding which factors may influence choice. Furthermore, results show how affective states are influenced by viewing such comparisons. Lastly, I discuss the value and limitations of the study and present opportunities for future work.

2 Related work and research opportunities

This section contains a preliminary literature review to explain the relevant concepts used in this research. First, social comparison theory and its presence in (online) education is explained. After that, the concepts of affect and motivation are explained, as well as their relation to social comparison.

2.1 Social comparison

According to Festinger [16], humans have an innate drive to compare themselves to others to assess their own capabilities and opinions. Especially when there is no objective way to measure one's abilities, social comparison is a useful tool for self-assessment.

Comparison targets

Comparisons can be made against different comparison targets, i.e. the other person or group one compares themselves with. Social comparisons in education often use the class average as a comparison target [23], though the group of social comparison targets can be adjusted. Two possible adjustments are the direction and distance of comparisons.

Direction. First, the direction of the social comparison can be adjusted. A common distinction is upward versus downward comparison. In an upward comparison, the comparison targets perform better than the source, whereas in a downward comparison, they perform worse [16, 39].

Festinger distinguished between abilities and opinions and their relation to the direction of social comparison. Regarding abilities, people have a drive to perform better and meet their upward comparison targets. Opinions, on the other hand, are more prone to be influenced by social processes. For example, in an argument, one can try to convince the other to change their opinion, rather than moving towards the other [16].

Distance. Second, the distance between the source and target of the comparison can differ. Festinger posed that comparisons are often made with targets that are similar to oneself [16], i.e. close comparison. The contrary is a far comparison in which the comparison target is dissimilar to oneself [25]. Festinger poses that close comparisons are more meaningful, as they represent a difference in abilities more accurately than far comparisons [16].

Distance can be related to the metric being compared. For example, far comparison would be comparing with someone with much higher grade, whereas a close comparison target scored similarly. Distance does not need to be related to the metric that is being compared. Choices for comparison targets are often based on other characteristics, such as demographics. Due to their similar characteristics, these people are expected to also have similar abilities [13].

Social comparison in (online) education

Social comparison has been thoroughly researched in traditional classroom settings, reporting both behavioural and psychological effects [13]. Technology provides new opportunities for education such as learning analytics, which is the collection, analysis, and reporting of learner data [29, 33]. Learning analytics often use social comparison as a reference frame for learners' progress and performance [23, 24]. However, learners have different attitudes and preferences regarding social comparison.

Usefulness of comparison. Despite its widespread use, the application of social comparison in learning analytics may not be appropriate. Jivet et al. [23] argue that social comparison defines learning too much as outperforming others rather than acquiring knowledge, and they advise to be careful with the use of social comparison without understanding individual differences [24].

Learners respond differently to social comparison. Research on different reference frames in learning analytics has shown that social comparison is the least preferred, compared to reference frames showing one's own goals or past achievements [19]. On the contrary, Huguette et al. [22] found that students responded more positively to social comparison than to personal reference frames. Qualitative research on different modules in learning analytics found that both peer comparison and personal performance were considered useful by most participants. However, two of the nineteen participants in this study indicated a dislike towards comparison and would rather not view it [1]. When combining personal and social comparisons, the latter receives more attention [27].

Preference in comparison direction. The influence of comparison direction and choice of comparison targets within educational settings has been thoroughly researched. Learners tend to choose upward comparison when deciding on a comparison target themselves [9, 22], which is in line with Festinger's theory [16]. Research by Joshi et al. [25] found that learners prefer upward comparison or find it equally preferable to downward. However, they found no strong preference for downward comparison on its own [25]. The research by Akhuseyinoglu et al. [3] investigated adaptive comparison ranges and also found a tendency for upward comparison. Several studies attributed this preference to self-improvement [8, 13, 22].

Research in both online and offline educational contexts has shown that social comparison preferences are not set in stone and that these vary as a result of performance or time [3, 22]. For example, the study on adaptive comparison ranges by Akhuseyinoglu et al. [3] found that students increased the range upward when their own performance improved. Furthermore, near the end of the course, they increased the range downward, adding more students who performed worse.

Design of social comparison

How social comparison can be communicated best is still open for question. No general guidelines exist, as the design depends largely on the information that

needs to be communicated. For example, textual descriptions can suffice for brief facts, whereas visualisations can convey complex data more efficiently. Existing research showed three overall themes: 1) specificity of comparison targets, 2) specificity of information, and 3) visualisation.

Specificity of comparison targets. Communication of a comparison is partially dependent on the specificity of the comparison target. In learning analytics dashboards, social comparison is most often shown as a comparison with the average of the peer group, with little to no detail on individual scores [23]. Comparison targets can be made more specific. Histograms, for example, contain a distribution of scores and can show where on this distribution the student lies [17].

More specific are ranks or leaderboards. Ranks can be non-specific, only visualizing them in text [6]. Leaderboards show the student's more specific position in comparison to all individuals [1,6]. Leaderboards can be adapted to include or exclude certain groups (e.g. only friends). In addition, individuals can be shown by name, though anonymous comparison is generally preferred [25].

Specificity of information. In addition to the specificity of comparison targets, the information displayed varies in detail. Some learning analytics dashboards show overall scores only, focusing on the bigger picture. However, they more often display performance on subdomains. Combinations also occur, such as in the Mastery Grids interface by Loboda et al. [28] (Figure 1).

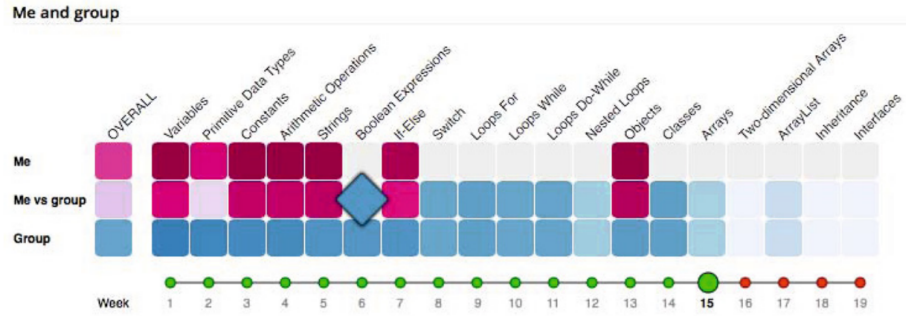


Fig. 1. Comparison in the Mastery Grids by [28]

Visualisation. Colour can be used to indicate absolute stance in comparison to the rest of the class. The student's score is often a different colour to differentiate between one's own score and that of the full group [17, 28]. Colour can also be used to indicate relative positions [25], using different colours for above, on

and below average scores. In addition, different intensities are used to show the distance from one's score to the average [2, 25, 28].

Social comparisons use different shapes. Bar or circle charts are commonly used, plotting the student's score against the average (Figure 2, Figure 3) [21, 27, 34]. Depending on the specificity of the information, several of these charts are shown, one for each subtopic. Spiders graphs can be used to present more information in a compact way [12].

Most often, combinations of colour, shape and text are used to visualise social comparison. For example, several studies use colour to differentiate between class average and the learner's score and use tiles to differentiate between different topics [2, 25, 28].

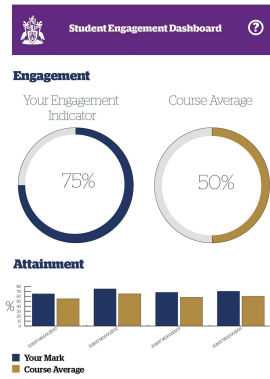


Fig. 2. Bar and circle charts by Harvey & Keyes [21]

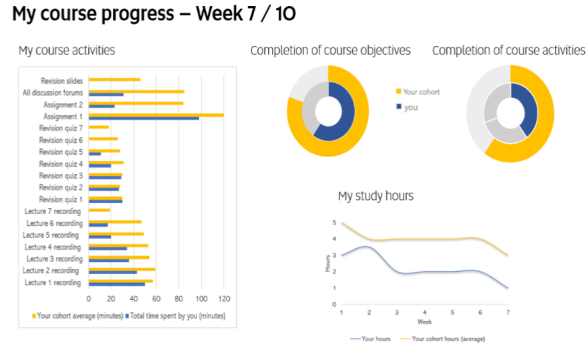


Fig. 3. Bar and circle charts by Lim et al. [27]

2.2 Affect

Social comparison is often used as a tool to foster motivation and engagement [18]. However, it can also impact other factors such as emotions in direct response to the comparison [1, 6]. These feelings can in turn affect for instance motivation, achievement and learning strategies [30].

Models for affect

Terminology around affect, emotion, moods, and feelings is not always clear in research, and the terms are often used interchangeably. In this research, the term affect is used. Affect describes feelings at any given time. It is always present, but it is not constant. It can fluctuate over time and in response to events. In comparison, emotions are short-lived, and are defined in more distinct categories [32].

Positive vs. negative affect First, affect can be classified in terms of positive and negative affect. Contrary to what the name suggest, positive and negative affect are not two ends of one dimension. Instead, affect can be scored on both levels, and positive and negative affect can co-exist [37]. Low positive affect indicates a state of tiredness, whereas high positive affect describes an energetic and pleasant state. Similarly, low negative affect indicates a sense of calm, whereas high negative affect describes distress [37, 40].

Positive and negative affect can be measured using adjectives. The PANAS Scales consists of two 10-item scales, one for positive affect and one for negative. Each scale consists of 10 adjectives. Participants rate to what extent they feel like the given words [37].

Dimensional model Second, affect can also be classified in terms of the dimensions valence and arousal (Figure 4). Valence ranges from pleasant to unpleasant, and arousal from activation to deactivation. According to the circumplex model of affect, all affective states can be captured on a two-dimensional plane by scoring them on both valence and arousal. More dimensions, such as dominance or motivation intensity, can be used [31]

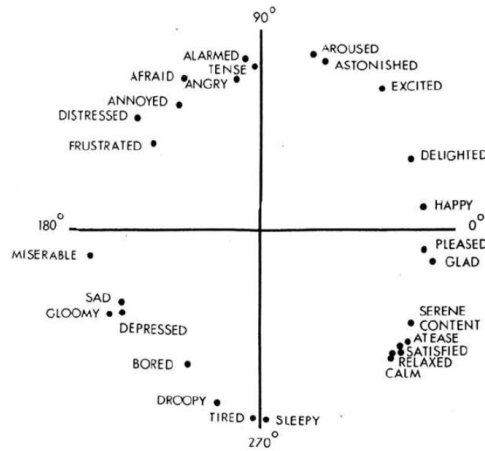


Fig. 4. Circumplex model of affect by Russell [31], the horizontal axis ranging from low to high valence (left to right), and the vertical axis from high to low arousal (top to bottom) .

The Self-assessment Manikin (SAM) is a widely used tool for measuring affect in terms of valence, arousal and dominance. It contains three scales, one for each dimension. Each scale contains five variations of the same figure, representing the levels on the scale. Participants are asked to select a level on the scale in a Likert-type fashion [10]. Though widely used, Betella and Verschure found the SAM to be outdated, and created and tested the Affective Slider as an alternative

(Figure 5) [7]. The Affective Slider measures valence and arousal using two slider controls, which are both accompanied by two emoticons, representing the outer ends of the scale. Two triangles are displayed beneath the slider to signify the strength of the emotion. The Affective Slider takes advantage of visual cues and digital interface elements and therefore requires little instructions before use.

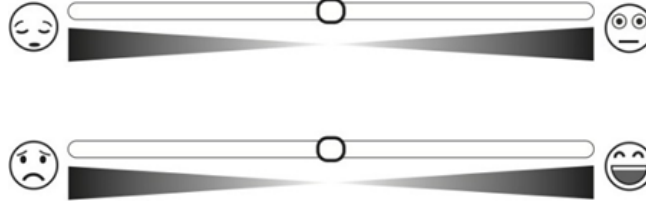


Fig. 5. Affective slider by Betella & Verschure [7].

Social comparison and affect

The comparison target can influence the affective response to social comparison. Research found that most participants' perception of their own performance changed from positive to disappointed after viewing an upward comparison [1]. This upward comparison can cause unpleasant feelings with varying levels of arousal. In the study by Lim et al. [11] participants reported feelings high in arousal such as anxiety and stress, and feelings low in arousal, such as depression. In contrast to upward comparison, downward comparison can make people feel better about themselves [39]. Furthermore, perceived importance of the task that is compared can impact the strength of these emotional effects [35].

However, the effects of comparison direction may change depending on how similar the comparison target is. Feeling similar to a comparison target can cause feelings of inspiration in upward comparison, whereas it can cause feelings of fear in downward comparison [13]. In addition, affective responses are only dependent on the comparison itself. Underlying factors such as expected performance can influence responses. For example, in the study by Bennett & Folley [6], participants who expected a high grade could still be disappointed for seeing a minority above them, whereas those who expected bad performance were content with being average.

2.3 Motivation

Social comparison can be used as a tool to boost motivation [13,18]. However, motivation is not only an effect, but also a predictor of preferences regarding social comparisons [13].

Achievement goals

Recent research has investigated how underlying motivation can affect preferences on social comparison [25] using the concept of achievement goals [15]. Elliot & McGregor [15] posed the 2x2 achievement goal framework, in which he distinguishes four types of achievement goals along two dimensions (Figure 6). The first dimension is performance versus mastery goals, which differ in the reference point of the goal. Performance goals focus on one's performance in relation to the performance of others, while mastery goals focus on performance compared to one's own past performance. Performance-oriented goals are more focused on outcomes, whereas mastery goals focus more on the process of learning and improving oneself [4, 14, 15].

		Definition	
		Absolute/ intrapersonal (mastery)	Normative (performance)
Valence	Positive (approaching success)	Mastery- approach goal	Performance- approach goal
	Negative (avoiding failure)	Mastery- avoidance goal	Performance- avoidance goal

Fig. 6. 2 x 2 Achievement Goal model by Elliot & McGregor [15].

The second dimension is approach versus avoidance goals. Approach goals work towards attaining something positive, whereas avoidance goals focus on moving away from something negative. Rather than wanting to achieve something, as with approach goals, avoidance goals focus on not wanting to lose something already attained [14, 15].

The model initially grouped mastery-approach and mastery-avoidance together, creating three subscales. Elliot & McGregor separated this further to attain four motivational goals: 1) performance-approach, 2) performance-avoidance, 3) mastery-approach, and 4) mastery-avoidance. Performance-approach goals focus on doing better than others, whereas performance-avoidance goals are focused on not doing worse. Mastery-approach goals focus on improving one's own abilities, whereas mastery-avoidance goals focus on not losing that ability or not missing any information [14, 15].

Social comparison and achievement goals

Research shows that all four achievement goals from Elliot & McGregor's model are linked to a tendency to socially compare [9], though perception of social comparison dashboards varies between the achievement goals. Research found that

people high in performance-approach are more comfortable with downward comparison [25]. Similarly, a different study found that those high in performance-avoidance engaged more in downward comparison [3].

Rather than testing personal achievement goals of participants, the research by Bounoua et al. [9] used achievement goals to frame the context of the study by phrasing the instructions in line with the goals. Overall, students chose to compare themselves to an upward target after performing a task, regardless of the condition they were in. However, they did find that those in the performance-avoidance group were less likely to choose an upward comparison, compared to all other groups.

Current implementations of learning analytics may focus too little on underlying achievement goals. With social comparison being the most common way to frame data, learning analytics are more catered towards performance-oriented individuals [23].

2.4 Research opportunities

The above literature indicated that social comparison is often included in learning analytics [23]. Yet, learners have individual preferences and responses regarding social comparisons. Recent research has investigated giving learners more autonomy over social comparison dashboards, by introducing adaptive comparison ranges [3].

However, little is known about what people do when they get the choice to compare themselves and why. In traditional education, several studies reported participants not choosing a comparison target, assuming this was due to a lack of interest in social comparison [8, 22]. In addition, studies in which participants could engage with social comparison in learning analytics reported participants not using these features, though the cause for this lack of engagement was not further analysed [5, 18].

Research question 1

The above literature inspired the first research question: *How do comparison direction, affective state, and achievement motivation affect the choice to compare oneself?* Three dimensions were used to explore a wide range of possible predictors: 1) social comparison direction, which is part of the design of the comparison itself, 2) emotion, which is a short-term response, and 3) motivational style, which is an underlying concept separate from the comparison itself.

This study explores the possible predictors of comparison choices. Despite its explorative nature, the following considerations can be made in trying to predict these effects. Based on the definition of the achievement goal model [15, 23], it is expected that those high in performance orientation are more likely to compare, whereas those high in mastery orientation are less likely to compare. No strong predictions were made on approach and avoidance motivations, and the aim is to explore any possible differences.

The effects of arousal and valence on choice behaviour are difficult to predict. This is because, upward comparison can cause unpleasant feelings of varying levels of arousal [1,11], yet it is still preferred over downward comparison [9,20,22], which can induce pleasant feelings [39].

Upward comparison may lead to higher odds of choosing comparison, as this is a natural tendency [16]. However, downward comparison can make us feel better about ourselves [39], which in turn could lead to wanting to view the comparison again.

Research question 2

In addition, this paper investigates whether previously tested factors affect response to comparison when that comparison is self-chosen. This led to the second research question: *How do comparison direction and achievement motivation affect affective state after self-chosen comparison?*

More is known about responses to social comparison than about the predictors of choice. However, prior research has shown these responses are also very personal and difficult to predict [6]. Based on prior research, upward comparison is expected to result in lower valence scores than downward comparison. The effects on arousal are difficult to predict. This is because upward comparison can lead to both anxiety and sadness [1,11], both of which are high in valence, but are different levels of arousal [14]. Similarly, downward comparison can lead to relief or pride [38], which are high in valence, but differ in arousal.

It is expected that those high in performance orientation experience stronger effects than those high in mastery orientation, as by definition their goals focus on performance in regards to others [14]. No strong predictions were made on approach and avoidance motivations, and the aim is to explore any possible differences.

3 Materials and methods

An experimental study was designed to investigate the interaction between emotional state and social comparison choice behaviour in the context of a cognitive task, a Dutch version of the New York Times Spelling Bee. A custom webapp was built to facilitate the experiment. First, a pilot study was performed to guide the design of the experiment task and social comparison frame. Next, the experiment was performed. This study was approved by the moderator of the Information Sciences bachelor thesis (Appendix A).

3.1 Experiment task and design process

As education is difficult to mimic in a short experiment, a low-stakes cognitive game was used as an alternative. Participants were provided with a Dutch version of the New York Times Spelling Bee(Figure 7) ¹, in which players need to find

¹ New York Times Spelling Bee

as many words as possible with a set of given letters. The Spelling Bee was used as it requires little explanation and no pre-existing knowledge, while still being cognitively stimulating. Solutions for the game were generated using the word list by OpenTaal ².

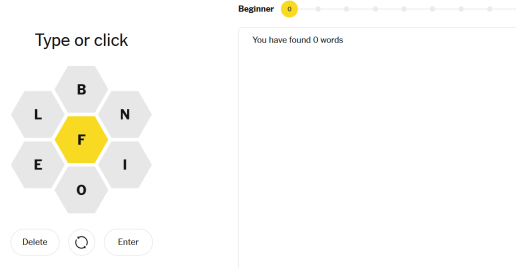


Fig. 7. New York Times Spelling Bee.

Pilot study

A pilot study was conducted to guide the design of the social comparison. Five participants took part in the pilot study. All were university students and were acquaintances of the researcher. Four participants were interviewed on location, and one remotely via a video call.

After signing the informed consent (Appendix B), the concept of the game was explained using the New York Times Spelling Bee (Figure 7). Next, participants were asked questions regarding their willingness to compare scores, and the preferred metrics for this in semi-structured interviews (Appendix B).

The interviews lasted between 5-10 minutes each. Audio of the interviews was recorded using a code-protected mobile phone, including the remote interview. Data was transferred to OneDrive for analysis. Transcripts were generated using Microsoft's Transcribe and adjusted where needed. Transcripts were coded to find recurring patterns.

Results for design

Two topics guided the design of the experiment task: 1) which statistics to show, and 2) when to show these. Results unrelated to the design are presented in section 4.1.

Statistics. All participants indicated a desire to see their own game statistics in some form. Two participants wanted to see their statistics to improve themselves. Metrics regarding the words, such as the number of correct guesses or the total

² OpenTaal License

number of words in the puzzle, were mentioned by four participants. Participants had mixed opinions about time metrics. One indicated to want to see time, while two mentioned it would be stressful or irrelevant to the game. One had mixed opinions, as it could be both interesting and stressful.

When to show statistics. One participant specifically mentioned wanting to see guessed words during the game, as these are useful for the gameplay. Participants had varying opinions about other statistics such as score and time. One thought that constantly showing would have the same effect as displaying it at (random) intervals. Three participants indicated a preference to see statistics at the end to prevent extra stress. One participant thought having free choice at all times would be the best option.

Conclusion and implications for design

Participants wanted to see statistics on their own performance, specifically regarding progress (e.g. score, total words) and performance (e.g. correct guesses) in the game. As a result, two metrics were included in the social comparison: score and percentage of correct guesses (Figure 9).

The score was visualised in relation to the maximum possible score as a progress measure. The percentage of correct guesses was visualised similarly, serving as a performance-oriented measure, as it measures how often a correct answer was given.

The visualisation was minimalistic to minimise confounding variables. Blue was chosen as a neutral alternative to colours like red and green, which are commonly associated with good and bad. The score of the average player was visualised with a vertical line, accompanied by the written score for clarity.

Guessed words were deemed useful for the game, and were therefore visible during gameplay (Figure 8). Participants' own scores and correct words were displayed alongside the game to prevent participants from choosing to view the social comparison out of curiosity for their own score (Figure 8).

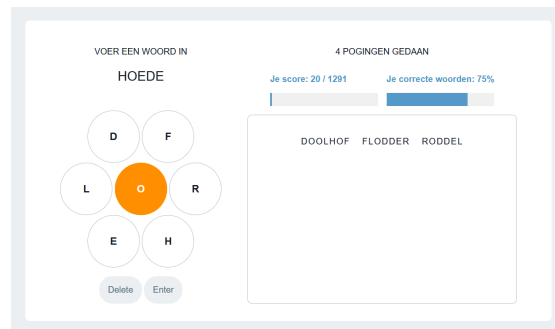


Fig. 8. Task interface.



Fig. 9. Comparison interface.

3.2 Experimental study

Design

The effects of affective state, achievement goals and social comparison direction on choice behaviour in social comparison were investigated in an online experiment.

All participants went through the same process (Figure 10). Participants first answered questions on demographics and achievement goals. Moreover, their current levels of valence and arousal were measured to establish a baseline.

After the questionnaire, an explanation of the task was given. In the main part of the experiment, participants went through a cycle of the following activities: 1) submit eight words in the game, 2) measure emotional state, 3a) present choice for social comparison, 3b) view comparison if applicable, 4) measure emotional state. This cycle was repeated three times. Participants were divided into two groups: upward and downward comparison. Based on the condition, the social comparison showed the average player performing above (upward) or below (downward) the participant.

Lastly, participants answered two open questions regarding the reasoning behind their choices for social comparison. Participants could return to the experiment. They received a username to log in with. Baseline affect was measured, after which the returning users proceeded to the game. The rest of the process was identical to that of first-time users.

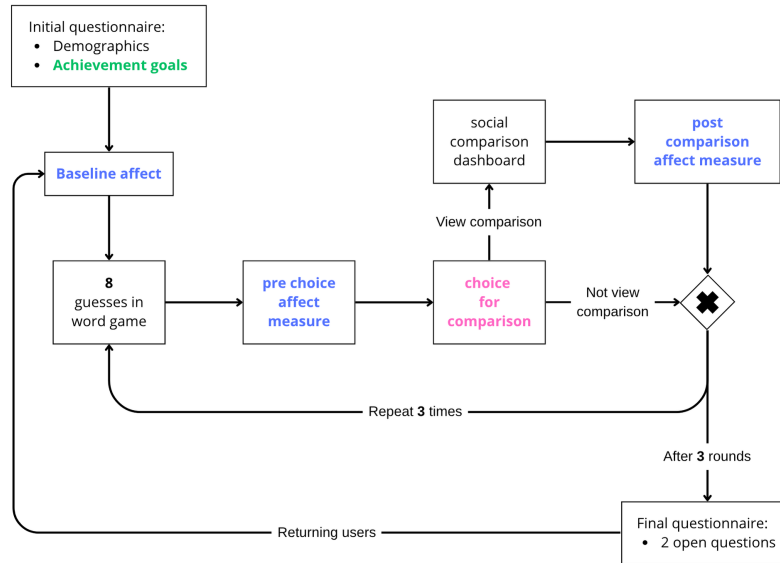


Fig. 10. Flowchart of the experiment setup.

3.3 Participants

In total, 96 participants with an average age of 22.6 (SD = 4.7, Min = 17, Max = 50) were recruited (male = 41, female = 49, other = 6). Most participants were students ($n = 75$). Convenience sampling was used, mainly due to time constraints. Though the aim was to recruit mostly students, non-students were also included in the research. Participants were recruited via WhatsApp, LinkedIn, face-to-face, and snowballing. Participants needed about 5-25 minutes to complete the experiment.

Measures

Social comparison direction Participants were randomly assigned to one of two conditions: 1) upward or 2) downward comparison. Average player scores were generated 8-12% above (Figure 12) or below (Figure 11) the participant's score. This margin was small enough to seem realistic for participants, yet big enough to visibly show a distance between the participant and the average. To enhance this distance, upward scores were plotted several pixels to the right, whereas downward scores were plotted several pixels to the left.

Participants were not informed beforehand that the comparison data was fake. At the end of the experiment, participants received a link to the results page, where the results and underlying goal would be explained two months after the experiment. This way, they were informed afterwards without disclosing the goal to potential returning users.



Fig. 11. Downward comparison.



Fig. 12. Upward comparison.

Achievement goals. Achievement goals were measured according to the 2x2 achievement goal framework [14], as it has been previously investigated in the context of social comparison. For non-students, additional instructions were

added to read the questions in the context of a previous study or work-related learning environment. A validated Dutch translation of the Achievement Goal Questionnaire (AGQ) was used [15, 36]. A control question was added to check for attentiveness of participants. All questions were measured on a five-point Likert scale.

Affect. Affect was measured according to the dimensional model, using the Affective Slider by Betella & Verschure [7]. This measure was quick and intuitive, allowing for several repetitions during the experiment without significantly extending the experiment’s duration.

Social comparison behaviour. Social comparison behaviour was measured in each round. The choice to view the comparison and the time needed to make this decision were logged, as this could provide insight into the choice process. Time spent viewing the comparison was logged if applicable, as this could provide insights into the attention people paid to the comparison.

Qualitative data. Two open questions were asked in the final questionnaire to gather qualitative data: 1) *Why did you do want to compare yourself?*, and 2) *Why did you not want to compare yourself?*. See Appendix C for all experiment materials.

Procedure

Data collection. Data was collected via a custom-built web app, which was hosted on a Virtual Private Server from Strato. Hosting via this third party was approved by the ethical moderator, as no personal data was collected. The database was stored on the servers, to which only the researcher had access. After the experiment, data was removed from the server and stored locally on a password-protected laptop.

New participants were directed to the informed consent form (Appendix C). The initial questionnaire contained demographic questions to gain insight into the participant pool. Next, the AGQ was administered and the Affective Sliders were used to measure baseline affect and arousal.

Participants first received instructions on the game, after which they went through three rounds. For each round, several statistics were logged. Game statistics were logged for potential analysis and for returning users. Both pre-scores on valence and arousal were logged. To measure decision behaviour, the choice (view or not-view comparison) was noted, as well as the time needed to make this decision. Lastly, the time spent viewing the comparison and the post-scores for the affect measures were recorded if applicable.

Pre-processing. In total, 110 respondents finished the experiment. Participants who answered the control question incorrectly ($n = 3$), whose data was incomplete ($n = 1$), who took exceptionally long for the full experiment (> 30 minutes, $n = 8$), who took exceptionally long for the initial questionnaire (> 20

minutes, $n = 1$) were excluded. One participant did not partake seriously based on the guessed words and answers to the open questions and was therefore excluded. This left 96 participants for data analysis.

Each subscale of the AGQ was scored by taking the average of the three corresponding items. Internal consistency was measured using Cronbach's alpha. Differences in pre- and post-comparison arousal scores were determined if applicable. For each participant, the total number of comparisons viewed, the average time spent deciding to compare, and the average time spent viewing the comparison were determined.

Quantitative data analysis. To assess which factors predict choice for social comparison, the full dataset was filtered to include only data points where the participant viewed the comparison at least once before. Otherwise, the comparison direction could not have been predictive. For each participant, the number of comparisons was transformed to the percentage of comparisons viewed. Absolute numbers of comparisons would not properly represent the choices following the comparison direction. Participants who viewed the comparison in the first round could base their choice for comparison on the direction two more times. Participants who first viewed the comparison in the second round could only base one more choice on the comparison direction. This is why percentages were chosen over absolute numbers.

Several generalized linear mixed-effects models were fitted to the data to determine the predictive value of arousal, valence, motivational goals, and condition on comparison choice. In addition, the difference in percentage of viewed comparisons between the upward and downward conditions was assessed using a Mann-Whitney U test as a non-parametric alternative to an independent samples t-test.

To determine the effect of the comparison on affective state, the full dataset was filtered to include only the data points where the participant viewed the comparison. Several linear mixed-effects models were fitted, taking into account pre-choice affective state, achievement goals, and social comparison direction. To further assess the difference in pre- and post-comparison arousal and valence, several Wilcoxon Signed-Rank tests were performed as a non-parametric to a paired samples t-test. The data was split into upward and downward subsets for these tests. Tests were performed individually for all three rounds. A final test was run on the combined data from all rounds.

All analyses were performed twice, including and excluding non-students to determine if non-students affected the results.

Qualitative data analysis. Qualitative data was analysed using a thematic analysis. Following Braun & Clarke [11], several rounds of coding were done to find recurring themes and group responses together. The occurrences of themes were tallied to identify the most common ones.

4 Results

4.1 Pilot study

Besides guiding the design of the experiment task, the pilot study was used to gain insight into the following topics: 1) desire to compare, 2) effect on emotion, 3) effect on motivation, and 4) effect in a different context.

Desire to compare

All participants indicated a desire to compare, though for different, and sometimes multiple, reasons. Three participants mentioned curiosity or fun, and three mentioned competitiveness. Two participants specifically mentioned comparison with friends. One participant mentioned he was curious about how the top players performed. He wants to understand why these people are at the top.

Effect on emotion

All participants gave a similar response: being above average would induce a positive feeling while scoring below average would cause a negative feeling. That being said, most indicated that it would not be a long-lasting feeling, as it is just a game. They would be able to let it go easily.

Effect on motivation

One participant indicated that comparison would be motivating to keep improving oneself, while others figured that scoring below others would be demotivating. However, one participant specifically indicated that scoring at the top would be demotivating, as there is no room to improve. The activity of other players was also mentioned as a factor influencing motivation.

Effect in different context

In a different context, such as education, participants would respond differently to social comparison. They expected the effects on emotions to have the same direction (where scoring below average leads to negative feelings and scoring above average leads to positive feelings). The situations differ in the intensity of the effect. Participants expected the effects to be stronger and longer lasting in a high-stakes context. However, participants expected weaker effects regarding motivation. One participant mentioned that initial motivation is higher and stronger for education. This would make it more difficult to demotivate. One participant expressed a strong dislike towards comparison of any kind in educational settings, whereas she did like it in the game. She thought education was not the appropriate setting for social comparison.

4.2 Experiment results

Of the 96 participants, 51 were in the downward condition, and 45 in the upward. 69.4% of participants decided to compare themselves in every round, and 4.1% never compared themselves (Figure 13, Appendix D.1). Comparison patterns were determined for the 26.5% who did not compare themselves all three times (e.g. 001 means not viewed in round one and two, viewed in round three) (Table 1, Table 2). Participants spent an average of 2.8 seconds on choosing to compare. This decreased over the rounds. On average, they viewed the comparison for 5.7 seconds. Both decreased over the rounds (Table 3, Table 4). Two participants returned to the experiment once. However, this was insufficient data to perform analysis on returning users.

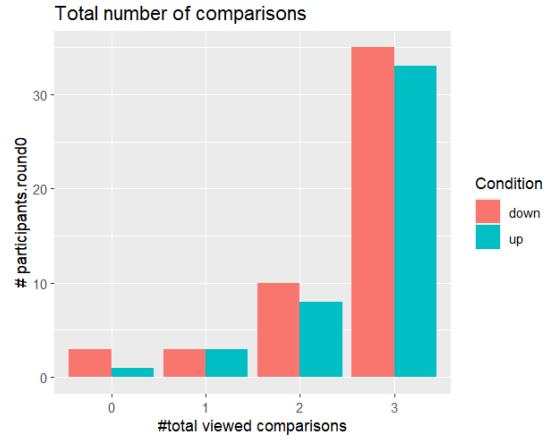


Fig. 13. Total number of comparisons, separated by condition

Table 1. Comparison patterns for participants who viewed the comparison once

	Pattern			Total
	001	010	100	
Downward	0	1	2	3
Upward	2	0	1	3
Total	2	1	3	6

Table 2. Comparison patterns for participants who viewed the comparison twice

	Pattern			Total
	011	101	110	
Downward	4	4	2	10
Upward	1	2	5	8
Total	5	6	7	18

Table 3. Time spent making the choice to compare in seconds

Round	Mean	SD	Min	Max
All ($n = 288$)	2.8	2.1	0.6	15.3
1 ($n = 96$)	4.5	2.3	1.9	15.3
2 ($n = 96$)	2.1	1.1	0.7	6.5
3 ($n = 96$)	1.9	1.5	0.6	8.6

Table 4. Time spent viewing the comparison in seconds

Round	Mean	SD	Min	Max
All ($n = 246$)	5.7	3.5	1.5	32.7
1 ($n = 84$)	7.4	3.0	2.8	20.2
2 ($n = 81$)	5.1	4.0	2.0	32.7
3 ($n = 81$)	4.4	2.7	1.5	17.3

4.3 Effect on choice

The analysis on comparison choice used the first filtered set ($n_{participants} = 90$, $n_{datapoints} = 174$). Alpha scores for all but performance avoidance were good ($\alpha > 0.8$). Performance avoidance was lower ($\alpha = 0.694$), but still acceptable. Average scores were comparable between the upward and downward comparison. However, performance avoidance scores were higher on average in the downward condition. No major differences in average and alpha scores occurred after excluding non-students ($n_{participants} = 70$, $n_{datapoints} = 135$). Again, only performance avoidance had a lower alpha score ($\alpha = 0.707$). Scores for the four subscales were distributed similarly between the conditions. The difference in performance avoidance was smaller than when including non-students (Appendix D.2).

Total number of comparisons per condition

The upward group viewed the comparison 89.3% ($SD = 23.5$) of the time, whereas the downward group viewed it 87.5% ($SD = 28.2$) of the time. However, Mann-Whitney U test showed this difference was not statistically significant ($W = 1003.5$, $p = 0.96$). Using only the students subset, upward had a mean of 89.1% ($SD = 24.5$) and downward a mean of 88.2% ($SD = 24.5$). Again, no significant differences were found between the two groups ($W = 607.5$, $p = 1$).

Generalized linear mixed-effects model

First, generalized linear mixed-effects models with only one predictor were ran to predict whether a comparison was viewed (1) or not (0). The participant was added as random effects term. One predictor was significant: pre choice arousal ($\beta = 0.04$, $p = 0.01$). For every increase in arousal pre choice, the odds of the participant choosing to view the comparison increases with 4.2% (Table 5). Next, a generalized linear mixed-effects model was fitted, using pre choice arousal and valence, the four motivational goals, and condition as predictors. The scale variables were standardized. No significant predictors were found in this combined model (Appendix D.2).

The same procedure was repeated for the student subset only. This resulted in two significant predictors: pre choice arousal ($\beta = 0.04$, $p = 0.02$), and mastery avoidance ($\beta = 1.16$, $p = 0.03$). For pre-choice arousal, this is similar to the full

Table 5. Coefficients for choice predictors including non-students

model	term	β	std.error	statistic	p	e^β
m1	Performance approach	-0.003	0.80	-0.004	1.00	1.00
m2	Performance avoidance	-0.07	0.67	-0.11	0.92	0.93
m3	Mastery approach	0.22	0.81	0.27	0.79	1.24
m4	Mastery avoidance	0.86	0.71	1.21	0.22	2.37
m5	conditionup	0.14	1.29	0.11	0.91	1.15
m6	Post arousal (last round)	0.02	0.03	0.64	0.52	1.02
m7	Post valence (last round)	-0.03	0.03	-0.97	0.33	0.97
m8	Pre arousal	0.04	0.02	2.48	0.01	1.04
m9	Pre valence	0.05	0.03	1.59	0.11	1.05

dataset, where every increase in arousal increases the odds of comparison with 3.9%. Mastery avoidance, on the other hand, has a much larger effect, with every unit-increase in mastery avoidance resulting in a 219.0% increase in the odds of choosing comparison (Table 6). As with the full dataset, the combined model did not result in significant predictors (Appendix D.2).

Table 6. Coefficients for choice predictors excluding non-students

model	term	β	std.error	statistic	p	e^β
m1	Performance approach	0.31	0.88	0.35	0.72	1.36
m2	Performance avoidance	0.19	0.74	0.26	0.80	1.21
m3	Mastery approach	0.43	0.86	0.50	0.62	1.54
m4	Mastery avoidance	1.16	0.55	2.12	0.03	3.19
m5	Condition (up)	0.06	1.46	0.04	0.96	1.07
m6	Post arousal (last round)	0.02	0.04	0.62	0.54	1.02
m7	Post valence (last round)	-0.05	0.04	-1.19	0.23	0.95
m8	Pre arousal	0.04	0.02	2.26	0.02	1.04
m9	Pre valence	0.04	0.04	1.18	0.24	1.04

4.4 Effect of comparison on affect state

The analysis on effect of the comparison used the second filtered set ($n_{participants} = 92$, $n_{datapoints} = 246$). Within the subset, internal consistency of all but performance-avoidance were good ($\alpha > 0.8$). Performance avoidance was lower ($\alpha = 0.686$), but still acceptable. Average scores were comparable between the upward and downward comparisons. However, performance-avoidance scores were higher on average in the downward condition. No major differences were found in average and alpha scores after excluding non-students ($n_{participants} = 72$, $n_{datapoints} = 192$). Again, only performance avoidance had a lower alpha score ($\alpha = 0.695$). Scores for the four subscales were again distributed similarly, with only a larger difference in average in performance-avoidance (Appendix D.2).

Linear mixed-effects model

First, linear mixed-effects models with only one predictor were fitted to predict arousal after comparison. The participant was added as a random effects term, and the initial and pre-arousal variables were added as covariates. The same procedure was repeated for valence. For both arousal and valence, the coefficient for the condition was significant. None of the achievement goals was significant (Appendix D.3, Appendix D.3).

Finally, a linear mixed-effects model was run using all motivational goals and the condition, again accounting for the participant and initial and pre-comparison affect scores. The condition term was significant for the valence model ($\beta_{conditionUp} = -18.6$, $p < 0.05$), and for the arousal model ($\beta_{conditionUp} = -10.4$, $p < 0.05$) (Appendix D.3, Appendix D.3). Being in the upward condition predicted lower scores for valence and arousal than being in the downward condition. The effect was stronger for valence than for arousal.

Table 7. Coefficients of the mixed model for valence, including non-students.

term	β	std.error	statistic	p
(Intercept)	69.22	1.74	39.73	<0.01
Initial valence*	5.25	1.35	3.90	<0.01
Pre valence*	8.30	0.96	8.69	<0.01
Performance approach*	1.13	1.30	0.87	0.39
Performance avoidance*	0.29	1.29	0.23	0.82
Mastery approach*	0.26	1.29	0.20	0.84
Mastery avoidance*	1.31	1.26	1.04	0.30
Condition (up)	-18.34	2.58	-7.10	<0.01
SD (Intercept)	10.04			
SD Observation	9.89			

Note: starred terms are standardized.

The same procedure was repeated using the student subset. Again, only the condition was a significant predictor both in the single-predictor linear mixed-effects models, as in the combined model. As with the full dataset, being in the upward condition predicted lower valence ($\beta_{conditionUp} = -19.7, p < 0.05$) and arousal scores ($\beta_{conditionUp} = -10.2, p < 0.05$).

Differences pre and post-comparison

Plotting valence and arousal scores over the measures, it was observed that, on average, valence increased after each comparison in the downward condition (Figure 14, Figure 15), whereas it decreased in the upward condition. A similar trend was for arousal in the downward condition, though in the upward condition, scores remained similar over all measures.

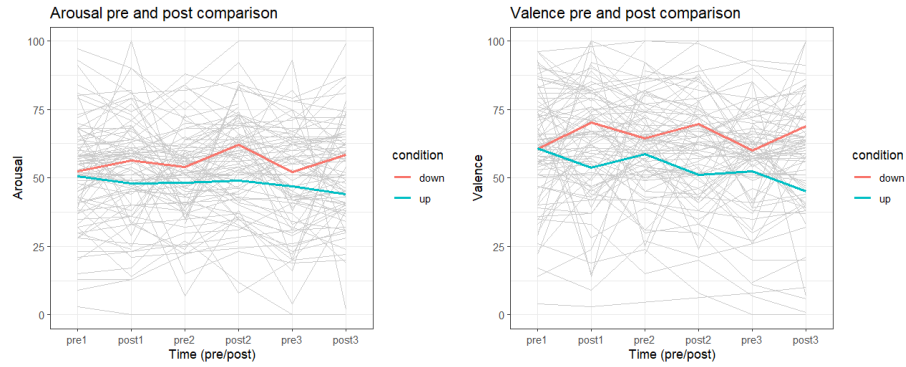


Fig. 14. Arousal pre and post over three rounds. **Fig. 15.** Valence pre and post over three rounds.

Wilcoxon signed-rank tests were performed to determine differences between pre and post-comparison scores in both groups for all rounds (Table 8, Table 9, Table 10). In all rounds, arousal did not significantly change in the upward condition. Furthermore, arousal did not change significantly after viewing a downward comparison in the first round. However, in the second and third round, arousal significantly increased after downward comparison. Valence significantly increased in the downward condition in all three rounds, whereas it decreased in the upward condition.

As seen in the regression, downward comparison affected valence and arousal positively. Valence and arousal were on average similar between conditions before the first comparison ($\Delta mean_{arousal} = 1.8$, $\Delta mean_{valence} = 0.3$). This difference was much larger after the third comparison ($\Delta mean_{arousal} = 14.4$, $\Delta mean_{valence} = 23.8$). In addition, changes in valence remained similar over the

rounds in downward comparison. However, in the upward condition, pre and post-comparison scores decreased over rounds.

Using the student subset, only one difference was found: valence did not significantly change after a downward comparison in the second round ($V = 171$, $p = 0.13$). The same trends regarding differences in means appeared (Appendix D.3).

Table 8. Results of Wilcoxon signed-rank tests - Round 1.

Name	Statistic p		Mean pre	Mean post	Mean difference
Upward arousal	443	0.46	50.6	48	-2.6
Downward arousal	255.5	0.06	52.4	56.3	4.0
Upward valence	541.5	<0.01	60.9	53.8	-7.0
Downward valence	166	<0.01	60.6	70.1	9.4

Table 9. Results of Wilcoxon signed-rank tests - Round 2.

Name	Statistic p		Mean pre	Mean post	Mean difference
Upward arousal	278.5	0.98	48.3	49.1	0.8
Downward arousal	146.5	<0.01	54.0	62.1	8.0
Upward valence	495.5	<0.01	58.7	51.1	-7.6
Downward valence	219	0.03	64.5	69.6	5.2

Table 10. Results of Wilcoxon signed-rank tests - Round 3.

Name	Statistic p		Mean pre	Mean post	Mean difference
Upward arousal	381	0.28	46.9	44.1	-2.8
Downward arousal	171.5	<0.01	52.1	58.5	6.4
Upward valence	468	<0.01	52.5	45.1	-7.4
Downward valence	131	<0.01	59.9	68.9	9.0

Leaving repeated measures over rounds out of consideration, similar effects were found (Table 11). Arousal significantly increased after downward comparison ($V = 1696$, $p < 0.01$), though no significant change was found after upward comparison ($V = 3230$, $p = 0.29$) (Figure 16). Valence increased significantly after downward comparison ($V = 1509$, $p < 0.01$), whereas it decreased after upward comparison ($V = 4417$, $p < 0.01$) (Figure 17). As the regression showed,

downward comparison led to higher scores. The biggest difference is seen in post-comparison valence, where the average scores differed by 19.4 Table 11. The same trends appeared for the student subset (Appendix D.3).

Table 11. Results of Wilcoxon signed-rank tests - All rounds combined.

Name	Statistic	p	Mean pre	Mean post	Mean difference
Upward arousal	3230	0.29	48.7	47.1	-1.6
Downward arousal	1695.5	<0.01	52.8	59.0	6.1
Upward valence	4417	<0.01	57.5	50.1	-7.3
Downward valence	1508.5	<0.01	61.6	69.5	7.9

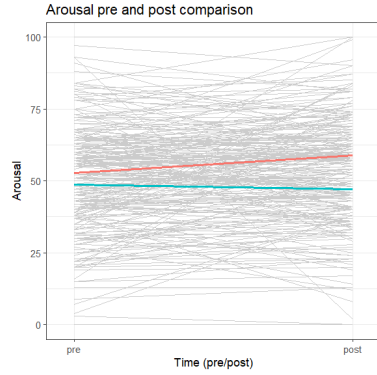


Fig. 16. Arousal scores pre and post-comparison.

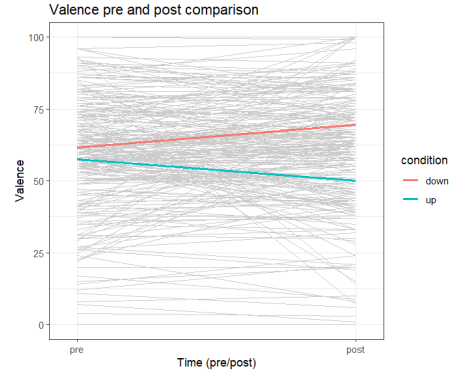


Fig. 17. Valence scores pre and post-comparison.

4.5 Qualitative data

Reasons to compare

Four participants did not compare themselves at all, and all four did not provide an answer to the question. Four other participants were classified as ‘other’, giving reasons such as thinking the comparison would mean the end of the experiment. Curiosity was often mentioned, though in 18 cases, it was not specified what the participant was curious about. The remaining 72 responses were coded according to six themes (Table 12). Some responses could fit multiple themes. See Appendix E for the original Dutch quotes from participants.

Reference for score. Most participants used the comparison as a reference frame for their own score. Subtle differences were observed in the phrasing of

Table 12. Occurrence per theme for viewing the comparison

Theme	Downward (n=49)	Upward (n=45)	Total (n=94)
Reference for score	26	21	47
Curious about others	5	9	14
Validation	11	1	12
Difficulty of the game	5	7	12
Self-improvement/motivation	3	4	7
Competitiveness	3	1	4

this reference frame. Some participants phrased it as wanting to see if they did better, while others phrased it more generally, wanting to know where they stand compared to others. One participant stated: *“I wanted to see if I did better than the average”*, and another said: *“I wanted to see how I was doing compared to others to see if my score was better or worse than average”*.

Curiosity about others. Curiosity about other people’s performance was often mentioned. The focus was on how others performed, rather than how the participants performed compared to them: *“Because I was curious about how the others made the puzzle”*.

Validation. Several participants used the comparison to seek validation for their own (perceived) score. Interestingly, this topic occurred a lot more in downward comparison than it did in upward comparison (Table 12).

Again, differences could be found in the phrasing. Some used this validation to confirm a good score: *“Because I thought it went well and I was looking for that confirmation.”*. One participant specifically mentioned that scoring above average made her feel better: *“I wanted to see how I did compare to others to see if I met the average. If I do better, then I feel like I did well.”*.

Not only good scores were validated. The comparison was also used to seek out validation of (perceived) low scores *“Because I had the feeling it was not going amazing and to then see that others are scoring similarly takes away from the feeling of being less”*.

Difficulty of the game. Several participants mentioned the difficulty of the game as a reason to compare. This was often combined with curiosity about others, desire to compare score, and/or validation. Participants wanted to see if others were also struggling to find words as well. *“I was curious how I did, I noticed I found it quite difficult to find words with more than three letters.”*.

Self-improvement. Several participants mentioned self-improvement and/or motivation. They used the comparison to evaluate whether they needed to step up their game: *“It motivates me to keep going at the same pace, or to step up a notch.”*.

Competitiveness. Last, competitiveness was mentioned by four participants as a reason to compare. They described a desire to ‘win’ the game.

Reasons to not compare.

68 participants compared themselves every round. However, 25 of them still provided an answer to the question, providing potential reasons why they would not compare/why they doubted. Three main themes were found: bad scores, keeping the focus on oneself, and not caring about the comparison (Table 13). No major differences were observed between the upward and downward condition.

Table 13. Occurrence per theme for not viewing the comparison

Theme	Not viewed all		Viewed all		Total
	Down (n = 16)	Up (n = 12)	Down (n = 13)	Up (n = 11)	
Bad scores	7	5	8	9	29
Focus on oneself	4	2	1	0	7
Did not care	1	4	2	1	8

Avoiding bad scores. The most common reason to not compare was due to (expecting) below-average scores. Participants were insecure or afraid to see a lower score. Some expected a bad score, and did not want confirmation: *“Because I thought I performed below average and I did not want to be confronted with that.”*.

In addition, participants did not compare, because they expected or experienced negative feelings after comparison as a result of scoring lower before. One participant stated: *“I thought I did quite well, but when I saw that I did less good than others, I felt less good about my own ‘achievement’.”*. Moreover, several participants stopped caring about the comparison after receiving a low score: *“I was less good than average so then it does not matter anymore”*.

Focus on oneself. The second theme was keeping the focus on oneself. Several participants indicated they did not want to view the comparison, because they wanted to try the game for themselves first, without getting distracted by the performance of others: *“Because I am trying to achieve a high score for myself. What others do does not matter yet, that will come later when I am actually good at it.”*. This theme mainly occurred among participants who did not view all comparisons.

Did not care. The third theme was participants not caring about comparison. This was focused towards the comparison in general (not caring enough for the experiment), as well as not attaching value to the comparison. One participant did compare every time but indicated that the actual score did not matter that much to them.

5 Discussion

In the following section, I interpret the results and answer the research questions of this study. The implications for social comparison in learning analytics and opportunities for future research are discussed.

5.1 Comparison is the most common choice

RQ1 was concerned with the possible predictors for choosing to compare oneself. The results show that the majority of participants compared themselves all three times, with curiosity being a common reason for comparison. Novelty bias [26] may have caused this tendency to view the comparison, and the effects possibly differ if this sense of novelty wears off. In addition, the experiment was short and the comparison could be quickly viewed, lowering the bar to choose comparison. Besides curiosity, the qualitative results showed that comparison is often used as a reference frame for one's own score, which is in line with Festinger's theory [16].

Generalized linear mixed-effects models indicated that arousal had a slight positive effect on the choice to compare oneself when no other predictors were included. Feeling activated can induce feelings of urgency or motivation to keep going or improve, whereas low arousal may disengage or demotivate people [30]. This possibly explains the positive effect of arousal. Similarly, qualitative results showed that several participants did not care and therefore did not compare themselves.

Unexpectedly, the data showed a significant and strong effect of mastery avoidance on the choice to compare. Based on the nature of the achievement goals [15, 23], it was expected that high scores on the performance orientations would positively predict this choice, rather than mastery. However, when combining multiple predictors, coefficients for both arousal and mastery avoidance were not significant.

5.2 Social comparison direction affects affective state

RQ2 was concerned with the effect of achievement goals and comparison direction on affective state. No significant effects of the achievement goals were found. However, social comparison direction significantly affected both valence arousal. Downward comparison correlates with higher scores on both scales.

As expected, valence increased after downward comparison, which is in line with downward comparison theory [39]. In all rounds, valence decreased after viewing an upward comparison, which is in line with prior studies in which upward comparison caused unpleasant or negative feelings [1, 11]. Arousal increased after viewing downward comparison in two out of three rounds. However, arousal did not significantly change after upward comparison. Prior research found that viewing an upward social comparison caused emotions of both high and low arousal [11]. Individual responses may have been stronger, yet varied in directions.

The results showed that arousal and valence were on average similar before the first comparison. However, over time the difference in average scores increased for both valence and arousal. A difference was observed between upward and downward comparison for valence scores. Valence increased after viewing a downward comparison. It decreased before the next comparison, and increased again after. A reversed pattern was observed after viewing an upward comparison. However, valence did not recover between the second and third round. A possible explanation is that participants were motivated to improve themselves after the first upward comparison, and ‘failed’ to do so in the second round. Upward comparison is often used for self-improvement [8, 13, 22], yet this was not possible in the experiment, as they were always shown an upward comparison.

From the qualitative results, it became evident that negative feelings do not always necessarily stop participants from viewing comparisons. Seventeen participants indicated a doubt to compare themselves to avoid confrontation with a bad score. Yet, they did not follow these doubts, and they proceeded to compare themselves all three times.

Though nearly all significant, the absolute differences in valence and arousal after downward comparison were minimal. A possible explanation is the low stakes of the experiment task. As noted in the pilot study, the emotional and motivational effects may be similar in a low-stakes context, yet stronger when the stakes are high.

5.3 Implications for learning analytics

This study found significant differences in affective state between upward and downward conditions. As emotions can affect motivation, performance, and learning strategies [30], these findings may help inform a social comparison design to foster motivation. However, feelings do not always align with behaviour. The qualitative data showed that participants doubted comparing themselves, yet they still chose to view the comparison. In addition, individual factors such as identification with the comparison target and expectations of performance can cause different emotional responses than those observed in this study [6, 13].

These individual factors make it difficult to generalize the results of this study to all learners. Furthermore, though two possible predictors were found for comparison choice, the models were not robust enough to draw strong conclusions. This lack of predictors shows that comparison choices cannot be simplified to a few variables. With these findings, this study underlines the work by Joshi et al. [25] and Akhuseyinoglu et al. [3], which highlighted the need to understand individual differences in social comparison preferences and behaviours.

5.4 Limitations & Future research

Despite its interesting insights, this study had several limitations. First, the task was not fully representative of an educational task. Education is difficult to mimic in short-term research due to the lack of control over participants’ pre-existing knowledge and intrinsic motivation for the topic. Where some studies

opt for hypothetical scenarios [1, 25], this study required participants to perform a low-stakes cognitive game. The downside was that the experiment was less focused on education. In addition, the Achievement Goal Questionnaire [15, 36] was used due to its prior usage in the context of social comparison [9, 25]. However, the questionnaire is context-specific and does not fully align with the task at hand. This was known beforehand, yet it was used, as potential useful patterns could be discovered. Future research in an educational setting is needed to draw stronger conclusions on social comparison choices in learners. A longitudinal study can embed the social comparison choices into a course. Similar studies have been performed [2, 3, 18, 28], though these focussed mainly on the effects of social comparison rather than why and when learners view such comparisons. In addition, longitudinal research can shed light on the changes in behaviour over time.

Second, quantitative data analysis provided several significant effects. However, results should be interpreted with caution. The majority of participants compared themselves all three times. The predictive models of comparison choice may not be reliable, as choosing not to view the comparison was underrepresented in the dataset. In addition, choices had to be made regarding the predictor variables for the models for both RQ1 and RQ2 to prevent overfitting. Interaction terms were not tested in this study. As a result, possible effects of interactions between achievement goals, affective state and condition may have gone unnoticed.

Third, only two out of 96 participants returned to the experiment, meaning no data analysis could be done on this part. Low return rates can be caused by various factors unrelated to the comparison, such as forgetfulness or lack of interest. The comparison itself could have contributed to the lack of returns, though future research is needed to determine possible correlations. Research has shown that preferences regarding social comparison can change over time [3, 22], making returning users an interesting subject of study, as their choices may also change. Furthermore, the results showed a possible effect over time in the upward comparison. Future research can investigate if these effects present themselves in the long-term. Future replications can be set up as repeated experiments over days or weeks. This can be designed as a diary study to gather more qualitative data on choices and how these change over time.

Fourth, in this study, both students and non-students were included due to limited time for data gathering. Analysis was performed on the full sample and the subset containing only students. No major differences were identified. However, this mixed participant pool should be considered when interpreting the results in educational contexts. Future replications of this study can investigate the differences in choice behaviour between different types of learners, such as university students, online students, and workplace learners, which were used in the study by [19]. These are all adult subjects. However, it could be interesting to investigate effects in primary school students. Research found that young children already compare themselves. However, as they get older, the motive to compare shifts from understanding social norms to evaluating one's perfor-

mance [13]. Future research could investigate the choices young learners make regarding social comparison, why they make these choices, and how this shifts as they grow older.

Fifth, The social comparison used a fixed margin for the comparison targets. This potentially caused the difference in average valence and arousal scores over the rounds, as participants could never improve themselves. Future research can investigate how allowing for choice of comparison targets affects the overall choice to compare. Including adaptive ranges, such as in the studies by Akhuseyinoglu et al. [2, 3] which can give insights into how autonomy over social comparison affects the choice to view the comparison.

Lastly, the social comparison had a minimalistic design. The metrics shown followed from the pilot study. However, the sample size of this pilot study was small. In addition, there are many different types of information and many ways to visualise this [23]. Future research could investigate the effect of different metrics and visualisations on the choice to view the comparison over time. For example, in the pilot study, participants indicated that time comparisons could be stressful. Future research can investigate if such a metric indeed induces stress, and how this affects the choice to compare oneself.

6 Conclusion

Recent research on social comparison in online learning settings has emphasised the need to understand and design for individual preferences of learners. However, little is known about the behaviour of learners when they get the choice to compare themselves. Therefore, this research aimed to answer the following research questions: *RQ1: How do comparison direction, affective state, and achievement motivation affect the choice to compare oneself?* and *RQ2: How do comparison direction and achievement motivation affect affective state after self-chosen comparison?*. To answer this question, an experimental study was performed in which participants could compare their scores on a low-stakes cognitive task.

The majority of participants compared themselves every possible time, mainly out of curiosity or to get a reference for their score. Little significant predictive effects of affective state, comparison direction, and achievement goals were found on the choice to compare were found. However, using generalized linear mixed-effects models, arousal was found to have a slight significant positive effect. In addition, mastery avoidance was significant for the student subset, though the effect was not significant when predictors were combined.

Comparison direction was found to significantly affect both valence and arousal. Those who saw an upward comparison felt less pleasant, whereas both arousal and valence increased for those who saw a downward comparison. Several participants doubted to compare themselves due to (expecting) bad scores, yet this eventually did not stop them from comparing themselves. However, no significant effects of achievement goals were found.

To conclude, though this study had several limitations, and future research is needed to verify effects in true educational settings, it provided an insight into an under-researched topic: the choice to compare oneself or not. The findings of this research provide opportunities for future work to better understand which choices learners make and why.

7 Declarations

Generative AI (specifically ChatGPT and GitHub CoPilot) was used during the creation of the webapp. Generative AI was only used for coding purposes. No participant data was provided to the AI tools, and no text was generated using AI.

References

1. Aghaei, K., Hatala, M., Mogharrab, A.: How Students' Emotion and Motivation Changes After Viewing Dashboards with Varied Social Comparison Group: A Qualitative Study. In: LAK23: 13th International Learning Analytics and Knowledge Conference. pp. 663–669. ACM, Arlington TX USA (Mar 2023). <https://doi.org/10.1145/3576050.3576107>, <https://dl.acm.org/doi/10.1145/3576050.3576107>
2. Akhuseyinoglu, K., McDonald, E., Klasnja Milicevic, A., Demmans Epp, C., Brusilovsky, P.: Exploring Adaptive Social Comparison for Online Practice. In: Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization. pp. 374–379. ACM, Cagliari Italy (Jun 2024). <https://doi.org/10.1145/3631700.3664899>, <https://dl.acm.org/doi/10.1145/3631700.3664899>
3. Akhuseyinoglu, K., Milicevic, A.K., Brusilovsky, P.: Who are My Peers? Learner-Controlled Social Comparison in a Programming Course. In: Hilliger, I., Muñoz-Merino, P.J., De Laet, T., Ortega-Arranz, A., Farrell, T. (eds.) *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption*. vol. 13450, pp. 1–15. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-16290-9_1, https://link.springer.com/10.1007/978-3-031-16290-9_1, series Title: Lecture Notes in Computer Science
4. Ames, C., Archer, J.: Mothers' beliefs about the role of ability and effort in school learning. *Journal of Educational Psychology* **79**(4), 409–414 (Dec 1987). <https://doi.org/10.1037/0022-0663.79.4.409>, <https://doi.apa.org/doi/10.1037/0022-0663.79.4.409>
5. Beheshitha, S.S., Hatala, M., Gašević, D., Joksimović, S.: The role of achievement goal orientations when studying effect of learning analytics visualisations. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. pp. 54–63. ACM Press, Edinburgh, United Kingdom (2016). <https://doi.org/10.1145/2883851.2883904>, <http://dl.acm.org/citation.cfm?doid=2883851.2883904>
6. Bennett, L., Folley, S.: Students' Emotional Reactions to Social Comparison via a Learner Dashboard. In: Sahin, M., Ifenthaler, D. (eds.) *visualisations and*

- Dashboards for Learning Analytics, pp. 233–249. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-81222-5_11, https://link.springer.com/10.1007/978-3-030-81222-5_11, series Title: Advances in Analytics for Learning and Teaching
7. Betella, A., Verschure, P.F.M.J.: The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLOS ONE* **11**(2), e0148037 (Feb 2016). <https://doi.org/10.1371/journal.pone.0148037>, <https://dx.plos.org/10.1371/journal.pone.0148037>
 8. Blanton, H., Buunk, B.P., Gibbons, F.X., Kuyper, H.: When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology* **76**(3), 420–430 (Mar 1999). <https://doi.org/10.1037/0022-3514.76.3.420>, <https://doi.apa.org/doi/10.1037/0022-3514.76.3.420>
 9. Bounoua, L., Cury, F., Regner, I., Huguet, P., Barron, K.E., Elliot, A.J.: Motivated use of information about others: Linking the 2×2 achievement goal model to social comparison propensities and processes. *British Journal of Social Psychology* **51**(4), 626–641 (Dec 2012). <https://doi.org/10.1111/j.2044-8309.2011.02027.x>, <https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/j.2044-8309.2011.02027.x>
 10. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* **25**(1), 49–59 (Mar 1994). [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9), <https://linkinghub.elsevier.com/retrieve/pii/0005791694900639>
 11. Braun, V., Clarke, V.: Using thematic analysis in psychology. *Qualitative Research in Psychology* **3**(2), 77–101 (Jan 2006). <https://doi.org/10.1191/1478088706qp063oa>, <http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa>
 12. Davis, D., Jivet, I., Kizilcec, R.F., Chen, G., Hauff, C., Houben, G.J.: Follow the successful crowd: raising MOOC completion rates through social comparison at scale. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. pp. 454–463. ACM, Vancouver British Columbia Canada (Mar 2017). <https://doi.org/10.1145/3027385.3027411>, <https://dl.acm.org/doi/10.1145/3027385.3027411>
 13. Dijkstra, P., Kuyper, H., Van Der Werf, G., Buunk, A.P., Van Der Zee, Y.G.: Social Comparison in the Classroom: A Review. *Review of Educational Research* **78**(4), 828–879 (Dec 2008). <https://doi.org/10.3102/0034654308321210>, <https://journals.sagepub.com/doi/10.3102/0034654308321210>
 14. Elliot, A.J.: Approach and avoidance motivation and achievement goals. *Educational Psychologist* **34**(3), 169–189 (Jun 1999). https://doi.org/10.1207/s15326985ep3403_3, http://www.tandfonline.com/doi/abs/10.1207/s15326985ep3403_3
 15. Elliot, A.J., McGregor, H.A.: A 2×2 achievement goal framework. *Journal of Personality and Social Psychology* **80**(3), 501–519 (2001). <https://doi.org/10.1037/0022-3514.80.3.501>, <https://doi.apa.org/doi/10.1037/0022-3514.80.3.501>
 16. Festinger, L.: A Theory of Social Comparison Processes. *Human Relations* **7**(2), 117–140 (May 1954). <https://doi.org/10.1177/001872675400700202>, <https://journals.sagepub.com/doi/10.1177/001872675400700202>
 17. Fleur, D.S., Van Den Bos, W., Bredeweg, B.: Learning Analytics Dashboard for Motivation and Performance. In: Kumar, V., Troussas, C. (eds.) *Intelligent Tutoring Systems*, vol. 12149, pp. 411–419. Springer International Publishing,

- Cham (2020). https://doi.org/10.1007/978-3-030-49663-0_51, http://link.springer.com/10.1007/978-3-030-49663-0_51, series Title: Lecture Notes in Computer Science
18. Fleur, D.S., Van Den Bos, W., Bredeweg, B.: Social comparison in learning analytics dashboard supporting motivation and academic achievement. *Computers and Education Open* **4**, 100130 (Dec 2023). <https://doi.org/10.1016/j.caeo.2023.100130>, <https://linkinghub.elsevier.com/retrieve/pii/S2666557323000083>
 19. Gallagher, T., Slof, B., Van Der Schaaf, M., Arztmann, M., Fracaro, S.G., Kester, L.: Learning analytics dashboard design: Workplace learner preferences for reference frames in immersive training in practice. *Journal of Computer Assisted Learning* **40**(6), 2840–2855 (Dec 2024). <https://doi.org/10.1111/jcal.13042>, <https://onlinelibrary.wiley.com/doi/10.1111/jcal.13042>
 20. Gerber, J.P., Wheeler, L., Suls, J.: A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin* **144**(2), 177–197 (Feb 2018). <https://doi.org/10.1037/bul0000127>, <https://doi.apa.org/doi/10.1037/bul0000127>
 21. Harvey, A.J., Keyes, H.: How do I compare thee? An evidence-based approach to the presentation of class comparison information to students using Dashboard. *Innovations in Education and Teaching International* **57**(2), 163–174 (Mar 2020). <https://doi.org/10.1080/14703297.2019.1593213>, <https://www.tandfonline.com/doi/full/10.1080/14703297.2019.1593213>
 22. Huguet, P., Dumas, F., Monteil, J.M., Genestoux, N.: Social comparison choices in the classroom: further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology* **31**(5), 557–578 (Sep 2001). <https://doi.org/10.1002/ejsp.81>, <https://onlinelibrary.wiley.com/doi/10.1002/ejsp.81>
 23. Jivet, I., Scheffel, M., Drachsler, H., Specht, M.: Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) *Data Driven Approaches in Digital Education*. pp. 82–96. Springer International Publishing, Cham (2017)
 24. Jivet, I., Scheffel, M., Specht, M., Drachsler, H.: License to evaluate: preparing learning analytics dashboards for educational practice. In: *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. pp. 31–40. ACM, Sydney New South Wales Australia (Mar 2018). <https://doi.org/10.1145/3170358.3170421>, <https://dl.acm.org/doi/10.1145/3170358.3170421>
 25. Joshi, A., Molenkamp, B., Sosnovsky, S.: Student Perception of Social Comparison in Technology Enhanced Learning. In: Viberg, O., Jivet, I., Muñoz-Merino, P., Perifanou, M., Papathoma, T. (eds.) *Responsive and Sustainable Educational Futures*. vol. 14200, pp. 118–132. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-42682-7_9, https://link.springer.com/10.1007/978-3-031-42682-7_9, series Title: Lecture Notes in Computer Science
 26. Koch, M., von Luck, K., Schwarzer, J., Draheim, S.: The novelty effect in large display deployments – experiences and lessons-learned for evaluating prototypes. In: *Proceedings of 16th European Conference on Computer-Supported Cooperative Work - Exploratory Papers*. European Society for Socially Embedded Technologies (EUSSET) (2018). https://doi.org/10.18420/ecscw2018_3
 27. Lim, L., Dawson, S., Joksimovic, S., Gašević, D.: Exploring students' sensemaking of learning analytics dashboards: Does frame of reference make a difference? In: *Proceedings of the 9th International Conference on Learning Analytics & Knowl-*

- edge. pp. 250–259. ACM, Tempe AZ USA (Mar 2019). <https://doi.org/10.1145/3303772.3303804>, <https://dl.acm.org/doi/10.1145/3303772.3303804>
28. Loboda, T.D., Guerra, J., Hosseini, R., Brusilovsky, P.: Mastery Grids: An Open Source Social Educational Progress visualisation. In: Rensing, C., De Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) *Open Learning and Teaching in Educational Communities*, vol. 8719, pp. 235–248. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-11200-8_18, http://link.springer.com/10.1007/978-3-319-11200-8_18, series Title: Lecture Notes in Computer Science
 29. Park, Y., Park, S.W.: Goal orientations and social comparison: The role of different motivations in affiliation preferences. *Motivation and Emotion* **41**(5), 617–627 (Oct 2017). <https://doi.org/10.1007/s11031-017-9634-6>, <http://link.springer.com/10.1007/s11031-017-9634-6>
 30. Pekrun, R., Goetz, T., Titz, W., Perry, R.P.: Academic Emotions in Students’ Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist* **37**(2), 91–105 (Jan 2002). https://doi.org/10.1207/S15326985EP3702_4, https://www.tandfonline.com/doi/full/10.1207/S15326985EP3702_4
 31. Russell, J.A.: A circumplex model of affect. *Journal of Personality and Social Psychology* **39**(6), 1161–1178 (Dec 1980). <https://doi.org/10.1037/h0077714>, <https://doi.apa.org/doi/10.1037/h0077714>
 32. Russell, J.A.: Core affect and the psychological construction of emotion. *Psychological Review* **110**(1), 145–172 (2003). <https://doi.org/10.1037/0033-295X.110.1.145>, <https://doi.apa.org/doi/10.1037/0033-295X.110.1.145>
 33. Siemens, G., Gasevic, D.: Guest Editorial - Learning and Knowledge Analytics. *Educational Technology and Society* **15**, 1–2 (Jul 2012)
 34. Sosnovsky, S., Fang, Q., De Vries, B., Luehof, S., Wiegant, F.: Towards Adaptive Social Comparison for Education. In: Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M. (eds.) *Addressing Global Challenges and Quality Education*. vol. 12315, pp. 421–426. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-57717-9_38, http://link.springer.com/10.1007/978-3-030-57717-9_38, series Title: Lecture Notes in Computer Science
 35. Suárez Vázquez, A., Suárez Álvarez, L., Del Río Lanza, A.B.: Is comparison the thief of joy? Students’ emotions after socially comparing their task grades, influence on their motivation. *The International Journal of Management Education* **21**(2), 100813 (Jul 2023). <https://doi.org/10.1016/j.ijme.2023.100813>, <https://linkinghub.elsevier.com/retrieve/pii/S1472811723000514>
 36. Van Yperen, N.: Constructie en validatie van de Nederlandstalige Doeloriëntatie Vragenlijst (NDV). *Gedrag & Organisatie* **17**, 387–401 (2004), publisher: UIT-GEVERIJ LEMMA B V
 37. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* **54**(6), 1063–1070 (1988). <https://doi.org/10.1037/0022-3514.54.6.1063>, <https://doi.apa.org/doi/10.1037/0022-3514.54.6.1063>
 38. Wehrens, M.J., Buunk, A.P., Lubbers, M.J., Dijkstra, P., Kuyper, H., Van Der Werf, G.P.: The relationship between affective response to social comparison and academic performance in high school. *Contemporary Educational Psychology* **35**(3), 203–214 (Jul 2010). <https://doi.org/10.1016/j.cedpsych.2010.01.001>, <https://linkinghub.elsevier.com/retrieve/pii/S0361476X10000044>

39. Wills, T.A.: Downward comparison principles in social psychology. *Psychological Bulletin* **90**(2), 245–271 (Sep 1981). <https://doi.org/10.1037/0033-2909.90.2.245>, <https://doi.apa.org/doi/10.1037/0033-2909.90.2.245>
40. Zevon, M.A., Tellegen, A.: The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology* **43**(1), 111–122 (Jul 1982). <https://doi.org/10.1037/0022-3514.43.1.111>, <https://doi.apa.org/doi/10.1037/0022-3514.43.1.111>

A Ethics and privacy Quick Scan

The Ethics and Privacy Quick Scan of the Utrecht University research Institute of Information and Computing Sciences was filled out before conducting the pilot study and experiment. The Quick Scan identified warnings with regards to the experimental study. However, the project was allowed to proceed with permission from the moderator (Figure 19). As discussed, participants were provided a link to the results page, on which they were informed about the underlying purpose of the research Figure 27. As not personal data was stored, the use of a third party (Strato) was allowed Figure 18.



Fig. 18. Contact with moderator regarding Quicksan warnings - 1

Groet,
Fleur

From: OZP Informatiekunde [redacted]
Sent: donderdag 1 mei 2025 08:52
To: Leenders, F. (Fleur) [redacted]
Subject: RE: Ethics and Privacy Quick Scan: Issues with thesis project

Hoi Fleur,

Ook nog even voor jouw administratie via de mail: Zoals besproken, als je de deelnemers achteraf vertelt dat de vergelijking niet op echte data gebaseerd was en wat het doel van het onderzoek was, dan zie ik verder geen probleem.

Succes met het onderzoek!

Met vriendelijke groet,
Imke

From: Qualtrics Survey Software [redacted]
Sent: vrijdag 11 april 2025 16:01
To: OZP Informatiekunde [redacted]
Subject: Ethics and Privacy Quick Scan: Issues with thesis project

Dear moderator,

Fleur Leenders (email: [redacted] supervisor email: [redacted]) completed the Quick Scan with ethics and/or privacy issues. Please consider the student's elaboration provided on the issues, and determine whether (1) the research can go ahead, (2) changes to the research design are needed, or (3) a fuller privacy assessment by the data steward / full ethical review is required.

For help with ethics issues, you can consult our members of the ethical review committee. For help with privacy issues, you can consult the faculty data steward [redacted]. For any issues with the Quick Scan, please contact the research director [redacted]. Please let the student (cc their supervisor and [redacted]) know the outcome as soon as possible, preferably within one working week for bachelor theses and 10 working days for master theses.

Please find the report below. Thank you for your help.

[Download as PDF](#)

URL to view Results [\[Click Here\]](#)

Response Summary:

Section 1. Research projects involving human participants

2

Fig. 19. Contact with moderator regarding Quickscan warnings - 2

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from X, Reddit) without directly recruiting participants, please answer no.

• Yes

Recruitment

P2. Does your project involve participants younger than 16 years of age?

• No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

• No

P4. Is your project likely to involve participants engaging in illegal activities?

• No

P5. Does your project involve patients?

• No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

• No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

• Yes

P9. Is it made clear to potential participants that not participating will in no way impact them (e.g. it will not directly impact their grade in a class)?

• Yes

Fig. 20. Contact with moderator regarding Quicksan warnings - 3

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent prior to collecting data from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on [the website](#).)

• Yes

PC2. Will you tell participants that their participation is voluntary?

• Yes

PC3. Will you obtain explicit consent for participation?

• Yes

PC4. Will you obtain explicit consent for any sensor readings, eye tracking, photos, audio, and/or video recordings?

• Yes

PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?

• Yes

PC6. Will you give potential participants time to consider participation?

• Yes

PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?

• Yes

PC8. Does your project involve concealment or deliberate misleading of participants?

• Yes

Fig. 21. Contact with moderator regarding Quicksan warnings - 4

Ethics Warning.  As you plan to use concealment or misleading (yes to PC8), and this may impact participants' rights to informed consent, a fuller ethical review is required. Please provide more information on the concealment/misleading here:

Participants perform a small task (a word puzzle), after which their score is shown in comparison to others. This happens several times in a row. The data of the others in the ranking is fake and is manipulated so the participant is placed on a specific point in the ranking. This is done to simulate different types of social comparison (upwards and downwards) for different study conditions.

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of **personal data** (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?

• Yes

High-risk data

DR1. Will you process personal data that could jeopardize the physical health, (physical, social, etc) safety, or legal rights of individuals in the event of a personal data breach?

• No

DR2. Will you combine, compare, or match personal data obtained from multiple sources, in a way that exceeds the reasonable expectations of the people whose data it is?

• No

DR3. Will you use any personal data of children or vulnerable individuals for marketing, profiling, automated decision-making, or to offer online services to them?

• No

DR4. Will you profile individuals on a large scale?

• No

Fig. 22. Contact with moderator regarding Quicksan warnings - 5

DR5. Will you systematically monitor individuals in a publicly accessible area on a large scale (or use the data of such monitoring)?

• No

DR6. Will you use special category personal data, criminal offense personal data, or other sensitive personal data on a large scale?

• No

DR7. Will you determine an individual's access to a product, service, opportunity, or benefit based on an automated decision or special category personal data?

• No

DR8. Will you systematically and extensively monitor or profile individuals, with significant effects on them?

• No

DR9. Will you use innovative technology to process sensitive personal data?

• No

Data minimization

DM1. Will you collect only personal data that is strictly necessary for the research?

• No

DM2. Will you only collect not strictly necessary personal data because it is (1) technically unfeasible not to collect it when collecting necessary data, or (2) needed as a source of necessary data?

• Yes

DM3. Will you (1) extract any necessary data as soon as possible from the collected not strictly necessary data and (2) delete the not strictly necessary data immediately after any required extraction?

• Yes

Fig. 23. Contact with moderator regarding Quicksan warnings - 6

DM4. Will you anonymize the data wherever possible?

• Yes

DM5. Will you pseudonymize the data if you are not able to anonymize it, replacing personal details with an identifier, and keeping the key separate from the data set?

• Yes

Using collaborators or contractors that process personal data securely

DC1. Will any organization external to Utrecht University be involved in processing personal data (e.g. for transcription, data analysis, data storage)?

• No

International personal data transfers

DI1. Will any personal data be transferred to another country (including to research collaborators in a joint project)?

• No

Fair use of personal data to recruit participants

DF1. Is personal data used to recruit participants?

• No

Participants' data rights and privacy information

DP1. Will participants be provided with privacy information? (Recommended is to use as part of the information sheet: For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information at www.uu.nl/en/organisation/privacy.)

• Yes

7

Fig. 24. Contact with moderator regarding Quicksan warnings - 7

DP2. Will participants be aware of what their data is used for?

- Yes

DP3. Can participants request that their personal data be deleted?

- Yes

DP4. Can participants request that their personal data be rectified (in case it is incorrect)?

- Yes

DP5. Can participants request access to their personal data?

- Yes

DP6. Can participants request that personal data processing is restricted?

- Yes

DP7. Will participants be subjected to automated decision-making based on their personal data with an impact on them beyond the research study to which they consented?

- No

DP8. Will participants be aware of how long their data is being kept for, who it is being shared with, and any safeguards that apply in case of international sharing?

- Yes

DP9. If data is provided by a third party, are people whose data is in the data set provided with (1) the privacy information and (2) what categories of data you will use?

- Not applicable

Using data that you have not gathered directly from participants

DE1. Will you use any personal data that you have not gathered directly from participants (such as data from an existing data set, data gathered for you by a third party, data scraped from the internet)?

Fig. 25. Contact with moderator regarding Quicksan warnings - 8

•No

Secure data storage

DS1. Will any data be stored (temporarily or permanently) anywhere other than on password-protected University authorized computers or servers?

•Yes

DS2. Does this only involve data stored temporarily during a session with participants (e.g. data stored on a video/audio recorder/sensing device), which is immediately transferred (directly or with the use of an encrypted and password-protected data-carrier (such as a USB stick)) to a password-protected University authorized computer or server, and deleted from the data capture and data-carrier device immediately after transfer?

•Yes

DS4. Excluding (1) any international data transfers mentioned above and (2) any sharing of data with collaborators and contractors, will any personal data be stored, collected, or accessed from outside the EU?

•No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

H1. Does your project give rise to a realistic risk to the national security of any country?

•No

H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?

•No

H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)

•No

Fig. 26. Contact with moderator regarding Quicksan warnings - 9

H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)

• No

H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?

• No

H6. Does your project give rise to a realistic risk of harm to the researchers?

• No

H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?

• Yes

H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?

• No

H9. Is there a realistic risk of other types of negative externalities?

• No

Ethics Warning.  As you replied yes to one (or more) of H1-H9, a fuller ethical review is required. Please provide more detail here on the potential harm, and how you will minimize risk and impact:

H7: As the experiment tests the effect of social comparison on emotions, possible negative effects on participants' emotions are expected. Participants could experience psychological discomfort from seeing they scored low in the social comparison ranking. No psychological harm is expected. It is expected that this discomfort is minimal, as the stakes of the test are low for participants, as their task is a word puzzle.

To minimize the risk further, participants receive a link to a website where an update of the research is posted after it is done. On this website, the misleading and the reason behind it is explained. It is a deliberate choice to not immediately tell them this, as participants get the chance to return to the experiment and perform the task again. If they know the purpose, this would affect the data of returning participants.

Fig. 27. Contact with moderator regarding Quicksan warnings - 10

Section 4. Conflicts of interest

C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?

- No

C2. Is there a direct hierarchical relationship between researchers and participants?

- No

Section 5. Your information.

This last section collects data about you and your project so that we can register that you completed the Ethics and Privacy Quick Scan, sent you (and your supervisor/course coordinator) a summary of what you filled out, and follow up where a fuller ethics review and/or privacy assessment is needed. For details of our legal basis for using personal data and the rights you have over your data please see the [University's privacy information](#). Please see the guidance on the [ICS Ethics and Privacy website](#) on what happens on submission.

Z0. Which is your main department?

- Information and Computing Science

Z1. Your full name:

Fleur Leenders

Z2. Your email address:

f.leenders@uu.nl

Z3. In what context will you conduct this research?

- As a student for my bachelor thesis, supervised by::

Jeroen Ooge

Z4. Bachelor programme for which you are doing the thesis

- Information Science (Informatiekunde)

Z6. Email of the course coordinator or supervisor (so that we can inform them that you filled this out and provide them with a summary):

Fig. 28. Contact with moderator regarding Quicksan warnings - 11

Z7. Email of the moderator (if you are studying a programme from Information and Computing Sciences, please leave the value provided, otherwise change this to the moderator as provided by the coordinator of your thesis project):

Z8. Title of the research project/study for which you filled out this Quick Scan:
(Working title) The relation between achievement goals, emotions and social comparison .

Z9. Summary of what you intend to investigate and how you will investigate this (200 words max):

The aim of the study is to investigate the effect of social comparison orientation and motivational goals on emotions and choice behaviour regarding the social comparison, as well as the effect emotions and choice behaviour on each other. The study consists of two parts: preliminary interviews to help build the research setup. 5-10 participants are asked a few questions on their preferences regarding social comparison in the context of a Spelling Bee inspired game. They are asked about their preferences regarding the visualisation and frequency of the comparison. This will help to expand the setup of the actual experiment.

In the actual experiment, participants are asked to fill in an online survey and play a Spelling Bee inspired game. This will be hosted on the UU webspace server. Data is stored on the server too. Participants first get a pre-questionnaire on their motivation and current emotional state. After that, they go through a sequence of playing the Spelling Bee, testing emotion changes, and viewing a social comparison of their score. Participants are split into two groups: upwards and downwards comparison. In every round, they have a choice to view the comparison. They receive a post-questionnaire with open questions for qualitative data.

Z10. In case you encountered warnings in the survey, does supervisor already have ethical approval for a research line that fully covers your project?

• No

Scoring

- Privacy: 0
 - Ethics: 2
-

Fig. 29. Contact with moderator regarding Quicksan warnings - 12

B Pilot study materials

B.1 Informed consent

Information letter

Welkom bij dit onderzoek. Dit onderzoek wordt uitgevoerd als onderdeel van het afrondende onderzoeksproject voor de bachelor informatiekunde aan de Universiteit Utrecht. Dit onderzoek wordt uitgevoerd door Fleur Leenders, onder begeleiding van Jeroen Ooge.

Dit onderzoek dient als vooronderzoek voor een experiment naar de effecten van persoonlijk motivatie en sociale vergelijkingen op emoties. Het betreft een kort interview over jouw voorkeuren voor sociale vergelijking in de context van een woordspel (Spelling Bee). Je krijgt eerst een uitleg over het spel en de context. Daarna zal de onderzoeker je een aantal vragen stellen over jouw voorkeuren. Het interview duurt ongeveer 10-15 minuten.

De interviews worden opgenomen met een mobiele telefoon. De gesprekken worden getranscribeerd en de opnames worden daarna direct verwijderd. Om je privacy te waarborgen wordt de getranscribeerde data anoniem opgeslagen en verwerkt. Hiervoor worden een vergrendelde persoonlijke computer en de Microsoft OneDrive van de Universiteit Utrecht gebruikt.

Aan dit onderzoek zitten geen verdere risico's verbonden.

Deelname aan dit onderzoek is geheel vrijwillig. Je kunt op elk moment je toestemming intrekken door dit aan te geven aan de onderzoeker. Ook heb je het recht om je toestemming achteraf in te trekken. Hiervoor kan je contact opnemen met de onderzoeker.

Vragen kan je direct stellen aan de onderzoeker, of je kan contact opnemen via f.leenders@uu.nl.

Consent form

Ik heb de informatiebrief over deelname aan het onderzoek gelezen. Ik kon aanvullende vragen stellen. Mijn vragen zijn genoeg beantwoord. Ik had genoeg tijd om te beslissen of ik meedoe.

*Ja/Nee**

Ik weet dat meedoen helemaal vrijwillig is. Ik weet dat ik op ieder moment kan beslissen om toch niet mee te doen. Daarvoor hoef ik geen reden te geven.

*Ja/Nee**

Ik geef toestemming om mijn gegevens te gebruiken, voor de doelen die in de informatiebrief staan.

*Ja/Nee**

Ik geef toestemming om de verzamelde gegevens minimaal 10 jaar na afloop van dit onderzoek te bewaren.

*Ja/Nee**

Ik bevestig dat ik 16 jaar of ouder ben.

*Ja/Nee**

Ik wil meedoen aan dit onderzoek.

*Ja/Nee**

** Doorhalen wat niet van toepassing is.*

Naam deelnemer:

Handtekening:

Datum:

B.2 Questionnaire

Table 14. Pilot study questionnaire

Question English	Question Dutch
Would you want to see information about your own performance? If yes, what kind?	Zou je informatie over je performance willen zien? Zo ja, welke informatie?
Would you only want to see this for yourself, or also in comparison with others?	Zou je dit alleen voor jezelf willen zien of in vergelijking met anderen?
When would you want to see it?	Wanneer zou je het willen zien?
How often would you want to see it?	Hoe vaak zou je het willen zien?
Would you always want to see it?	Zou je het altijd willen zien?
Do you think it would affect your mood?	Denk je dat het je gemoedstoestand zou beïnvloeden?
Do you think it would affect your motivation in the long term?	Denk je dat het op lange termijn impact zou hebben op je motivatie?
Do you think you would answer these questions differently in a different context?	Denk je dat je andere antwoorden zou geven in een andere context?

C Experimental study materials

C.1 Informed consent

Welkom bij dit onderzoek over de effecten van sociale vergelijking op emoties. Dit onderzoek wordt uitgevoerd als onderdeel van het afrondende onderzoeksproject voor de bachelor informatiekunde aan de Universiteit Utrecht. Dit onderzoek wordt uitgevoerd door Fleur Leenders (f.leenders@uu.nl), onder begeleiding van Jeroen Ooge (j.ooge@uu.nl).

Je wordt eerst een aantal meerkeuzevragen gesteld. Daarna volgt een woordpuzzel gebaseerd op de Spelling Bee van The New York Times. Tussen het spel door worden vragen gesteld. Daarna volgt een korte vragenlijst van twee open vragen. Het onderzoek duurt maximaal 15 minuten.

Data wordt verzameld via en opgeslagen op een virtuele privéserver van Strato. Voor het gebruik van deze derde partij is expliciet toestemming gevraagd aan de universiteit. De data blijft maximaal een maand op de server staan. Daarna wordt het van de servers gehaald en lokaal op de laptop van de onderzoeker opgeslagen. Om het experiment goed te laten verlopen worden tijdelijk cookies in de browser geplaatst. Deze bevatten geen persoonlijke gegevens en worden na het sluiten van het tabblad automatisch verwijderd.

Alleen de onderzoeker heeft toegang tot de server. Deze is zo goed mogelijk beschermd, maar een datalek van de server is niet volledig uit te sluiten. Vul daarom geen persoonlijke gegevens in in de open vragen.

Deelname aan dit onderzoek is geheel vrijwillig. Je kunt op elk moment je toestemming intrekken door de browser te sluiten. Je hebt het recht om data te laten corrigeren of verwijderen. Let op: zodra de data geanonimiseerd is, kan deze niet meer gewijzigd of verwijderd worden.

Voor vragen of opmerkingen kan je contact opnemen via f.leenders@uu.nl.

Door op onderstaande knop te klikken, bevestig je het volgende:

- Ik weet dat meedoen helemaal vrijwillig is. Ik weet dat ik op ieder moment kan beslissen om toch niet mee te doen. Daarvoor hoef ik geen reden te geven.
- Ik geef toestemming om mijn gegevens te gebruiken, voor de doelen zoals bovenstaand vermeld.
- Ik geef toestemming om de verzamelde gegevens minimaal 10 jaar na afloop van dit onderzoek te bewaren.
- Ik bevestig dat ik 16 jaar of ouder ben.

Ik heb bovenstaande gelezen en wil meedoen aan het onderzoek.

C.2 Initial questionnaire

Table C.2 shows the questions asked regarding demographics. Table 16 shows the questionnaire used for scoring the achievement goals.

Table 15. Demographics questions

Question English	Question Dutch	Answer options
What is your gender identity?	Wat is je gender-identiteit?	Vrouw, Man, Anders
How old are you?	Hoe oud ben je?	Numeric input
Are you a student?	Ben je student?	Ja, Nee

Figure 30 shows the interface for the Affective Slider, adapted from the original source code [7]. In the initial questionnaire, extra context was provided. During the rest of the experiment, only the sliders and the one line above were shown.

Gevoel

Geef hieronder aan hoe je je voelt. Tijdens het experiment zal je verschillende keren moeten aangeven hoe je je voelt. Let op: de sliders verschijnen daarbij in willekeurige volgorde.

Geef aan hoe **plezierig** je je op dit moment voelt.

Geef aan hoe **opgewonden** je je op dit moment voelt.

Volgende

Fig. 30. Affective Slider interface

Table 16: Achievement goal questionnaire

Subscale	Original by [15]	AGQ Dutch translation by [36]
Performance approach	It is important for me to do better than other students.	Het is belangrijk voor mij dat ik beter presteer dan studenten in mijn directe omgeving.
	It is important for me to do well compared to others in this class.	Het is belangrijk voor mij dat ik in vergelijking met anderen in mijn studie goed presteer.
	My goal in this class is to get a better grade than most of the other students.	In mijn studie ben ik erop gericht om betere beoordeling te krijgen dan mijn medestudenten.
Mastery approach	I want to learn as much as possible from this class.	Ik wil zo veel mogelijk leren in mijn studie.
	It is important for me to understand the content of this course as thoroughly as possible.	Ik heb er behoefte aan om alles wat mijn studie aangaat zo goed mogelijk te beheersen.
	I desire to completely master the material presented in this class.	Het is belangrijk voor mij om de inhoud van mijn studie zo grondig mogelijk te begrijpen.
Performance avoidance	I just want to avoid doing poorly in this class.	Ik wil gewoon voorkomen dat ik het slecht doe in mijn studie.
	My goal in this class is to avoid performing poorly.	Mijn doel in mijn studie is te voorkomen dat ik slecht presteer.
	My fear of performing poorly in this class is often what motivates me.	Mijn angst om slecht te presteren in mijn studie is vaak wat mij motiveert.

Mastery avoidance	I worry that I may not learn all that I possibly could in this class.	Ik maak mij zorgen dat ik inhoudelijk niet alles uit mijn studie haal wat mogelijk is.
	Sometimes I'm afraid that I may not understand the content of this class as thoroughly as I'd like.	Soms ben ik bang dat ik de inhoud van mijn studie niet zo grondig begrijp als ik wel zou willen.
	I am often concerned that I may not learn all there is to learn in this class.	Ik ben vaak bezorgd dat ik niet alles wat er te leren valt in mijn studie er ook werkelijk uithaal.

C.3 Final questionnaire

Table C.3 shows the open questions asked in the final questionnaire.

Table 17. Final questionnaire

Question English	Question Dutch
Why did you do want to compare yourself?	Waarom wilde je jezelf wel vergelijken?
Why did you not want to compare yourself?	Waarom wilde je jezelf niet vergelijken?

D Additional results

D.1 Descriptives

Table 18. Number of comparisons per condition

Condition	# Viewed comparisons				Total
	0	1	2	3	
Downward	3	3	10	35	51
Upward	1	3	8	33	45
Total	4	6	18	68	96

D.2 Effect on choice

Achievement goals data

Table 19 shows the mean scores and internal consistency on achievement goals within the subset used for the effect on comparison choice Figure 31 shows the distribution of goal scores for both conditions. Table 20 and Figure 32 show the same, but for the student subset.

Table 19. Distribution of scores on the achievement goal subscales including non-students

Scale	Mean	SD	Min	Max	α
Performance approach	3.15	0.83	1	5	0.809
Mastery approach	3.81	0.79	1.33	5	0.835
Performance avoidance	3.39	0.97	1.33	5	0.694
Mastery avoidance	3.31	0.96	1.33	5	0.84

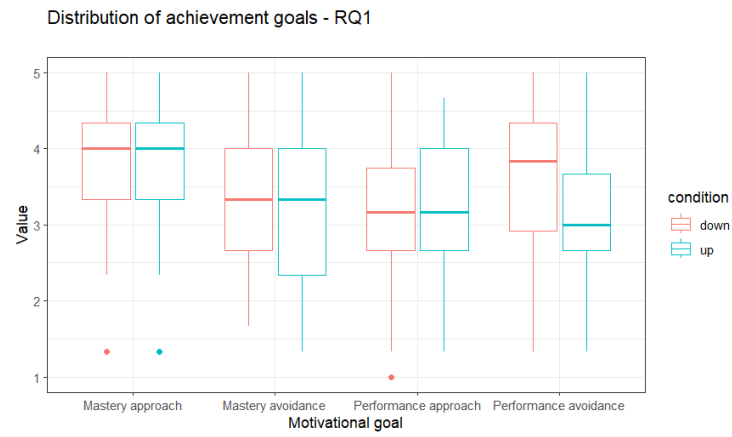


Fig. 31. Distribution of achievement goals including non-students

Table 20. Distribution of scores on the achievement goal subscales excluding non-students

Scale	Mean	SD	Min	Max	α
Performance approach	3.11	0.86	1	5	0.817
Mastery approach	3.77	0.84	1.33	5	0.854
Performance avoidance	3.38	0.99	1.33	5	0.707
Mastery avoidance	3.38	0.94	1.33	5	0.813

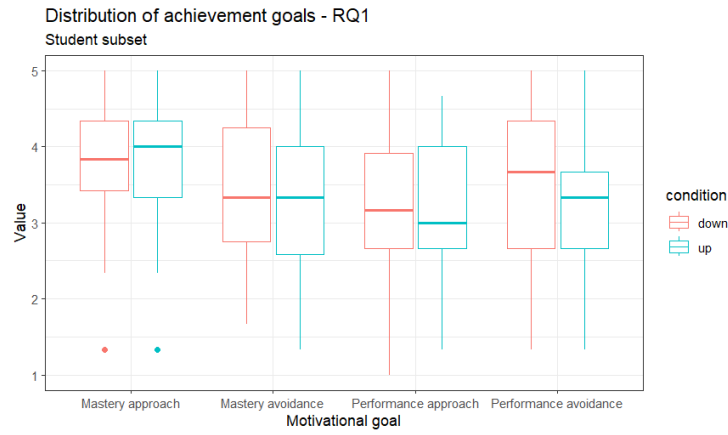


Fig. 32. Distribution of achievement excluding non-students

Generalized Linear Mixed-Effects Models

Table 21 shows the coefficients of the models ran using just one variable. Table ?? shows the combined model. Table 23 and Table 24 shows the same information for the student subset.

Table 21. Coefficients for choice predictors including non-students

model	term	β	std.error	statistic	p	e^{β}
m1	Performance approach	-0.003	0.80	-0.004	1.00	1.00
m2	Performance avoidance	-0.07	0.67	-0.11	0.92	0.93
m3	Mastery approach	0.22	0.81	0.27	0.79	1.24
m4	Mastery avoidance	0.86	0.71	1.21	0.22	2.37
m5	conditionup	0.14	1.29	0.11	0.91	1.15
m6	Post arousal (last round)	0.02	0.03	0.64	0.52	1.02
m7	Post valence (last round)	-0.03	0.03	-0.97	0.33	0.97
m8	Pre arousal	0.04	0.02	2.48	0.01	1.04
m9	Pre valence	0.05	0.03	1.59	0.11	1.05

Table 22. Mixed model with multiple choice predictors including non-students

term	β	std.error	statistic	p
(Intercept)	2.73	0.85	3.21	< 0.01
Performance approach*	-0.02	0.33	-0.05	0.96
Performance avoidance*	-0.20	0.35	-0.58	0.56
Mastery approach*	0.11	0.35	0.32	0.75
Mastery avoidance*	0.69	0.37	1.86	0.06
Condition (up)	0.45	0.68	0.66	0.51
Pre arousal*	0.62	0.34	1.81	0.07
Pre valence*	0.25	0.32	0.78	0.44
SD (Intercept)	1.25			

Note: starred terms are standardized.

Table 23. Coefficients for choice predictors excluding non-students

model	term	β	std.error	statistic	p	e^{β}
m1	Performance approach	0.31	0.88	0.35	0.72	1.36
m2	Performance avoidance	0.19	0.74	0.26	0.80	1.21
m3	Mastery approach	0.43	0.86	0.50	0.62	1.54
m4	Mastery avoidance	1.16	0.55	2.12	0.03	3.19
m5	conditionup	0.06	1.46	0.04	0.96	1.07
m6	Post arousal (last round)	0.02	0.04	0.62	0.54	1.02
m7	Post valence (last round)	-0.05	0.04	-1.19	0.23	0.95
m8	Pre arousal	0.04	0.02	2.26	0.02	1.04
m9	Pre valence	0.04	0.04	1.18	0.24	1.04

Table 24. Mixed model with multiple choice predictors excluding non-students

term	β	std.error	statistic	p
(Intercept)	2.38	0.94	2.54	0.01
Performance approach*	0.09	0.33	0.28	0.78
Performance avoidance*	0.11	0.32	0.33	0.74
Mastery approach*	0.19	0.34	0.57	0.57
Mastery avoidance*	0.77	0.43	1.78	0.07
Condition (up)	0.62	0.70	0.89	0.37
Pre arousal*	0.59	0.34	1.76	0.08
Pre valence*	0.20	0.31	0.64	0.52
SD (Intercept)	0.55			

Note: starred terms are standardized.

D.3 Effect of comparison

Achievement goals data

Table D.3 shows the mean scores and internal consistency on achievement goals within the subset used for analysis on the effect of comparison. Figure 33 shows the distribution of goal scores for both conditions. Table D.3 and Figure 34 show the data for the student subset.

Table 25. Motivation subscales

Scale	Mean	SD	Min	Max	α
Performance approach	3.16	0.82	1	5	0.804
Mastery approach	3.82	0.8	1.33	5	0.833
Performance avoidance	3.37	0.97	1.33	5	0.686
Mastery avoidance	3.3	0.98	1.33	5	0.848

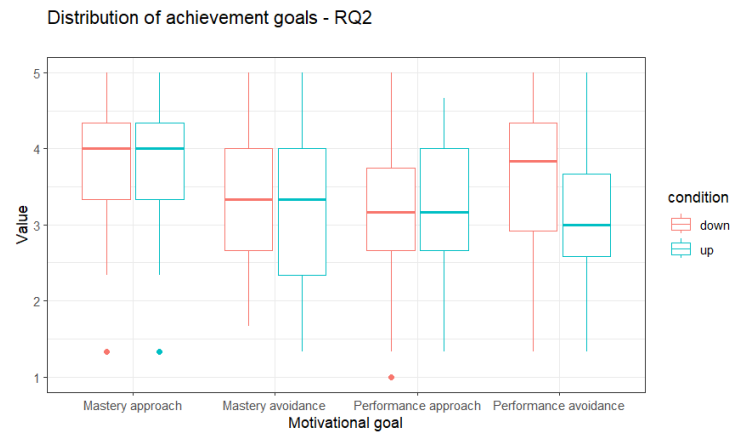


Fig. 33. Distribution of achievement goals

Table 26. Motivation subscales - student subset

Scale	Mean	SD	Min	Max	α
Performance approach	3.12	0.85	1	5	0.811
Mastery approach	3.78	0.84	1.33	5	0.851
Performance avoidance	3.36	0.99	1.33	5	0.695
Mastery avoidance	3.37	0.97	1.33	5	0.825

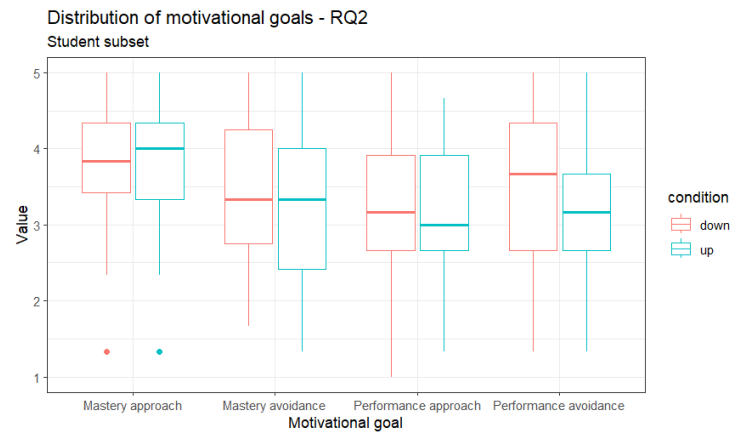


Fig. 34. Distribution of achievement goals

Linear Mixed-Effects Models - Valence

Table D.3 shows the coefficients for the models using one variable. Table D.3 shows the coefficients of the combined model. Table D.3 and Table D.3 show the same data for the analysis on the student subset.

Table 27. Coefficients valence predictor - all

model	term	estimate	std.error	statistic	p.value
m1	Performance approach	1.05	1.94	0.54	0.59
m2	Performance avoidance	2.89	1.60	1.81	0.07
m3	Mastery approach	0.43	1.98	0.22	0.83
m4	Mastery avoidance	2.41	1.62	1.49	0.14
m5	Condition (up)	-18.62	2.46	-7.56	<0.01

Table 28. Coefficients full mixed model valence

term	β	std.error	statistic	p
(Intercept)	69.22	1.74	39.73	<0.01
Initial valence*	5.25	1.35	3.90	<0.01
Pre valence*	8.30	0.96	8.69	<0.01
Performance approach*	1.13	1.30	0.87	0.39
Performance avoidance*	0.29	1.29	0.23	0.82
Mastery approach*	0.26	1.29	0.20	0.84
Mastery avoidance*	1.31	1.26	1.04	0.30
Condition (up)	-18.34	2.58	-7.10	<0.01
sd__(Intercept)	10.04			
sd__Observation	9.89			

Note: starred terms are standardized.

Table 29. Coefficients valence predictor - student subset

model	term	β	std.error	statistic	p
m1	Performance approach	2.19	2.26	0.97	0.34
m2	Performance avoidance	2.40	1.89	1.27	0.21
m3	Mastery approach	1.93	2.25	0.86	0.39
m4	Mastery avoidance	2.21	1.96	1.13	0.26
m5	Condition (up)	-19.73	2.92	-6.75	<0.01

Table 30. Coefficients full mixed model valence

term	β	std.error	statistic	p
(Intercept)	69.62	2.04	34.21	<0.01
Initial valence*	6.75	1.59	4.25	<0.01
Pre valence*	8.00	1.04	7.66	<0.01
Performance approach*	1.57	1.55	1.01	0.32
Performance avoidance*	0.53	1.51	0.35	0.72
Mastery approach*	0.96	1.52	0.63	0.53
Mastery avoidance*	0.89	1.46	0.61	0.55
Condition (up)	-19.43	3.01	-6.45	<0.01
SD (Intercept)	11.10			
SD Observation	8.91			

Note: starred terms are standardized.

Linear Mixed-Effects Models - Arousal

Table D.3 shows the coefficients for the models using one variable. Table D.3 shows the coefficients of the combined model. Table D.3 and Table D.3 show the same data for the analysis on the student subset.

Table 31. Coefficients arousal predictor - all

model	term	β	std.error	statistic	p
m1	Performance approach	0.52	1.75	0.30	0.77
m2	Performance avoidance	0.61	1.45	0.42	0.68
m3	Mastery approach	-0.35	1.78	-0.20	0.84
m4	Mastery avoidance	1.32	1.45	0.91	0.37
m5	Condition (up)	-10.35	2.59	-4.00	<0.01

Table 32. Coefficients full mixed model valence

term	β	std.error	statistic	p
(Intercept)	58.29	1.84	31.65	<0.01
Initial arousal*	3.61	1.38	2.61	0.01
Pre arousal*	8.22	0.94	8.70	<0.01
Performance approach*	0.65	1.38	0.47	0.64
Performance avoidance*	-0.95	1.37	-0.69	0.49
Mastery approach*	-0.43	1.37	-0.31	0.76
Mastery avoidance*	0.79	1.32	0.60	0.55
Condition (up)	-10.66	2.73	-3.90	<0.01
SD (Intercept)	10.97			
SD Observation	9.51			

Note: starred terms are standardized.

Table 33. Coefficients arousal predictor - student subset

model	term	β	std.error	statistic	p
m1	Performance approach	1.14	2.00	0.57	0.57
m2	Performance avoidance	0.16	1.69	0.10	0.92
m3	Mastery approach	0.28	2.00	0.14	0.89
m4	Mastery avoidance	1.68	1.75	0.96	0.34
m5	Condition (up)	-10.24	3.12	-3.28	<0.01

Table 34. Coefficients full mixed model arousal - student subset

term	β	std.error	statistic	p
(Intercept)	58.03	2.19	26.45	< 0.01
Initial arousal*	3.82	1.69	2.26	0.03
Pre arousal*	8.23	1.21	6.78	< 0.01
Performance approach*	0.93	1.66	0.56	0.58
Performance avoidance*	-0.94	1.63	-0.57	0.57
Mastery approach*	-0.22	1.64	-0.14	0.89
Mastery avoidance*	1.11	1.58	0.70	0.48
Condition (up)	-10.36	3.26	-3.18	< 0.01
SD (Intercept)	11.77			
SD Observation	10.13			

Note: starred terms are standardized.

Analysis of conditions

Below are shown the analyses of pre- and post-comparison scores over the rounds for the student subset.

Table 35. Results of Wilcoxon signed-rank tests - Round 1.

Name	Statistic	p	Mean pre	Mean post	Mean difference
Upward arousal	212	0.91	48.1	47.7	-0.4
Downward arousal	164.5	0.10	50.5	54.9	4.4
Upward valence	350	< 0.01	59.5	50.5	-8.9
Downward valence	96.5	< 0.01	61.3	70.7	9.4

Table 36. Results of Wilcoxon signed-rank tests - Round 2.

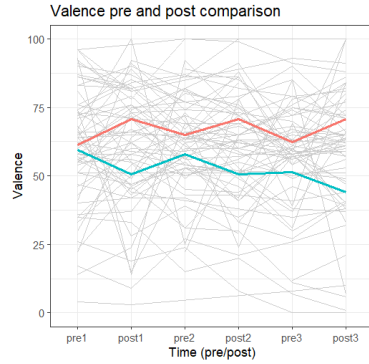
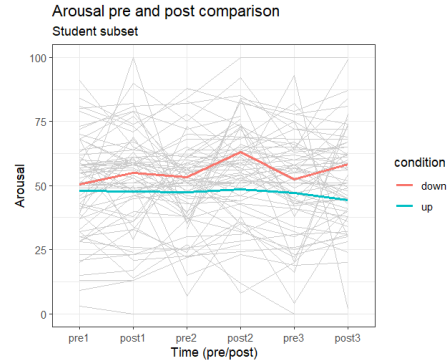
Name	Statistic	p	Mean pre	Mean post	Mean difference
Upward arousal	158	0.91	47.4	48.7	1.3
Downward arousal	70	< 0.01	53.3	63.0	9.8
Upward valence	286.5	< 0.01	57.8	50.5	-7.3
Downward valence	171	0.13	65.0	70.6	5.6

Table 37. Results of Wilcoxon signed-rank tests - Round 3.

Name	Statistic	p	Mean pre	Mean post	Mean difference
Upward arousal	228	0.36	47.2	44.3	-2.9
Downward arousal	133	0.02	52.5	58.3	5.8
Upward valence	282.5	0.03	51.3	44.0	-7.3
Downward valence	93	0.001	62.3	70.8	8.5

Table 38. Results of Wilcoxon signed-rank tests - All rounds combined.

Name	Statistic	p	Mean pre	Mean post	Mean difference
Upward arousal	1760.5	0.64	47.6	46.9	-0.6
Downward arousal	1079.5	<0.01	52.1	58.7	6.6
Upward valence	2678	<0.01	56.3	48.4	-7.8
Downward valence	1020	<0.01	62.8	70.7	7.8

**Fig. 35.** Valence pre and post - student subset.**Fig. 36.** Arousal pre and post - student subset.

E Translated quotes

Table 39 contains all original quotes and their translations used in the results section.

Table 39: Translated quotes.

Dutch quote	English translation
Ik wou kijken of ik het beter had gedaan dan gemiddeld	I wanted to see if I did better than the average.

Ik wilde zien hoe ik het deed in vergelijking met anderen, om te zien of mijn score beter of slechter was dan gemiddeld	I wanted to see how I was doing compared to others to see if my score was better or worse than average
Omdat ik dacht dat het goed ging en ik die bevestiging zocht	Because I thought it went well and I was looking for that confirmation.
Ik wil kijken hoe ik het doe ten opzichte van andere mensen om te kijken of ik voldoe aan het gemiddelde. Als ik het beter doe dan voel ik dat ik het goed heb gedaan.	I wanted to see how I did compared to others to see if I met the average. If I do better, then I feel like I did well.
Omdat ik het gevoel had dat het niet geweldig goed ging en om dan te zien dat andere ongeveer hetzelfde scoren dan haalt dat de minderwaardiging weg	Because I had the feeling it was not going amazing and to then see that others are scoring similarly takes away from the feeling of being less
Dit motiveert om door te gaan op hetzelfde tempo, of juist een tandje bij te steken.	It motivates me to keep going at the same pace, or to step up a notch.
Omdat ik nieuwsgierig naar was naar hoe anderen de puzzel maakte	Because I was curious about how the others made the puzzle
Omdat ik een high score voor mezelf probeer te krijgen. Wat anderen doen maakt mij nog niks uit, dat komt later wel als ik er daadwerkelijk goed in ben	Because I am trying to achieve a high score for myself. What others do does not matter yet, that will come later when I am actually good at it
Ik was benieuwd hoe ik het deed, ik merkte namelijk dat ik het nogal lastig vond om woorden te vinden die meer dan 3 letters hebben.	I was curious how I did, I noticed I found it quite difficult to find words with more than three letters.
Omdat ik dacht dat ik onder gemiddelde presteerde en ik wilde daar niet mee geconfronteerd worden	Because I thought I performed below average and I did not want to be confronted with that.
Ik vond dat ik het best goed deed, maar wanneer ik zag dat ik het minder goed deed dan anderen voelde ik me minder goed over mijn eigen "prestatie".	I thought I did quite well, but when I saw that I did less good than others, I felt less good about my own "achievement".
Ik was toch minder goed dan gemiddeld dus dan maakt het niet meer uit	I was less good than average so then it does not matter anymore

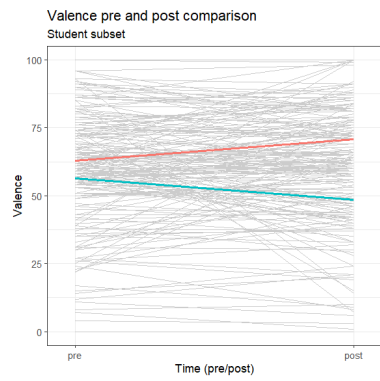


Fig. 37. Valence pre and post - student subset.

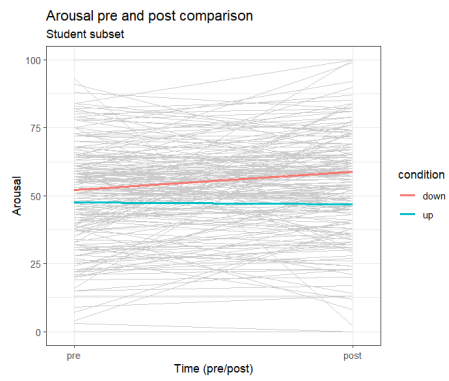


Fig. 38. Arousal pre and post - student subset.