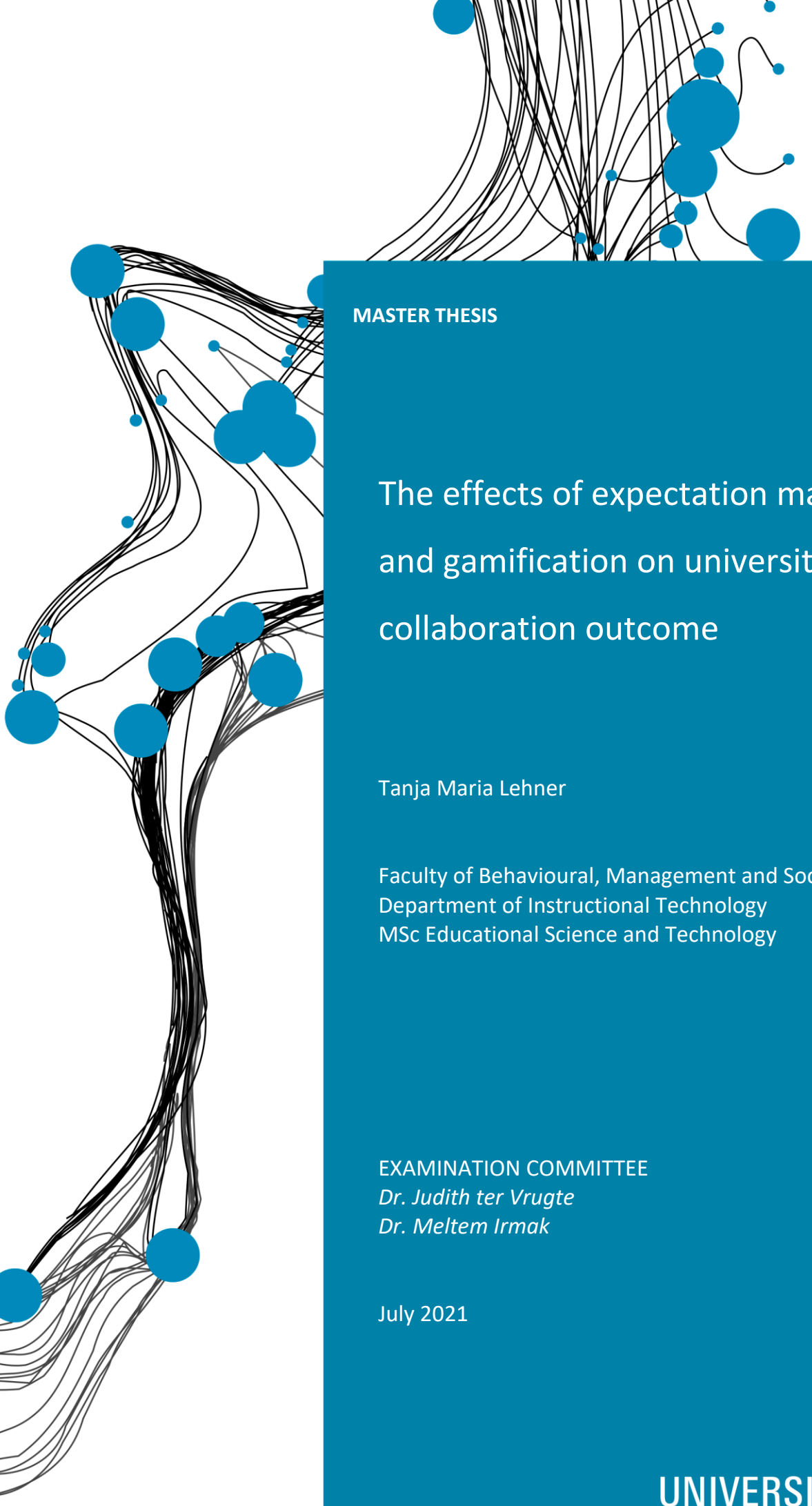


MASTER THESIS

The effects of expectation management and gamification on university students' collaboration outcome

Tanja Maria Lehner

Faculty of Behavioural, Management and Social Sciences
Department of Instructional Technology
MSc Educational Science and Technology

EXAMINATION COMMITTEE
Dr. Judith ter Vrugte
Dr. Meltem Irmak

July 2021

UNIVERSITY OF TWENTE.

Acknowledgement

I would like to express my gratitude to a number of people, without whom I would have not been able to finish this thesis. First of all, a big thank you to Dr. Judith ter Vrugte, who supported me throughout the entire process, from the initial brainstorming for this project, through preparing and conducting the data collection, to writing the final version of my thesis. All the time and effort she put into continuously providing me with valuable and challenging feedback has motivated me to bring my thesis to the quality it has now. I would also like to express my gratitude to my second supervisor, Dr. Meltem Irmak, for her support and for her insightful and valuable inputs on my thesis. Special thanks also go to my fellow student Diane Zievinger, who was a very reliable and dedicated partner for the intense period of data collection. It was a pleasure working with all of you. Last but not least, I would like to thank my friends, my boyfriend and my family. I am very grateful for all your support, encouragement and belief in me during this project and always.

Abstract

Differences in students' communication styles and interaction patterns and associated mismatched expectations can lead to less optimal collaboration. Supporting students in managing their expectations is considered a promising strategy. However, sometimes students do not succeed in engaging in additional support at a level where they could benefit from its added value. To achieve engagement in instructional support gamification can be used. Thus, this study aimed to investigate to what extent expectation management and the combination of expectation management and gamification influence university students' collaboration outcome. To research this, a quasi-experimental design with two time-points for a post-test was used. The sample included 322 first-year psychology bachelor students of a Dutch university. Students in the two experimental conditions (non-gamified expectation management support and gamified expectation management support) were compared to a control condition without expectation management and gamification. Students' collaboration outcome was measured through their products and behaviour. The results of the collaboration products showed that students benefit from instructional collaboration support that integrates expectation management and gamification. Although this effect was not reflected in the data of the collaboration behaviour, we carefully conjecture that the combination of expectation management and gamification can positively influence university students' collaboration outcome, as indicated by the collaboration products. Another finding regarding the two post-tests of the study is that in most cases analysed, students did not differ in their collaboration outcome over time and in their change over time. Explanations for these findings as well as limitations, implications and ideas for future research are discussed.

Keywords: expectation management, gamification, collaboration, university education

Table of contents

Introduction **4**
Theoretical framework **6**
Expectation management 6
Gamification 8
Present study **10**
Research design and methods **12**
Research design 12
Participants 12
Course assignments (collaboration products) 13
Materials 14
Expectation management questionnaire 14
Group discussion prompt 16
RIDE assessment tool (collaboration behaviour) 17
RIDE characteristics introduction video 17
Online collaboration support environment 18
Gamification 20
Differences between conditions 22
Procedure 25
Data analysis 26
Collaboration products 26
Collaboration behaviour 28
Results **30**
Effects on the collaboration products 30
Individual collaboration product 30
Group collaboration product 34
Effects on the collaboration behaviour 37
Extra – Difference between self- and peer-assessments 40
Discussion and conclusion **47**
Research findings 47
Effects on the collaboration products 47
Effects on the collaboration behaviour 50
Limitations, implications and future research 55
References **57**
Appendices **63**
Appendix A. Assessment forms and grading rubric for the collaboration products 63
Appendix B. Expectation management questionnaire items and presented results 66
Appendix C. Factor analysis of expectation management questionnaire items 70

Introduction

Collaboration is an important part of the curriculum in university education. However, having two or more students work together on a task creates additional challenges for students (Blumenfeld, Marx, Soloway, & Krajcik, 1996). To address this added complexity, it is widely recognised that universities need to provide instructional support to students (e.g. Bird & Osland, 2005; Lawrence, 2013; Popov, 2013). This is because if students are not adequately supported in working together, it can lead to less optimal collaboration resulting in poor outcomes (Järvelä et al., 2016). Jones (1996) points out that suboptimal collaboration might also cause students to develop negative attitudes towards collaboration or disruptive collaboration styles; which in the long run defeats the purpose of letting students work together during their education. Therefore, this study aims to contribute to understanding how students can be adequately supported in their collaboration.

One of the many challenges in collaboration arises, according to Bird and Osland (2005), from the underlying idea that people expect others to be similar to them. These expectations, however, do not correspond to reality. Students can differ in their attitudes, communication styles and behaviour patterns (Blumenfeld et al., 1996; Popov, 2013). This mismatch between expectations and reality can then lead to misunderstandings resulting in less optimal collaboration (Bird & Osland, 2005; Lawrence, 2013). To overcome this challenge instructional support should help students to create appropriate expectations and understand the similarities and differences that exist in the group (Bird & Osland, 2005). It is assumed that being aware of these characteristics and knowing what to expect from each other helps to build a shared understanding of how to work together in the group. This common ground is, according to Popov (2013), an important basis of effective collaboration. And therefore, instructional support that includes clarifying for all group members what can realistically be expected from each other before the collaboration starts, i.e. expectation management of the students, seems to be a promising strategy. However, research shows that even if students receive support, they might be reluctant to engage at a level where they can benefit from this added value (Lajoie, 2005).

A promising method to facilitate this necessary student engagement in instructional support is gamification. Gamification enables people to engage in tasks they would otherwise not find attractive by making them more game-like (Sailer, Hense, Mayr, & Mandl, 2017). Elements of game design, such as avatars or badges, can be used to transfer the motivational power of games to other contexts such as education (Sailer & Homner, 2019). Besides the motivational power, Falloon (2010) demonstrates another possible advantage of gamification in this context, namely that the use of avatars can help students feel more comfortable communicating with others about personal

information, such as characteristics or expectations, which is essential for students' expectation management. Sillaots (2014) adds that the use of avatars can create a safe environment and influence attitude. Previous research has also shown that gamification can affect action and behaviour (Landers & Landers, 2015) and create an enjoyable and satisfying experience (Bakhanova, Garcia, Raffe, & Voinov, 2020). It can also provide opportunities to foster deep and frequent learning processes (Tobias, Fletcher, Bediou, Wind, & Chen, 2005) and positively impact learning outcomes (Sailer & Homner, 2019; van Roy & Zaman, 2017).

Thus, this research aims to contribute to the understanding of how students can be adequately supported in their collaboration, specifically by investigating the effects of expectation management and gamification on university students' collaboration outcome. To research this unique combination, expectation management and a combination of expectation management and gamification will be integrated into online instructional support in collaboration and compared to a control condition (without expectation management and gamification). The outcome of this study will contribute to a better understanding of how to design (online) instructional support for students' collaboration.

Theoretical framework

Expectation management

Generally, expectations are defined as beliefs or feelings about what will or should happen (Oxford University Press, 2020a). Specifically, people have the expectation that others are similar to them (Bird & Osland, 2005) and thus that others will show similar styles and patterns when working together. This, however, does not align with reality as students show differences in their interaction (Blumenfeld et al., 1996; Popov, 2013). Research shows that if expectations are not met due to this discrepancy between expectations and reality; frustration, misunderstanding and conflict can arise during interaction with others (Bird & Osland, 2005; Popov, 2013). This, in turn, can negatively impact students' collaboration (Popov, 2013).

More specifically, previous research has shown that a major challenge for students in collaboration is that they can show differences in their attitudes, communication styles and behaviours for interacting with each other (Blumenfeld et al., 1996; Popov, 2013). These differences arise, according to Razmerita and Brun (2011), for example, from different personalities, educational experiences or cultural backgrounds. Cultural differences in communication and interaction between individuals are widely recognised in research. A taxonomy that describes precisely these differences is the one by Meyer (2014) and it conceptualises potential conflict points in collaboration at the individual level, reflected in eight dimensions. These dimensions are communicating, evaluating, persuading, leading, deciding, trusting, disagreeing and scheduling (Meyer, 2014). Although the taxonomy is informed by cultural differences, it is important to note that the dimensions also fit collaboration regardless of culture because they align with activities or skills for effective student collaboration identified in the literature (e.g. Mickan & Rodger, 2000; Saab, van Joolingen, & van Hout-Wolters, 2007).

In particular, the eight dimensions are bipolar continua:

- Communicating varies from high context (implicit) to low context (explicit) communication;
- Evaluating ranges from indirect negative feedback to direct negative feedback;
- Persuading varies from concept-first (deductive) to application-first (inductive) type of reasoning;
- Leading differs from a (perceived) hierarchical to a (perceived) egalitarian structure;
- Deciding ranges from individual to consensual decision-making;
- Trusting differs from relationship-based (social, personal) to task-based (practical, pragmatic);
- Disagreeing ranges from avoiding confrontation to a confrontational approach;
- Scheduling varies from flexible to linear or structured task and time management (Meyer, 2014).

Since these seven dimensions have ranges, it means that students within a group can vary in terms of communicating, evaluating, persuading, leading, deciding, trusting, disagreeing and scheduling. To overcome the challenge caused by these student differences and the associated mismatched expectations, Cagiltay, Bichelmeyer, and Kaplan Akilli (2015) suggest that support is needed specifically in raising students' awareness and managing their expectations.

Managing expectations is the attempt to avoid frustration or disappointment by determining what can be provided, achieved or done before actually starting to work on a particular task (Oxford University Press, 2020b). Wick (2013) adds that expectation management implies open communication to clarify for everyone involved what can realistically be expected of each other. Bird and Osland (2005) point out that people often have these expectations without being aware of them. By making students aware of their own traits and preferences in advance and by making them aware of the traits and preferences of their group members through sharing, expectations can be aligned and a common ground can be found in the group. This shared understanding is an important aspect of effective collaboration according to Popov (2013). Hence, raising awareness of students' own and each other's expectations is considered a promising strategy to overcome the challenge of subconscious mismatched expectations that affect collaboration (Cagiltay et al., 2015). Previous research has suggested that students can potentially gain more from collaborating if they know what they can expect from each other in specific situations (Bird & Osland, 2005; Cagiltay et al., 2015; Popov, 2013). This means that supporting students in managing their expectations might potentially positively affect their collaboration outcome.

Supporting students in expectation management can however bring new challenges. A widely acknowledged challenge in providing students with instructional support identified in the literature is their engagement (e.g. Challco, Mizoguchi, & Isotani, 2016; Lajoie, 2005). In particular, students do not always succeed in engaging in additional support at a level where they could actually benefit from its added-value (Lajoie, 2005). Active engagement of students in instructional support is an overarching goal of gamification in education (Dicheva & Dichev, 2015; Giannetto, Chao, & Fontana, 2013) and thus seems to respond to this challenge. Specifically, Giannetto et al. (2013) explain that the addition of game design elements is intended to make a task more attractive to students and thus lead to completion and higher engagement. An additional difficulty in the context of expectation management is that students might be hesitant to communicate personal characteristics with others (Pickard, Roster, & Chen, 2016), this, however, is an essential part of managing expectations in collaboration. To create a safe space to foster sharing this kind of information, gamification can also be used. Studies of Sillaots (2014) and Falloon (2010) suggest that avatars could be especially useful for this in online settings as in the present study. Baccon,

Chiarovano, and MacDougall (2019) point out that avatars have the power to combine the advantages of face-to-face and online settings. In particular, avatars could give students the feeling that they do not talk to a "real person" and feel less judged than when talking to a real human (Pickard et al., 2016). Moreover, in online environments, avatars can make it easier to bond and build trust between students compared to an environment without avatars, as students feel more self-present in the online environment (Pan & Steed, 2017), which is known to lead to greater self-disclosure (Hooi & Cho, 2014). To sum up, the benefits of gamification appear to be twofold; facilitate both an engaging as well as a safe environment for expectation management.

Gamification

Gamification is defined as adding game design elements to any activity outside of games (Deterding, Dixon, Khaled, & Nacke, 2011) to make it a more game-like experience (Werbach, 2014). Game design elements are elements that often occur in games and therefore characterise them (Deterding et al., 2011). Deterding et al. (2011) argue that they are the fundamental components of any gamified environment. The elements most used in education are points, badges, and leader boards (Dicheva & Dichev, 2015). Avatars, however, have received little attention in literature until now. Avatars represent the students in the gamified environment and can typically be selected from a predefined set, adapted or created by the students (Werbach & Hunter, 2012). In games, avatars typically have specific strengths and weaknesses as well as series of traits and abilities that can be developed throughout the gameplay (Kromand, 2007). One way to evolve avatars over time and show their increased experience is through the use of badges. The widely used badges are visualised accomplishments that can be earned (Werbach & Hunter, 2012). Sailer et al. (2017) argue that they can be seen as feedback for demonstrated performance or behaviour, and they can also be status symbols (Antin & Churchill, 2011), for example, to represent the avatars' skills.

By using game design elements, gamification aims to motivate behaviour and attitudes in a goal-oriented, desired manner (Dicheva & Dichev, 2015; Sailer & Homner, 2019), and promote performance for the gamified task (Sailer et al., 2017). A theory that explains the influence of gamification on behaviour, and outcomes is the theory of gamified learning of Landers (2015). This theory describes that gamification directly influences behaviour and attitude, which in turn influence the learning outcome (Landers, 2015). Research has already applied and confirmed this theory (e.g. Landers & Landers, 2015; Sailer & Homner, 2019). However, it is important to note that the behaviour that gamification is intended to facilitate must itself be related to the desired (learning) outcome (Landers, 2015). In the case of collaboration, this can be done, for example, by using avatars to foster the communication of personal traits and preferences, as this can help in finding a common

ground within the group improving collaboration. Another example would be rewarding respectful communication, as this is behaviour contributing to effective collaboration according to Saab et al. (2007). To conclude gamification might potentially positively affect students' behaviour and outcomes in the collaboration.

Present study

Based on the presented theoretical framework we conclude that supporting students in expectation management could benefit students' collaboration and contribute to the value of collaboration in education. Gamification might add to this value and further strengthen this effect, by harnessing an engaging and safe setting.

Therefore, the present study adopts a unique combination of instructional support, expectation management and gamification, which based on current theory and research findings in fields of behavioural change, psychology and educational science is promising, but which, to the best of our knowledge, has not been studied yet. Besides the scientific relevance, the study could also contribute to practice by providing practitioners with a better understanding of how to design (online) instructional support for student collaboration in higher education.

In summary, this study aimed to investigate the effectiveness of expectation management and a combination of expectation management and gamification in online collaboration support. In particular, the following overarching research question was formulated:

To what extent do expectation management and the combination of expectation management and gamification influence university students' collaboration outcome?

To answer this research question, the current study was conducted with students from the University of Twente located in the Netherlands. For the participating students, the study was part of their regular coursework. As part of this, the students worked on two consecutive projects in which their collaboration outcomes were measured. The collaboration outcome in this study consisted of the collaboration products and collaboration behaviour of the participating university students. To investigate the effects of expectation management and gamification on these collaboration outcomes, three conditions were needed: two experimental and one control condition. The conditions differed in terms of the online collaboration support they received. In the first experimental condition, students received support on expectation management (without gamification). In the second one, students received expectation management support with gamification in the form of avatars and badges. In the control condition, students received no expectation management support (and no gamification was used).

Therefore, the hypotheses of this study were the following:

Hypothesis collaboration products – Difference among the three conditions (H1a). Based on the presented literature, it was hypothesised that students in the gamified expectation management

condition have the best collaboration products, followed by the non-gamified expectation management condition, and the control condition in the last place.

Hypothesis collaboration products – Difference over time (H1b). As the participating students worked on two consecutive projects, it was expected that students in all three conditions improve in their collaboration products over time (between the two projects), but that students in the gamified expectation management condition improve more than students in non-gamified expectation management condition and the control condition.

The overall improvement (all conditions) was hypothesised because in the second project students in all conditions already know each other and have experience in working together, which could allow them to collaborate more efficiently benefitting the collaboration products (Løvold, Lindsjørn, & Stray, 2020; Zambrano, Kirschner, Sweller, & Kirschner, 2019). The reason why it was expected that the gamified expectation management condition increases more compared to the other two conditions is the design of the gamification in the collaboration support of this study. Specifically, only one of the two game design elements incorporated in the study was introduced to the students at the time the data for *project one* was collected, i.e. avatars were introduced to them, but badges were only introduced to them after finishing *project one* (see the section on the online collaboration support environment for more information on the gamification design). For this reason, it was expected that the gamification effect gets a boost and unfold its full potential during *project two*.

Hypothesis collaboration behaviour – Difference among the three conditions (H2a). Based on the presented literature same as for the collaboration products, it was hypothesised that students in the gamified expectation management condition perform best in terms of their collaboration behaviour, followed by the non-gamified expectation management condition, and the control condition in the last place.

Hypothesis collaboration behaviour – Difference over time (H2b). It was also expected that students in all conditions improve in their collaboration behaviour over time (between the two projects), but that students in the gamified expectation management condition improve more than in the other two conditions.

Students' overall improvement was also expected for this measurement because previous research suggests that supporting students over a longer period of time and repeatedly experiencing collaboration benefits students' collaboration behaviour (Kollar, Fischer, & Slotta, 2007). The reason why the gamified expectation management condition was expected to increase more in terms of the collaboration behaviour compared to the other two conditions was, same as for the collaboration products, that the gamification design was fully introduced to the students only after *project one* was completed and thus a boost was expected during *project two*.

Research design and methods

Research design

To investigate the extent to which expectation management and the combination of expectation management and gamification influence university students' collaboration outcome, a quasi-experimental design with two time-points for a post-test was used. The independent variable was the condition students were in, i.e. the type of support students received, and consisted of three categorical groups: the non-gamified expectation management condition, the gamified expectation management condition, and the control condition. The dependent variable of this study was the collaboration outcome. The outcome consisted of the individual score for the individual collaboration products, the group scores for the group collaboration products, and the individual score of the self- and peer-assessed collaboration behaviour. These data were collected at two moments in time; once after *project one* and the second time after *project two* of the participating students. A quantitative approach was used to collect the data to allow for a large number of participants within a limited timeframe.

Participants

A total of 336 university students participated in this study. All first-year psychology bachelor students of the University of Twente attending the Design and Evaluation course in the Cognition and Development module were selected as participants based on convenience sampling. Eight students did not give consent to use their data in this study and six additional students did not complete any task of the study. Therefore the sample included 322 students[1]. Of these, 213 were female (67.4%), 102 male (32.3%) and one other (0.3%). The participants ranged in age from 17 to 38 years (M= 20.39 years, SD = 2.25). 88 of them identified themselves as Dutch, 179 as German and 49 as other. The largest group within the other category were Romanian with six students and Spanish, Greek and Finnish with three students each.

To fulfil the two group projects, the teachers of the course formed 63 groups of three to six students each by random assignment. The groups were assigned to the conditions by the researchers. All groups met with a tutor weekly during tutor meetings (directed at students welfare and study conditions). Multiple groups would join a meeting at the same time. Though the tutors did not play a role in the current study it was assumed that students might casually talk about the support environment during those meetings. Therefore, researchers ensured that students that joined the same tutor meeting were assigned to the same condition. This was done to safeguard

[1] Demographic data were available for 316 students.

validity by minimising the risk that students could talk about differing tasks and uncover the different conditions. After allocation the non-gamified expectation management condition consisted of 20 groups (103 students), the gamified expectation management condition had 19 groups (95 students), and the control condition included 24 groups (124 students). More detailed information about students' demographics in the three conditions is shown in Table 1.

Table 1

Demographic information of participants per condition

	Non-gamified expectation management condition (N = 103)	**Gamified expectation management condition (N = 95)**	**Control condition (N = 124)**
Gender			
% Female	66.0	64.2	71.2
% Male	34.0	35.8	28.0
% Other	.0	.0	.8
Age			
Mean	20.49	20.37	20.32
Standard deviation	2.00	2.93	1.77
Nationality			
% Dutch	24.3	28.4	30.5
% German	57.3	58.9	54.2
% Other	18.4	12.6	15.3

Note. N = Number of students

Course assignments (collaboration products)

The assignments that students worked on were part of the normal coursework of the participating students in the third quartile of the academic year. There were two assignments; *project one* and *project two*, which were in total running over 10 weeks. Each project was complemented with an individual and a group collaboration product. These were scored by the teachers of the course and ranged from a minimum of 1 to a maximum of 10, according to the Dutch grading system. The assessment forms and rubric used for grading can be found in Appendix A.

Project one focused on collaboratively writing a scientific article on the replication of a cognitive experiment. The students received the original study and data and were asked to replicate and report the results in form of a scientific article (group collaboration product). In addition, students were required to write an individual speed essay about the content of this project (individual collaboration product).

Project two was about collaboratively evaluating and redesigning a product. The students were asked to conduct a heuristic evaluation of a product and, based on the results, to redesign the product and create a digital prototype, which was then tested in a usability assessment. All these steps had to be described in form of a report (group collaboration product). At the end of this project, students also had to write an individual speed essay about the project content (individual collaboration product).

Materials

Expectation management questionnaire

An expectation management questionnaire was designed to help students become aware of their own traits and preferences when collaborating and to serve as a feed for the group discussion to find a common ground within the group.

Initial version. The initial version of the questionnaire was adapted from a questionnaire developed by Ravi Shankar (2017) based on the eight dimensions of Meyer (2014): communicating, evaluating, persuading, leading, deciding, trusting, disagreeing and scheduling. It consisted of 24 items representing these eight dimensions on a bipolar Likert scale ranging from -3 to +3. A pilot was conducted to test the quality of the expectation management items of this initial version of the questionnaire. This was done with 32 biomedical engineering master students of the University of Twente attending the Applied Cell Biology course. The internal consistency of the items was tested and indicated that they were not reliable. Reasons for the low internal consistency were that the items lacked a clear direction and the phrasing seemed too difficult or inappropriate for the target population of this study i.e. students, as the original questionnaire and the dimensions stem from an organisational setting. Based on these insights, the questionnaire was completely revised in line with literature in the context of student collaboration in higher education. First of all the leading dimension was renamed to power distancing and the persuading dimension was completely removed from the questionnaire because it did not fit well enough into the context of the study. The other dimensions were deemed valuable and were therefore retained. With the seven remaining dimensions in mind, the items were carefully reviewed. Overall, this resulted in eight items remaining the same, nine items being adjusted, seven being removed and six items being added. During the

revision process, it was also sought to have more balanced items per dimension, from two to five items per dimension to three to four items per dimension.

Final version. The revised and final version of the questionnaire was then used for the study and consisted of seven dimensions and a total of 23 bipolar items. The seven dimensions inspired by Meyer (2014) were as follows: communicating (from implicit to explicit), evaluating (from indirect to direct), power distancing (from hierarchical to egalitarian), deciding (from individual to consensual), trusting (from relationship-based to task-based), disagreeing (from avoiding confrontation to confrontational) and scheduling (from flexible to linear). More information about the dimensions and their items is shown in Table 2. For each item, students rated where they would fit on a 7-point bipolar Likert scale ranging from -3 to +3.

After the students had rated all items, their results were presented to them at the end of the questionnaire and separately in a personalised email. The presented results included a brief explanation of the dimensions and how they scored on each scale. A score of <-1 indicated one side of the scale, between -1 and 1 no outspoken preference was shown and >+1 indicated the other side of the scale. These scores were not used for analysing purposes but to give students an indication of where they lie on the respective scale and become aware of their preferences. All items of the questionnaire and how the results were communicated to the students can be found in Appendix B.

Regarding the reliability of the final expectation management questionnaire, Cronbach's alpha showed that the reliability of the dimensions was good in the case of evaluating and scheduling, acceptable in case of disagreeing, poor in case of communicating, power distancing and deciding and unacceptable for trusting (see Table 2). In terms of the validity, a factor analysis with principal axis factoring using oblique rotation was performed and revealed that six items did not load as expected (one in trusting, one in deciding, one in power distancing, two in communicating, and one in disagreeing). The results can be found in Appendix C. However, it needs to be mentioned that these results were not processed further, as this questionnaire was only used as a treatment and not for analysis purposes in the current study.

Table 2

Overview expectation management questionnaire dimensions: Number of bipolar items, example items and Cronbach's alpha

Dimension	Number of items	Example item (bipolar)	Cronbach's alpha
Communicating	4	"My communication is sophisticated, layered and nuanced" – "My communication is clear, concise and direct"	.58
Evaluating	4	"If I have done poor work, I prefer to get soft and subtle feedback" – "If I have done poor work, I prefer to get frank, blunt, and honest feedback"	.76
Power distancing	3	"If I do not agree with my teachers, I will not express my opinion to them" – "If I do not agree with my teachers, I will express my opinion to them"	.58
Deciding	3	"I think decisions can be made by an individual in the group" – "I think decisions should involve everybody in the group"	.59
Trusting	3	"I prefer to invest time just getting to know my fellow group members – without discussing group work much" – "I prefer to talk only about group work with my fellow group members"	.47
Disagreeing	3	"I believe that open debate and discussion is likely to ruin relationships" – "I believe that open debate and discussion is an indicator or a healthy group"	.64
Scheduling	3	"I value flexibility over organization and structure" – "I value organization and structure over flexibility"	.72

Group discussion prompt

At the beginning of the collaboration student groups were prompted to discuss factors to take into account during their upcoming collaboration. These prompts were either unstructured or structured (see the section on differences between conditions in the online collaboration support environment for more information). Then, the student groups were asked to write down how the discussed factors would impact their collaboration and how the group planned to address them. These entries were not used for analysing purposes but to ensure that the groups were discussing.

RIDE assessment tool (collaboration behaviour)

To measure the collaboration behaviour for *project one* and *project two*, the RIDE characteristics of Saab et al. (2007) were used. RIDE stands for 1) respect, 2) intelligent collaboration, 3) deciding together, and 4) encouraging (Saab et al., 2007) and according to Eshuis et al. (2019), these four characteristics or behaviours underpin successful collaboration. Self- and peer-assessments of these RIDE characteristics were already used in studies to reflect and assess collaborative behaviour (e.g. Eshuis et al., 2019). In this study, the self- and peer-assessment was done with an existing tool.

In particular, the existing RIDE assessment tool consisted of three steps. The first step within the tool was for students to rate the collaboration behaviour of themselves and their group members based on each of the four RIDE characteristics on a ten-point scale. These ratings ranged from 1 to 10 each, with 1 indicating very poor collaboration behaviour and 10 perfect collaboration behaviour. If students needed more information regarding the RIDE characteristics they could click on information buttons. After all group members finished this step and clicked “ready”, they were shown the second step, in which students could view the outcomes of their evaluation in the form of graphs. These showed the average group score for each RIDE characteristic and through clicking on a particular characteristic also the self- and peer ratings for each group member. After all group members clicked “ready”, the final step asked them to discuss what went well and what could be improved regarding each RIDE characteristic and to write down their group goals for future collaboration, so what they will continue to do or what they will improve. Besides these guiding questions, students could see the graphs from the previous step and information about the RIDE characteristics. The entries in the text boxes were not used for analysing purposes, only the scores from the self- and peer evaluation were used. These self- and peer-ratings were computed to an average score for each RIDE characteristic for each student.

RIDE characteristics introduction video

To ensure that students were able to follow and evaluate each other based on the four RIDE characteristics students were introduced to them in form of a video. This video was a screencast or digital recording of a PowerPoint presentation with voice-over narration and was created by the researchers. It first explained why the RIDE characteristics are important for successful collaboration. This was followed by a description of each characteristic and its sub-rules (see Table 3), which were slightly adjusted from the study of Eshuis et al. (2019) and Saab et al. (2007) to fit the current context. Finally, the video ended with a summary.

Table 3

RIDE characteristics and its sub-rules

RIDE characteristics	Sub-rules
Respect	• Give others the opportunity to contribute • Pay attention to what others say and do • Don't judge someone by a mistake they make • Take other's contribution serious
Intelligent collaboration	• Share all relevant information and theories • Proactively provide sufficient information • Ask for explanations and elaborations when you need it • Be critical towards all contributions (but don't judge somebody based on their contribution)
Deciding together	• Share what you are about to decide • Check with others before you make a decision • Actively contribute to the decisions being made
Encouraging	• Involve others in the process • Encourage others to participate • Give compliments when others make a meaningful contribution

Note. Sub-rules adjusted from the study of Eshuis et al. (2019) and Saab et al. (2007).

Online collaboration support environment

For the purpose of this study, an online collaboration support environment was created. It was developed with the authoring platform Graasp[2]. There were three different versions of the environment; one for each condition. This study was part of a bigger data collection that is why more components are described in this part. In general, as shown in Figure 1, the support environment was segmented into tabs about an overview, a series of four tasks, and contact. Besides these tabs, the created environment also included an outline of the collaboration group that students were in and a chat function.

[2] https://graasp.eu/

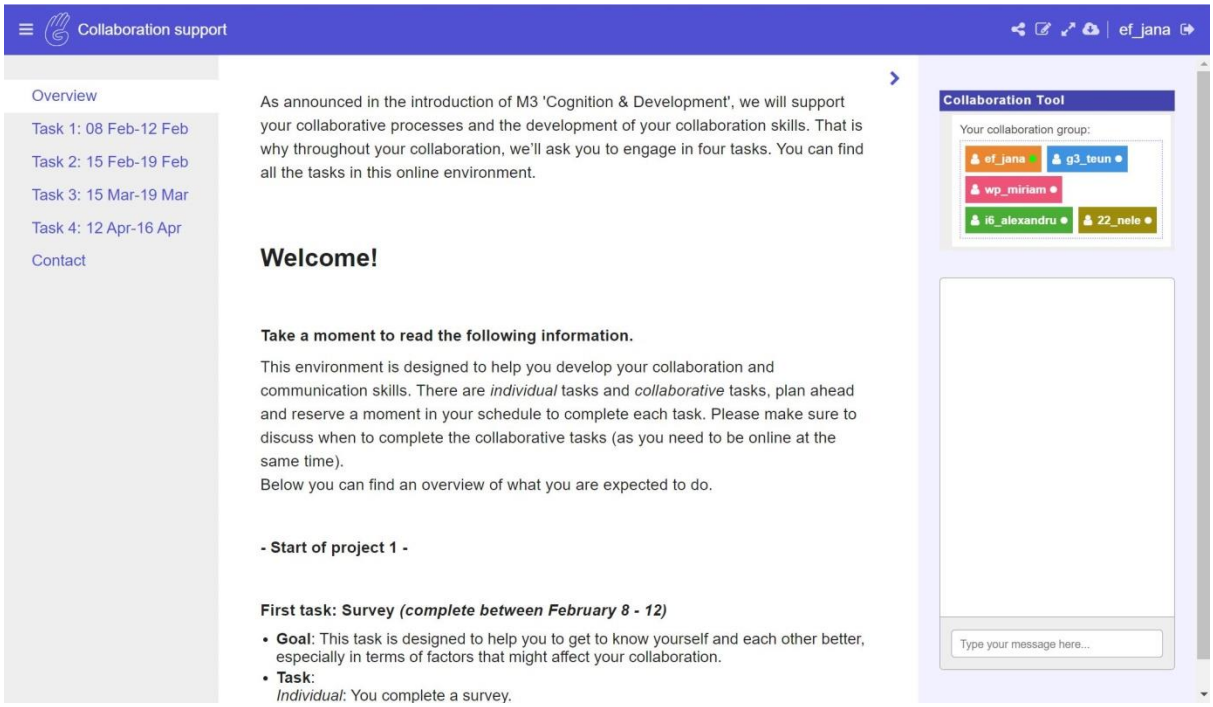


Figure 1. Online collaboration environment (non-gamified expectation management condition).

The online collaboration support environment of all three conditions included the following:

Overview. The first tab showed an introduction and an overview of the specific tasks in the environment, including the goal, whether the task needed to be performed individually or collaboratively, a short description, the estimated time to complete the task and when in the collaboration students were expected to do them.

Task 1. The second tab presented *Task 1* starting with an introduction and then outlining the steps of the task. This task consisted of an individual and a collaborative part. The first step was to click on a link to complete a questionnaire individually, followed by the group discussion prompt. In this questionnaire, students were asked for their consent and answered demographic questions (gender, age, nationality the students identify with), as this questionnaire was the first one in the study. Then, depending on the condition, the questionnaire included or omitted the expectation management items. In the end, it also included six items on students' knowledge sharing self-efficacy for all conditions, since the study was part of a larger project, as mentioned earlier.

Task 2. The third tab showed *Task 2*. It consisted of a brief introduction and one step in which the students were asked to individually watch the RIDE characteristics introduction video. The summary of the video was also provided in PDF format to allow students to quickly look something up without having to re-watch the entire video.

Task 3. The next tab presented *Task 3* and in this task, students were shown a brief introduction and the RIDE assessment tool for *project one*. To complete the tool students had to be online at the same time to work on the task collaboratively.

Task 4. The fifth tab showed *Task 4*. Similar to *Task 3*, it started with a brief introduction and the RIDE assessment tool. This time, however, students were asked to complete the tool together for *project two*. Since the study was part of a bigger data collection, in addition to the RIDE assessment tool, the students were asked to click on a link to complete a questionnaire as the final step. The questionnaire focused on how students perceived the overall collaboration in their group for *project one* and *project two*.

Contact. The last tab displayed contact information about the researchers, specifically their names, email addresses and pictures, with the intention of allowing students to easily reach out to the researchers in case of questions.

Gamification

The collaboration support environment was gamified in this study by using avatars and badges, specifically by integrating the creation of avatar passports and badges assignments into the online environment.

Avatar passport creation. Students were asked to create their own personal avatar that represented themselves. Based on previous research from Falloon (2010) and Sillaots (2014), avatars were chosen as game design elements as they might have especially an added value in the current study to create the necessary safe space for communicating with others about personal information, such as characteristics or expectations. All avatars of a group were displayed in the form of avatar passports in an additional tab in the online collaboration support environment. More specifically, the passports contained a column for each group member with their first names pre-filled by the researchers to avoid the situation of several students simultaneously entering information in the same column. Within these pre-assigned columns, there were predefined fields for uploading a picture of the avatar, its skillset and achievements as shown in Figure 2. To create the *picture*, students were provided with a link to a free website[3]. This website offered students multiple design options that they could choose from (e.g. different styles and colours for face, eye, hair, clothes, and background) and allowed for downloading the picture of their avatar without registration. After downloading, students could add the picture to their avatar passport. To fill in the avatar's *skillset* students were asked to enter the scores from the expectation management questionnaire. The *achievements* remained empty in the beginning and were filled during the badges assignments.

[3] https://avatarmaker.com/

Badges assignments. The achievements or badges were used to develop avatars over time (*project one* and *project two*) and connect desired student behaviour to the avatars. Previous research notes that it is important that the behaviour that gamification is intended to facilitate must itself be related to the desired (learning) outcome (Landers, 2015). This means that if collaboration should be facilitated, the gamified support should contain elements relevant to collaboration (Azmi, Iahad, & Ahmad, 2015). Relevant elements that contribute to effective collaboration are the RIDE characteristics (Eshuis et al., 2019). Therefore, this study incorporated four achievements or badges that students could earn to evolve their avatars, one for each of the RIDE characteristics. To award the badges, students were asked to discuss and decide together whose avatar deserves to evolve with which badge. Specifically, students were asked to decide for each of the four badges if and who earns it and to provide a short explanation. Students who received badges could download a picture of the respective badge and add it to their avatar passports (see Figure 2). Badges could be earned twice during the experiment; once after *project one* and once after *project two*.

Figure 2. Example of a filled-in avatar passport.

Differences between conditions

Generally, the environments for the three conditions differed mainly in 1) whether or not students received the expectation management items in the first questionnaire, 2) whether they were provided with structured or unstructured discussion prompts, 3) whether or not students were asked to create their avatar passports, and 4) whether or not they received the badges assignments (see Table 4). In addition, the gamified expectation management condition also used some different phrasing with the goal of making it more game-like. For example, instead of asking the students to discuss the outcomes of the questionnaire, they were asked to discuss the avatars' skillsets.

Table 4

Overview of the tasks in the three conditions

Tasks	Conditions		
	Non-gamified expectation management condition	Gamified expectation management condition	Control condition
Task 1	**(1) Expectation management questionnaire (individual)** **(2) Structured group discussion prompt (collaborative)**	**(1) Expectation management questionnaire (individual)** **(2) Avatar passport creation (individual)** **(3) Structured group discussion prompt (collaborative)**	(1) Questionnaire without expectation management items (individual) (2) Unstructured group discussion prompt (collaborative)
Task 2	(1) RIDE characteristics introduction video (individual)	(1) RIDE characteristics introduction video (individual)	(1) RIDE characteristics introduction video (individual)
Task 3	(1) RIDE assessment tool (collaborative)	(1) RIDE assessment tool (collaborative) **(2) Badges assignment (collaborative)**	(1) RIDE assessment tool (collaborative)
Task 4	(1) RIDE assessment tool (collaborative) *(2) perception questionnaire (individual)*	(1) RIDE assessment tool (collaborative) **(2) Badges assignment (collaborative)** *(3) perception questionnaire (individual)*	(1) RIDE assessment tool (collaborative) *(2) perception questionnaire (individual)*

Note. Differences between the control condition and the other two conditions appear in bold. Subtasks that were part of the larger data collection but were not used for this study appear in italic.

Specifically, the different conditions contained the following:

Control condition. The first tab presented an introduction and *Overview* of the tasks in the online environment.

In the second tab showing *Task 1,* students of this condition were asked to fill out a questionnaire individually that included only the demographic questions and the items on students'

knowledge sharing self-efficacy but omitted the items on expectation management. The group discussion prompt that followed was unstructured and therefore different from the other two conditions. Here, the students were only asked to think about whether there were certain factors to consider during their collaboration and if so, to write down how these things would impact their collaboration and how the group planned to address them.

The next tab, *Task 2* provided students with the video about the RIDE characteristics. *Task 3* provided students with the RIDE assessment tool for *project one*. The fifth tab presented *Task 4* and asked students to complete the RIDE assessment tool for *project two* and to fill out the perception questionnaire. And the last tab displayed *contact* information about the researchers.

Non-gamified expectation management condition. The first tab showed, same for the control condition, an introduction and an overview of the specific tasks in the online environment.

The second tab presented *Task 1*. Students in this condition were asked to complete the expectation management questionnaire individually, followed by the prompt for the group discussion. The prompt in this condition was structured and aimed to help students uncover the traits and preferences of their group members and find a common ground within their group. Specifically, students were asked to discuss in their group 1) what the results of the questionnaire were, 2) what the commonalities and differences were, and 3) what the identified communalities and differences meant for their collaboration. They were also asked to write down how the discussed factors would impact their collaboration and how they would take this into account.

The next tabs, which were *Task 2*, *Task 3*, T*ask 4* and the *Contact* tab, were the same as in the control condition. Specifically, *Task 2* provided students with the video about the RIDE characteristics, *Task 3* asked students to complete the RIDE assessment tool for *project one*, *Task 4* to complete the RIDE assessment tool for *project two* and to fill out the perception questionnaire, and the last tab also displayed contact information about the researchers.

Gamified expectation management condition. The first tab also showed an introduction and an *overview* of the specific tasks in the online environment.

In the second tab showing *Task 1* students were also asked to first complete the expectation management questionnaire. Then the students in this condition got the assignment to create their avatars and the instruction on how to fill out their avatar passports. In particular, they were asked to add the picture of their avatar and the skillset as described in the avatar passport creation section. Then, students were asked to go to the tab Avatar passport.

The tab *Avatar passport* was added as a sixth tab to the environment and provided the students with their group's blank avatar passports to fill in. In addition, this tab contained the

structured prompt for the group discussion, as described in the non-gamified expectation management condition.

The third tab presenting *Task 2* also provided the students in this condition with a video introducing the RIDE characteristics, as in the previously described conditions, to help them evaluate each other and award badges based on these characteristics after *project one* and *project two* as part of the next tasks.

In the next tabs showing *Task 3* and *Task 4*, students were also asked to do the RIDE assessment tool. Besides the purpose of measurement, this tool aimed to help students in this condition with their next assignment in this task. This assignment was about awarding badges and adding them to their avatar passport, if applicable, as described in the badges assignment section. In *Task 4*, students were also shown a link to complete the collaboration perception questionnaire.

Also in this condition, the last tab displayed *contact* information about the researchers.

Procedure

Before the data collection of this study, the Ethics Committee of the University of Twente was asked for permission. The experiment was part of students' regular coursework in the third quartile of the academic year and was running over 10 weeks. Prior to the experiment at the beginning of the module, the students were informed by their teacher about the goals of the experiment and the tasks involved, using PowerPoint slides provided by the researchers. Students' participation in this study lasted a total of one to two hours, depending on the condition they were in, and consisted of four tasks. Students performed these four tasks in an online environment, as described in the material section, at specific times of collaborating. At the beginning of the experiment, each group received the link to their condition's environment and their individual login names via email.

The students performed *Task 1* during the first week of *project one*, which took approximately 30-40 minutes for the gamified expectation management condition, 30 minutes for the non-gamified expectation management condition, and 15 minutes for the control condition. During this week, students were also asked for their consent and answered demographic questions. The following week, students in all conditions performed *Task 2*, which took about 10 minutes. *Task 3* was done the week following the end of *project one* and took approximately 30 minutes for the gamified expectation management condition and 20 minutes for the non-gamified expectation management and control condition. The final *Task 4* was completed by the students in the week following the end of *project two* and took approximately 35 minutes for the gamified expectation

management condition and 25 minutes for the non-gamified expectation management and control condition.

At the beginning of each of these four weeks, the students received an email prompt to perform the corresponding task and a (personalised) reminder at the end of the week. One week after the last task, students in the control condition received the expectation management questionnaire in order for all students to benefit equally from the research. After the course, the assignments were scored by the teacher and the data were made available for this study within four weeks.

Data analysis

For analysing the collaboration outcome, the individual collaboration product and the collaboration behaviour were analysed at an individual level and the group collaboration product was analysed at a group level. All data were quantitative and analysed via SPSS. For all analyses, the significance level was set at .05. The inclusion criteria for data usage differed for the collaboration behaviour and the individual as well as the group collaboration product. After deciding on the inclusion criteria, variables for the analysis were computed (if applicable) and the assumptions for each test to answer the research question and hypotheses were examined.

Collaboration products

Individual collaboration product

Inclusion criteria. A total of 322 students participated. However, two groups did not complete *Task 1* and 11 did not fully do *Task 3*. The criteria whether students completed a task were checked per condition, for example for *Task 3* in the gamified expectation management condition it was checked whether students did the whole RIDE assessment tool and the badges assignment. Students who did not perform *Task 4* were not removed since the individual collaboration products were handed in before this task. In total, 13 groups with 66 students were excluded (two groups with 10 students in the non-gamified expectation management condition, nine groups with 45 students in the gamified expectation management condition, and two groups with 11 students in the control condition). 18 additional students were removed because of the following reasons: five students left the module and therefore the study before *Task 3* or *Task 4* (two non-gamified expectation management, and three control condition) and eight students did not submit an individual product for *project one* (five non-gamified expectation management condition, and three control condition), seven students did not submit it for *project two* (four non-gamified expectation management condition, and three control condition), and three students did not submit it for both projects (one

gamified expectation management condition, and two control condition). The groups of these 18 additional students were kept because only one student of the group dropped out or did not submit the individual products. Therefore, 238 students (83 non-gamified expectation management condition, 50 gamified expectation management condition, and 105 control condition) were used for analysing the individual product as part of their collaboration.

Assumptions check. Before testing possible differences among the three conditions and over time in terms of students' individual collaboration product, the assumptions for a Mixed ANOVA were tested. The normality assumption was checked with the Shapiro-Wilk test. For *project one*, the results were not significant for the non-gamified expectation management condition with $W(83) = .976$, $p = .124$, the gamified expectation management condition with $W(50) = .955$, $p = .053$, and the control condition with $W(105) = .980$, $p = .112$. Regarding *project two*, the results were not significant for the non-gamified expectation management condition with $W(83) = .982$, $p = .285$ and the gamified expectation management condition with $W(50) = .970$, $p = .231$ and significant for the control condition with $W(105) = .960$, $p = .003$. Although the Shapiro-Wilk test revealed partial violation for *project two*, normality is assumed because ANOVA is quite robust to violations of normality for large group sizes as in this study. In addition, Levene's F test for the individual collaboration product showed that the assumption for homogeneity of variances was satisfied for *project one* with $F(2,235) = .245$, $p = .783$ and for *project two* with $F(2,235) = 1.205$, $p = .301$. Moreover, the homogeneity of the variance-covariance matrices assumption using Box's M test was fulfilled with $p = .905$. To investigate possible outliers in the data, the z-scores of the individual collaboration product for *project one* and *project two* were computed. One z-score was outside the range of -3 and +3 and was therefore considered an outlier. Since all other assumptions were met, a Mixed ANOVA was performed. The identified outlier in the individual collaboration product of *project two* did not influence the results of the Mixed ANOVA and therefore the results were reported on the full data including the outlier.

Group collaboration product

Inclusion criteria. A total of 63 groups participated. However, same as for the individual collaboration product two groups in the gamified expectation management condition had to be excluded because they did not complete *Task 1* and 11 additional groups had to be removed because they did not do *Task 3* (two non-gamified expectation management condition, seven gamified expectation management condition, and two control condition). Also here, the criteria of whether students completed a task were checked per condition and the groups who did not perform *Task 4* were not removed since the group collaboration products were handed in before this task. Therefore, 50 groups (18 non-gamified expectation management condition, 10 gamified expectation

management condition, and 22 control condition) were used for analysing the group collaboration product.

Assumptions check. Before testing possible differences among the three conditions and over time in terms of students' group collaboration product, the assumptions for a Mixed ANOVA were tested. The normality assumption was checked with the Shapiro-Wilk test. Regarding *project one*, the results were not significant for the non-gamified expectation management condition with $W(18) = .984$, $p = .979$, the gamified expectation management condition with $W(10) = .963$, $p = .817$, and the control condition with $W(22) = .943$, $p = .227$. Looking at *project two*, the results were not significant for the non-gamified expectation management condition with $W(18) = .984$, $p = .984$, the non-gamified expectation management condition with $W(10) = .886$, $p = .152$, and the control condition with $W(22) = .936$, $p = .166$. Also when looking at the Skewness and Kurtosis the scores were fairly close to zero and therefore normality was assumed. Furthermore, Levene's F test for the collaboration product showed that the assumption for homogeneity of variances was met for *project one* with $F(2,47) = 3.150$, $p = .052$ and for *project two* with $F(2,47) = 2.093$, $p = .135$. In addition, the homogeneity of the variance-covariance matrices assumption using Box's M test was fulfilled with $p = .316$. To examine possible outliers in the data, the z-scores of the group collaboration product were calculated for *project one* and *project two*. All z-scores were within the range of -3 and +3 and therefore no data point was considered an outlier. Since all assumptions were met, a Mixed ANOVA was performed.

Collaboration behaviour

Inclusion criteria. A total of 322 students participated. However, two groups did not complete *Task 1*, 11 did not do *Task 3* and seven did not do *Task 4*. Same as for the collaboration products, the criteria of whether students completed a task was checked per condition. In total, 20 groups with 105 students were excluded (two groups with 10 students in the non-gamified expectation management condition, nine groups with 45 students in the gamified expectation management condition, and nine groups with 50 students in the control condition). Four additional students were removed due to leaving the module (two non-gamified expectation management condition, and two control condition) but the groups were kept because only one student of the group left the module. Therefore, 213 students were used for analysing the collaboration behaviour (90 non-gamified expectation management condition, 51 gamified expectation management condition, and 72 control condition).

Variable computation. Before testing possible differences among the three conditions and over time in terms of students' collaboration behaviour, the average score for each RIDE

characteristic for each student (based on self- and peer-ratings) were manually extracted from the teacher dashboard of the collaboration environment. These four scores per project and student were used to compute two new variables: an average score based on the four RIDE characteristics for each student, once for *project one* and once for *project two*.

Assumptions check. Using these newly computed collaboration behaviour variables for *project one* and *project two* the assumptions for a Mixed ANOVA were checked. The normality assumption was tested with the Shapiro-Wilk test and showed that this assumption is partially violated for *project one* and violated for *project two*. Looking at *project one*, the results were significant for the non-gamified expectation management condition with $W(90) = .972$, $p = .050$ and the gamified expectation management condition with $W(51) = .819$, $p < .001$ and not significant for the control condition with $W(72) = .926$, $p = .400$. Regarding *project two*, the results were significant for all conditions; with $W(90) = .926$, $p < .001$ for the non-gamified expectation management condition, with $W(51) = .808$, $p < .001$ for the gamified expectation management condition and with $W(72) = .964$, $p = .036$ for the control condition. When inspecting the histograms, they did not show normality or equal shape. As ANOVA is quite robust to violations of normality for large group sizes as in this study normality could still be assumed. However, Levene's F test showed that the assumption for homogeneity of variances was not satisfied with $F(2,210) = 8.414$, $p < .001$ for *project one*, and with $F(2,210) = 8.593$, $p < .001$ for *project two*. In addition, the homogeneity of the variance-covariance matrices assumption using Box's M test was violated with $p < .001$. To investigate possible outliers in the data, the z-scores of the collaboration behaviour for *project one* and *project two* were computed. Seven z-scores were outside the range of -3 and +3 and were therefore considered outliers. Since the homogeneity of variances assumption and the homogeneity of the variance-covariance matrices assumption were violated and transformation did not yield stabilised homogeneity as well as considering the outliers that were considered plausible, non-parametric alternatives were conducted. In particular, Kruskal-Wallis H tests and Wilcoxon signed-rank tests were performed to test possible differences between the conditions, over time and their change over time.

Results

For analysing the collaboration outcome of the university students, the individual collaboration product, the group collaboration product and the collaboration behaviour were analysed in detail. An overview of the performed analyses can be found in Table 5.

Table 5

Overview of performed analyses

Outcome variables	Main analyses	Further analyses	
	Difference between condition Difference over time Difference in change over time	Difference between conditions per project	Difference over time per condition
Individual collaboration product	One Mixed ANOVA	Two One-Way ANOVAs	Three Paired samples t-tests
Group collaboration product	One Mixed ANOVA	Two One-Way ANOVAs	Three Paired samples t-tests
Collaboration behaviour	Two Kruskal Wallis H tests One Wilcoxon signed-rank test	Two Kruskal-Wallis H tests	Three Wilcoxon signed-rank tests

Effects on the collaboration products

Individual collaboration product

To examine whether students differed in their individual collaboration product between the three conditions (non-gamified expectation management condition, gamified expectation management condition, and control condition), over time (*project one* and *project two*) and in their change over time, a Mixed ANOVA was conducted as main analysis. In addition, further analyses with ANOVAs and Paired samples t-tests were conducted to zoom in on the difference between conditions per project and the difference over time per condition. The descriptive statistics are shown in Table 6.

Table 6

Descriptive statistics of the individual collaboration product

	Non-gamified expectation management condition (N= 83)	**Gamified expectation management condition (N= 50)**	**Control condition (N= 105)**	**Total (N=238)**
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Individual collaboration product *project one*	6.81 (1.23)	7.39 (1.29)	6.25 (1.25)	6.82 (1.31)
Individual collaboration product *project two*	6.58 (1.43)	7.38 (1.54)	6.50 (1.38)	6.82 (1.50)
Total individual collaboration product	6.69 (1.09)	7.39 (1.09)	6.37 (1.09)	

Note. N = Number of students

Main analysis (Mixed ANOVA)

A Mixed ANOVA was performed with the conditions as between-subject factor and the time with *individual collaboration product project one* and *individual collaboration product project two* as within-subject factor.

Difference between conditions. The results of the Mixed ANOVA revealed that there was a significant main effect of the condition on the individual collaboration product with $F(2, 235) = 14.563$, $p < .001$, $\eta_p^2 = .110$. This indicates a medium to large effect. The post hoc pairwise comparison using Bonferroni correction showed that there was a significant difference in the individual collaboration products between the gamified expectation management condition and the non-gamified expectation management condition ($p = .001$) as well as between the gamified expectation management condition and the control condition ($p < .001$). The difference between the non-gamified expectation management condition and control condition was not significant ($p = .143$). This indicates that ignoring the different time points, the gamified expectation management condition had better individual collaboration products compared to the non-gamified expectation management condition and to the control condition. And that the non-gamified expectation management condition performed similarly to the control condition in terms of the individual collaboration product when ignoring the different points of measurement.

Difference over time (between project one and project two). In contrast, there was no significant main effect of time on students' individual collaboration product with $F(1, 235) = .000$, $p = .990$, $\eta_p^2 < .001$. This indicates that ignoring the different conditions the participating students performed similarly in terms of their individual collaboration product for *project one* and *project two*.

Difference in change over time. In addition, there was no significant interaction between time and condition in terms of students' individual collaboration product with $F(2, 235) = 2.099$, $p = .125$, $\eta_p^2 = .018$. This indicates that changes in the individual collaboration product over time were not significantly different between the conditions. Figure 3 shows the profile plot.

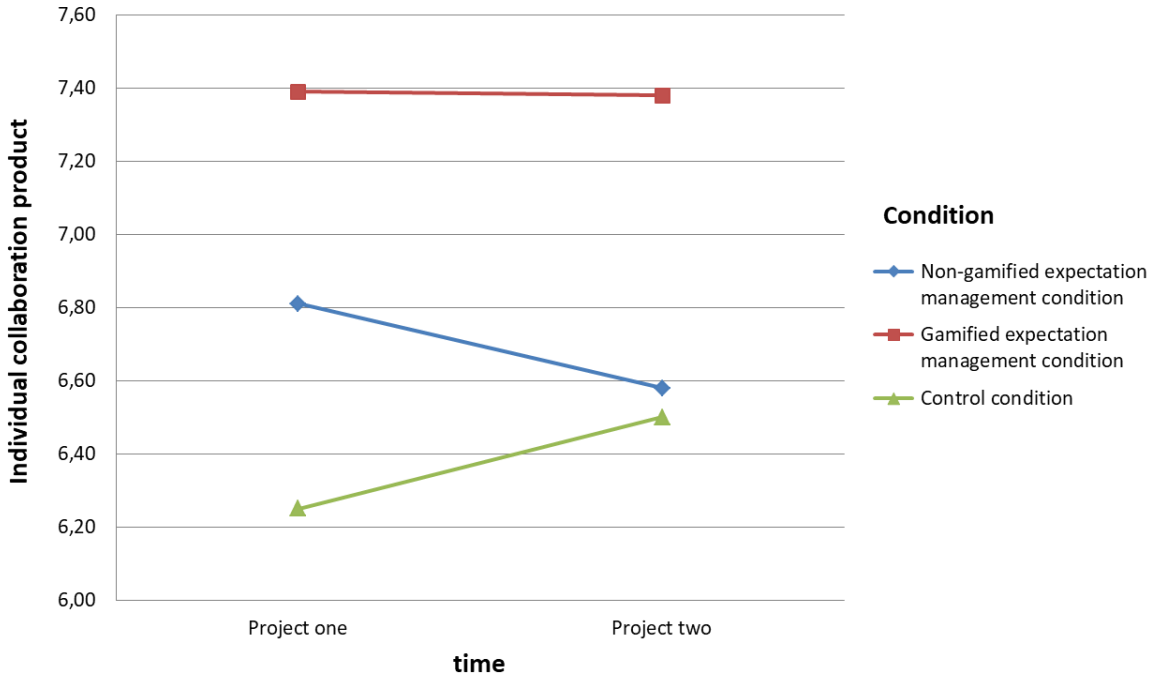


Figure 3. Profile plot individual collaboration product.

Further analyses (One-way ANOVAs, Paired samples t-tests)

Difference between conditions per project. To investigate the effect of the different conditions on the individual collaboration product further, two One-way ANOVAs were conducted; one for each project. For *project one*, a One-way ANOVA (with the *individual collaboration product project one* as dependent variable) revealed that there was a significant difference between the conditions, $F(2, 235) = 14.601$, $p < .001$, $\eta^2 = .111$, indicating a medium to large effect. Post hoc comparisons using Tukey HSD test showed that there was a significant difference between the gamified expectation management condition and non-gamified expectation management condition ($p = .027$), between the gamified expectation management condition and the control condition ($p < .001$) as well as between the non-gamified expectation management condition and control condition

(p = .008). This indicates that both the gamified expectation management condition and the non-gamified expectation management condition had better individual collaboration products in *project one* compared to the control condition. And that the gamified expectation management condition also performed better than the non-gamified expectation management condition in terms of the individual collaboration products in *project one*.

Looking at *project two*, a One-way ANOVA (with the *individual collaboration product project two* as dependent variable) revealed a significant difference between the conditions, $F(2, 235) = 7.006$, p = .001, η^2= .056, indicating a small to medium effect. Post hoc comparisons using Tuckey HSD test showed that there was a significant difference between the gamified expectation management condition and non-gamified expectation management condition (p = .006) as well as between the gamified expectation management condition and the control condition (p < .001). The difference between the non-gamified expectation management condition and control condition was not significant (p = .918). This indicates that the gamified expectation management condition had better individual collaboration products in *project two* compared to the non-gamified expectation management condition and to the control condition. And that there was no difference in the individual collaboration products between the non-gamified expectation management condition and control condition.

Difference over time (between project one and project two) per condition. To investigate an effect of time on the individual collaboration product further, three Paired samples t-tests with the *individual collaboration product project one* and the *individual collaboration product project two* as test pair were conducted; one for each condition. Paired samples t-tests for all three conditions revealed no significant difference between *project one* and *project two*, with $t(82) = 1.403$, p = .164, d = .17 for the non-gamified expectation management condition, with $t(49) = .044$, p = .965, d = .01 for the gamified expectation management condition, and with $t(104) = -1.527$, p = .130, d = -.18 for the control condition. This indicates that the students of all three conditions performed similarly in *project one* compared to *project two*.

To sum up, the individual collaboration product significantly differed between the three conditions overall and in *project one* and *project two.* In particular, the differences overall could hold when zooming in on *project two* but were different for *project one*. Meaning that overall and in *project one* and *project two* students in the gamified expectation management condition performed better compared to students in the non-gamified expectation management condition and the control condition. Furthermore, overall and in *project two* no difference could be found between students in the non-gamified expectation management condition and the control condition. However, in *project one,* students in the non-gamified expectation management condition performed better than

students in the control condition. When looking at students' development over time the individual collaboration product did not significantly differ between project one and project two, when ignoring the different conditions and when looking at the three different conditions separately. In addition, changes in the individual collaboration product over time were not significantly different when students were in the non-gamified expectation management condition, the gamified expectation management condition or the control condition.

Group collaboration product

To examine whether student groups differed in their group collaboration product between the three conditions (non-gamified expectation management condition, gamified expectation management condition, and control condition), over time (*project one* and *project two*) and their change over time, a Mixed ANOVA was conducted as main analysis. Moreover, further analyses with ANOVAs and Paired samples t-tests were conducted to zoom in on the difference between conditions per project and the difference over time per condition. The descriptive statistics are shown in Table 7.

Table 7

Descriptive statistics of the group collaboration product

	Non-gamified expectation management condition (N= 18)	**Gamified expectation management condition (N= 10)**	**Control condition (N= 22)**	**Total (N= 50)**
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Group collaboration product *project one*	6.92 (.59)	7.77 (.98)	7.05 (.51)	7.25 (.69)
Group collaboration product *project two*	7.09 (.63)	7.90 (.69)	7.22 (.50)	7.40 (.62)
Total group collaboration product	7.00 (.56)	7.84 (.56)	7.14 (.56)	

Note. N = Number of student groups

Main analysis (Mixed ANOVA)

A Mixed ANOVA was performed with the conditions as between-subject factor and the time with *group collaboration product project one* and *group collaboration product project two* as within-subject factor.

Difference between conditions. The results of the Mixed ANOVA revealed that there was a significant main effect of the condition on the group collaboration product with $F(2, 47) = 4.780$, $p = .001$, $\eta_p^2 = .246$. This indicates a large effect. The post hoc pairwise comparison using Bonferroni correction showed that there was a significant difference in the group collaboration products between the gamified expectation management condition and the non-gamified expectation management condition ($p = .001$) as well as between the gamified expectation management condition and the control condition ($p = .006$). The difference between the non-gamified expectation management condition and control condition was not significant ($p = 1.00$). This indicates that ignoring the different time points, the gamified expectation management condition had better group collaboration products compared to the non-gamified expectation management condition and to the control condition. And that the non-gamified expectation management condition performed similarly to the control condition in terms of the group collaboration product when ignoring the different points of measurement.

Difference over time (between project one and project two). In contrast, there was no significant main effect of time on students' group collaboration product with $F(1, 57) = 3.614$, $p = .063$, $\eta_p^2 = .071$. This indicates that ignoring the different conditions the participating groups performed similarly in terms of their group collaboration product for *project one* and *project two*.

Difference in change over time. In addition, there was no significant interaction between time and condition in terms of students' group collaboration product with $F(2, 47) = .020$, $p = .980$, $\eta_p^2 = .001$. This indicates that changes in the group collaboration product over time were not significantly different between the conditions. Figure 4 shows the profile plot.

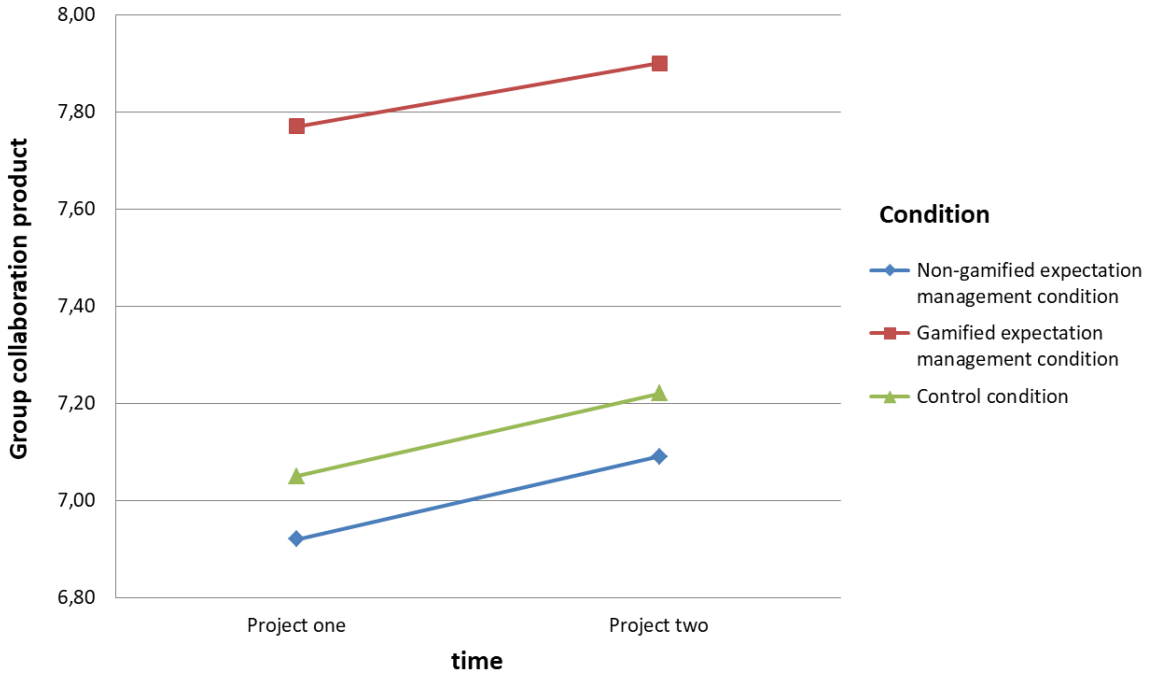


Figure 4. Profile plot group collaboration product.

Further analyses (One-way ANOVAs, Paired samples t-tests)

Difference between conditions per project. To investigate the effect of the different conditions on the group collaboration product further, two One-way ANOVAs were conducted; one for each project. For *project one*, a One-way ANOVA (with the *group collaboration product project one* as dependent variable) revealed that there was a significant difference between the conditions, $F(2, 47) = 5.900$, $p = .005$, $\eta^2 = .201$, indicating a large effect. Post hoc comparisons using Tuckey HSD test showed that there was a significant difference between the gamified expectation management condition and non-gamified expectation management condition ($p = .005$) as well as between the gamified expectation management condition and the control condition ($p = .016$). The difference between the non-gamified expectation management condition and control condition was not significant ($p = .786$). This indicates that the gamified expectation management condition had better group collaboration products in *project one* compared to the non-gamified expectation management condition and to the control condition. And that there was no difference in the group collaboration products between the non-gamified expectation management condition and control condition.

Looking at *project two*, a One-way ANOVA (with the *group collaboration product project two* as dependent variable) revealed a significant difference between the conditions, $F(2, 47) = 2.270$, $p = .003$, $\eta^2 = .208$, indicating a large effect. Post hoc comparisons using Tuckey HSD test showed that there was a significant difference between the gamified expectation management condition and non-gamified expectation management condition ($p = .003$) as well as between the gamified

expectation management condition and the control condition (p = .011). The difference between the non-gamified expectation management condition and control condition was not significant (p = .766). This indicates that the gamified expectation management condition had better group collaboration products in *project two* compared to the non-gamified expectation management condition and to the control condition. And that there was no difference in the group collaboration products between the non-gamified expectation management condition and the control condition.

Difference over time (between project one and project two) per condition. To investigate an effect of time on the group collaboration product further, three Paired samples t-tests with the *group collaboration product project one* and the *group collaboration product project two* as test pair were conducted; one for each condition. Paired samples t-tests for all three conditions revealed no significant difference between *project one* and *project two*, with $t(17) = -1.285$, $p = .216$, $d = -.28$ for the non-gamified expectation management condition, with $t(9) = -.707$, $p = .497$, $d = -.15$ for the gamified expectation management condition, and with $t(21) = -1.501$, $p = .148$, $d = -.34$ for the control condition. This indicates that the groups in all three conditions performed similarly in *project one* compared to *project two*.

To conclude, same as for the individual collaboration product, the group collaboration product significantly differed between the three conditions. However, for the group collaboration product, the differences overall could hold for *project one* and *project two*. Meaning that overall and in *project one* and *project two* groups in the gamified expectation management condition performed better compared to groups in the non-gamified expectation management condition and to groups in the control condition. Furthermore, no difference could be found between groups in the non-gamified expectation management condition and the control condition. When looking at the development over time, same as for the individual collaboration product, groups performed similarly in *project two* compared to *project one*, when ignoring the different conditions and when looking at the three different conditions separately. In addition, changes in the group collaboration product over time were not significantly different when groups were in the non-gamified expectation management condition, the gamified expectation management condition or the control condition.

Effects on the collaboration behaviour

To investigate whether students differed in their collaboration behaviour between the three conditions (non-gamified expectation management condition, gamified expectation management condition, and control condition) a Kruskal-Wallis H test was conducted. To examine possible differences over time (*project one* and *project two*) a Wilcoxon signed-rank test was performed as

well as a Kruskal-Wallis H test with the change of collaboration behaviour between *project one* and *project two* as a dependent variable to investigate a change over time between the conditions. In addition, further analyses with Kruskal-Wallis H tests and Wilcoxon signed-rank tests were conducted to zoom in on the difference between conditions per project and the difference over time per condition. The descriptive statistics are shown in Table 8.

Table 8

Descriptive statistics of collaboration behaviour

	Non-gamified expectation management condition (N = 90)	**Gamified expectation management condition (N = 51)**	**Control condition (N = 72)**	**Total (N = 213)**
	Mdn (IQR)	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*
Self- and peer-assessed collaboration behaviour *project one*	8.90 (.82)	9.00 (.47)	8.80 (.94)	8.90 (.70)
Self- and peer-assessed collaboration behaviour *project two*	8.98 (.55)	8.92 (.17)	8.84 (.89)	8.92 (.60)
Total self- and peer-assessed collaboration behaviour (average *project one* and *project two*)	8.92 (.73)	8.95 (.35)	8.79 (1.02)	
Change in self- and peer-assessed collaboration behaviour (*project two* - *project one*)	.05 (.40)	.00 (.37)	.04 (.27)	

Note. N = Number of students

Main analyses (Kruskal-Wallis H tests, Wilcoxon signed-rank test)

Difference between conditions. To examine possible differences between the conditions a Kruskal-Wallis H test was conducted with *total self- and peer-assessed collaboration behaviour* as the dependent variable. The Kruskal-Wallis H test revealed no significant difference in the collaboration behaviour between the three conditions, $H(2) = .681$, $p = .711$, $\eta^2 = .003$, with a mean rank score of 109.12 for the non-gamified expectation management condition, 110.10 for the gamified expectation management condition, and 102.15 for the control condition. This indicates that the three conditions did not differ in their collaboration behaviour when ignoring the different points of measurement.

Difference over time (between project one and project two). To investigate possible differences over time, a Wilcoxon signed-rank test was conducted with the *self- and peer-assessed collaboration behaviour project one* and *self- and peer-assessed collaboration behaviour project two* as test pair. The Wilcoxon signed-rank test showed that the collaboration behaviour did significantly differ between *project one* and *project two* with $Z = -2.573$, $p = .010$, $r = -.18$. This is considered a small effect and indicates that ignoring the different conditions the participating students performed better in terms of their collaboration behaviour in *project two* than *project one*.

Difference in change over time. A Kruskal Wallis H test with the *change in self- and peer-assessed collaboration behaviour* between *project one* and *project two* as a dependent variable revealed no significant difference between the conditions, $H(2) = 1.615$, $p = .446$, $\eta^2 = .008$, with a mean rank score of 112.63 for the non-gamified expectation management condition, 99.20 for the gamified expectation management condition, and 105.49 for the control condition. This indicates that changes in the collaboration behaviour over time were not significantly different between the conditions.

Further analyses (Kruskal-Wallis H tests, Wilcoxon signed-rank tests)

Difference between conditions per project. To investigate possible differences between the conditions on the collaboration behaviour further, two Kruskal-Wallis H tests were conducted; one for each project. For *project one*, a non-parametric Kruskal-Wallis H test (with the *self- and peer-assessed collaboration behaviour project one* as dependent variable) revealed no significant difference in the collaboration behaviour between the three conditions, $H(2) = .737$, $p = .692$, $\eta^2 = .003$, with a mean rank score of 107.08 for the non-gamified expectation management condition, 112.61 for the gamified expectation management condition, and 102.93 for the control condition. This indicates that the three conditions did not differ in their collaboration behaviour in *project one*.

Looking at *project two*, a non-parametric Kruskal-Wallis H test (with the *self- and peer-assessed collaboration behaviour project two* as dependent variable) showed no significant differences in the collaboration behaviour between the three conditions, $H(2) = 1.167$, $p = .558$, $\eta^2 =$

.006, with a mean rank score of 111.86 for the non-gamified expectation management condition, 106.37 for the gamified expectation management condition, and 101.37 for the control condition. This indicates that the three conditions did not differ in their collaboration behaviour also in *project two*.

Difference over time (between project one and project two) per condition. To investigate possible differences over time on the collaboration behaviour further, three Wilcoxon signed-rank tests with *self- and peer-assessed collaboration behaviour project one* and *self- and peer-assessed collaboration behaviour project two* as test pairs were conducted; one for each condition. For the non-gamified expectation management condition, a Wilcoxon signed-rank test revealed a significant difference between the collaboration behaviour of *project one* and *project two* with Z = -2.279, p = .023, r = -.24. This is considered a small effect and indicates that students in this condition performed better in their collaboration behaviour in *project two* compared to *project one*. However, the Wilcoxon signed-rank tests showed no difference in the collaboration behaviour between *project one* and *project two* for the gamified expectation management condition with Z= -.452, p = .652, r = -.06, and the control condition with Z= -1.391, p = .164, r = -.16. This indicates that students in these two conditions performed similarly in their collaboration behaviour in *project two* and *project one*.

To sum up, the self-and peer-assessed collaboration behaviour did not significantly differ between the three conditions overall and in *project one* and *project two*. When looking at students' development over time students of the non-gamified expectation management condition performed better in *project two* compared to *project one*. However, in the gamified expectation management condition and the control condition students performed similarly in *project two* compared to *project one*. Furthermore, changes in the collaboration behaviour over time were not significantly different when students were in the non-gamified expectation management condition, the gamified expectation management condition or the control condition.

Extra – Difference between self- and peer-assessments

To examine whether students differed in their collaboration behaviour between the three conditions, over time and in their change over time, the average scores of students' self-and peer-assessments were used. However, previous research suggests that students might differ in how they evaluate themselves and how others evaluate them (Dunning, Heath, & Suls, 2004), and thus the results might differ between the self- and peer-assessments. To substantiate the discussion, these student self-and peer-ratings are therefore now compared.

To investigate these possible differences between the self- and peer assessment of the collaboration behaviour, a Paired samples t-test was performed with *total self-assessed collaboration behaviour (average project one and project two)* and *total peer-assessed collaboration behaviour (average project one and project two)* as test pair. The Paired samples t-test revealed a significant difference between self- and peer assessments with $t(212) = -3.443$, $p = .001$, $d = -.27$. This is considered a small effect and indicates that the peer-ratings of the collaboration behaviour were higher ($M = 8.91$, $SD = .58$) than the self-ratings ($M = 8.71$, $SD = .86$). In addition, self- and peer assessment scores showed a medium correlation ($r = .361$, $p < .001$).

Since the test showed a difference between self- and peer-ratings with a medium correlation, the Kruskal-Wallis H tests and Wilcoxon signed-rank tests (to test for differences between conditions, over time and change over time) are run again for the collaboration behaviour, but this time separately for the self- and the peer-ratings. This is done to check if the results for the collaboration behaviour are different when examined separately. The descriptive statistics are shown in Table 9.

Table 9

Descriptive statistics of self- and peer-assessed collaboration behaviour

	Non-gamified expectation management condition (N = 90)	**Gamified expectation management condition (N = 51)**	**Control condition (N = 72)**	**Total (N = 213)**
	Mdn (IQR)	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*
Self-assessed collaboration behaviour *project one*	8.75 (1.75)	8.75 (1.25)	8.75 (1.50)	8.75 (1.50)
Self-assessed collaboration behaviour *project two*	8.88 (1.50)	8.75 (1.00)	9.00 (.57)	8.75 (1.25)
Total self-assessed collaboration behaviour (average *project one* and *project two*)	8.88 (1.28)	8.75 (.88)	8.88 (1.50)	
Change in self-assessed collaboration behaviour (*project two – project one*)	.00 (.75)	.00 (.75)	.00 (.75)	
Peer-assessed collaboration behaviour *project one*	8.84 (.87)	9.05 (.45)	8.89 (.94)	8.94 (.75)
Peer-assessed collaboration behaviour *project two*	9.00 (.57)	9.00 (.30)	8.84 (.92)	8.95 (.59)
Total peer-assessed collaboration behaviour (average *project one* and *project two*)	8.94 (.68)	9.05 (.32)	8.83 (.94)	
Change in peer-assessed collaboration behaviour (*project two – project one*)	.07 (.42)	-.05 (34)	0.00 (.42)	

Note. N = Number of students

Self-assessment

Main analyses:

Difference between conditions. To examine possible differences between the conditions a Kruskal-Wallis H test was conducted with the *total self-assessed collaboration behaviour* as the dependent variable. The Kruskal-Wallis H test revealed no significant difference in the self-assessed collaboration behaviour between the three conditions, $H(2) = .300$, $p = .861$, $\eta^2 = .001$, with a mean rank score of 109.65 for the non-gamified expectation management condition, 104.39 for the gamified expectation management condition, and 105.53 for the control condition. This indicates that the three conditions did not differ in their self-assessed collaboration behaviour when ignoring the different points of measurement.

Difference over time (between project one and project two). To investigate possible differences over time, a Wilcoxon signed-rank test was conducted with the *self-assessed collaboration behaviour project one* and *self-assessed collaboration behaviour project two* as test pair. The Wilcoxon signed-rank test showed that the self-assessed collaboration behaviour did not significantly differ between *project one* and *project two* with $Z = -.722$, $p = .470$, $r = -.05$. This indicates that ignoring the different conditions the participating students performed according to themselves similarly in terms of their collaboration behaviour in *project one* and *project two*.

Difference in change over time. A Kruskal Wallis H test with the *change of self-assessed collaboration behaviour* between *project one* and *project two* as a dependent variable revealed no significant difference between the conditions, $H(2) = .568$, $p = .753$, $\eta^2 = .003$, with a mean rank score of 110.62 for the non-gamified expectation management condition, 105.20 for the gamified expectation management condition, and 103.76 for the control condition. This indicates that changes in the self-assessed collaboration behaviour over time were not significantly different between the conditions.

Further analyses:

Difference between conditions per project. To investigate possible differences between the conditions on the self-assessed collaboration behaviour further, two Kruskal-Wallis H tests were conducted; one for each project. For *project one,* a non-parametric Kruskal-Wallis H test (with the *self-assessed collaboration behaviour project one* as dependent variable) revealed no significant difference in the self-assessed collaboration behaviour between the three conditions, $H(2) = .115$, $p = .944$, $\eta^2 = .001$, with a mean rank score of 108.66 for the non-gamified expectation management condition, 105.90 for the gamified expectation management condition, and 105.70 for the control condition. This indicates that the three conditions did not differ in their self-assessed collaboration behaviour in *project one*.

Looking at *project two,* a non-parametric Kruskal-Wallis H test (with the *self-assessed collaboration behaviour project two* as dependent variable) showed no significant differences in the self-assessed collaboration behaviour between the three conditions, $H(2) = .710$, $p = .701$, $\eta^2 = .001$, with a mean rank score of 111.01 for the non-gamified expectation management condition, 102.77 for the gamified expectation management condition, and 104.98 for the control condition. This indicates that the three conditions did not differ in their self-assessed collaboration behaviour also in *project two*.

Difference over time (between project one and project two) per condition. To investigate possible differences over time on the self-assessed collaboration behaviour further, three Wilcoxon signed-rank tests with the *self-assessed collaboration behaviour project one* and *self-assessed collaboration behaviour project two* as test pairs were conducted; one for each condition. The Wilcoxon signed-rank tests showed that the self-assessed collaboration behaviour did not differ between *project one* and *project two* for all three conditions, with $Z = -.689$, $p = .491$, $r = -.07$ for the non-gamified expectation management condition, with $Z = -.302$, $p = .763$, $r = -.04$ for the gamified expectation management condition, and with $Z = -.196$, $p = .845$, $r = -.02$ for the control condition. This indicates that students in all conditions performed according to themselves similarly in terms of their collaboration behaviour in *project one* and *project two*.

Peer-assessment

Main analyses:

Difference between conditions. To examine possible differences between the conditions a Kruskal-Wallis H test was conducted with *total peer-assessed collaboration behaviour* as the dependent variable. The Kruskal-Wallis H test revealed no significant difference in the peer-assessed collaboration behaviour between the three conditions, $H(2) = 1.156$, $p = .561$, $\eta^2 = .005$, with a mean rank score of 107.71 for the non-gamified expectation management condition, 113.52 for the gamified expectation management condition, and 101.50 for the control condition. This indicates that the three conditions did not differ in their peer-assessed collaboration behaviour when ignoring the different points of measurement.

Difference over time (between project one and project two). To investigate possible differences over time, a Wilcoxon signed-rank test was conducted with the *peer-assessed collaboration behaviour project one* and *peer-assessed collaboration behaviour project two* as test pair. The Wilcoxon signed-rank test showed that the peer-assessed collaboration behaviour did significantly differ between *project one* and *project two* with $Z = -2.880$, $p = .004$, $r = -.20$. This is considered a small effect and indicates that ignoring the different conditions the participating

students performed according to their peers better in terms of their collaboration behaviour in *project two* than *project one*.

Difference in change over time. A Kruskal Wallis H test with the *change of peer-assessed collaboration behaviour* between *project one* and *project two* as a dependent variable revealed no significant difference between the conditions, $H(2) = 4.923$, $p = .085$, $\eta^2 = .023$, with a mean rank score of 116.29 for the non-gamified expectation management condition, 92.49 for the gamified expectation management condition, and 105.66 for the control condition. This indicates that changes in the peer-assessed collaboration behaviour over time were not significantly different between the conditions.

Further analyses:

Difference between conditions per project. To investigate possible differences between the conditions on the peer-assessed collaboration behaviour further, two Kruskal-Wallis H tests were conducted; one for each project. For *project one,* a non-parametric Kruskal-Wallis H test (with the *peer-assessed collaboration behaviour project one* as dependent variable) revealed no significant difference in the peer-assessed collaboration behaviour between the three conditions, $H(2) = 1.653$, $p = .438$, $\eta^2 = .008$, with a mean rank score of 104.61 for the non-gamified expectation management condition, 116.61 for the gamified expectation management condition, and 103.18 for the control condition. This indicates that the three conditions did not differ in their peer-assessed collaboration behaviour in *project one*.

Looking at *project two,* a non-parametric Kruskal-Wallis H test (with the *peer-assessed collaboration behaviour project two* as dependent variable) showed no significant differences in the peer-assessed collaboration behaviour between the three conditions, $H(2) = 1.719$, $p = .423$, $\eta^2 = .008$, with a mean rank score of 112.48 for the non-gamified expectation management condition, 107.58 for the gamified expectation management condition, and 99.74 for the control condition. This indicates that the three conditions did not differ in their peer-assessed collaboration behaviour also in *project two*.

Difference over time (between project one and project two) per condition. To investigate possible differences over time on the peer-assessed collaboration behaviour further, three Wilcoxon signed-rank tests with the *peer-assessed collaboration behaviour project one* and *peer-assessed collaboration behaviour project two* as test pairs were conducted; one for each condition. A Wilcoxon signed-rank test revealed a significant difference between the peer-assessed collaboration behaviour of *project one* and *project two* for the non-gamified expectation management condition with $Z = -2.884$, $p = .004$, $r = -.30$. This is considered a small to medium effect and indicates that students in this condition performed according to their peers better in terms of their collaboration behaviour in

project two compared to *project one*. However, the Wilcoxon signed-rank tests showed that the peer-assessed collaboration behaviour did not differ between *project one* and *project two* for the gamified expectation management condition with Z = -.017, p = .986, r < -.01, and for the control condition with Z = -1.453, p = .146, r = -.17. This indicates that students in these two conditions performed according to their peers similar in terms of their collaboration behaviour in *project one* and *project two*.

In conclusion, the results of the peer-assessed collaboration behaviour were consistent with those of the self- and peer-assessed collaboration behaviour, but the results of the self-assessed collaboration behaviour could not hold for the difference over time (between *project one* and *project two*). Meaning that in all three cases the collaboration behaviour did not significantly differ between the three conditions overall and in *project one* and *project two*. Furthermore, in all three cases, changes in the collaboration behaviour over time did not differ significantly when students were in the non-gamified expectation management condition, the gamified expectation management condition or the control condition. In addition, when looking at students' development over time students in the gamified expectation management condition and the control condition performed similarly in *project two* compared to *project one* in all three cases. However, in the non-gamified expectation management condition students performed better in *project two* compared to *project one* when looking at the self- and peer-assessed and the peer-assessed collaboration behaviour but they performed similarly when looking at the self-assessed collaboration behaviour.

Discussion and conclusion

The unique contribution of this study is the combination of instructional support, expectation management and gamification to investigate to what extent expectation management and gamification influence university students' collaboration outcome. In this study, expectation management and a combination of expectation management and gamification were integrated into online instructional support in collaboration and compared to a control condition (without expectation management and gamification). Expectation management support was designed to help students in becoming aware of their own and group members' traits and preferences when working together to find a common ground within the group. Gamification was added to create a safe and engaging setting for the students. It was expected that students in the gamified expectation management condition have the best collaboration outcomes consisting of products (H1a) and behaviour (H2a), followed by the non-gamified expectation management condition, and the control condition in the last place. As the data was collected at two moments in time with the same student groups experiencing repeated collaborative situations, it was also hypothesised that students in all three conditions improve in their collaboration outcome consisting of products (H1b) and behaviour (H2b) over time, but more importantly that students in the gamified expectation management condition improve more than students in non-gamified expectation management condition and the control condition. This was expected due to the design of the gamification in the collaboration support of this study. To be more precise the badges, as the second game design element, were only introduced to the students after *project one* was completed and thus a boost was expected during *project two*.

Research findings

Effects on the collaboration products

Difference between conditions

Looking at the effects of the individual and the group collaboration products, the results suggest that the combination of expectation management and gamification have positively influenced students' collaboration products, as the gamified expectation management condition performed significantly better than the other two conditions. However, contrary to what was hypothesised, the results did not indicate a unique effect of expectation management alone, as there was no difference between the non-gamified expectation management condition and the control condition on students' collaboration products in most analysed cases. Only when zooming in on the individual collaboration product in *project one* this difference was significant. However, based on the

other results of the current study, it is questionable whether this difference in students' individual collaboration products between the non-gamified expectation management conditions and the control condition in *project one* is actually an expected effect. An indication for this is the fact that the difference could not be found in the overall analysis and is not evident when looking at the development of the three conditions over time.

The general finding that in most cases no difference was found between students in the non-gamified expectation management condition and the control condition is not in line with our hypothesis and with previous research. The literature suggests that students can potentially benefit more from collaboration if they know what to expect from each other in certain situations (Bird & Osland, 2005; Cagiltay et al., 2015; Popov, 2013). However, from the current study, it might be concluded that managing students' expectations, with the goal of building a shared understanding and subsequently influencing the collaboration, only seems to be effective in combination with gamification. A possible explanation why, in the current study, expectation management alone did not influence students' collaboration products could be that students in the non-gamified expectation management condition did not interact with the support on the right level where they could actually benefit from the added value of expectation management to show a difference. A lack of student engagement in instructional support is a well-known problem (e.g. Challco et al., 2016; Lajoie, 2005) that can negatively impact behaviour leading to decreased collaboration outcomes in the long run (Challco et al., 2016) and a reason why gamification is added to overcome this difficulty (Dicheva & Dichev, 2015; Giannetto et al., 2013). A possible reason why students might not engage in this specific instructional support could be that students had to share information about their traits and preferences when working together. This might feel personal especially at the beginning of a collaboration where the students only start to get to know each other. Previous research has shown that avatars could help in creating a safe setting to communicate personal information (Falloon, 2010; Pickard et al., 2016; Sillaots, 2014) through the feeling of self-presence in the online environment (Hooi & Cho, 2014). The avatars enhanced self-disclosure of traits and preferences related to collaboration might have helped the students in finding a common ground within the group, improving the collaboration and subsequently leading to the increased collaboration products. Thus, the overall effect could be seen for the gamified expectation management condition, but not for the expectation management condition (and the control condition).

Overall, the results of both the individual and the group collaboration products suggest that adding expectation management in combination with gamification in instructional support can benefit students in their collaboration, as the gamified expectation management condition could yield better results in terms of the collaboration products compared to the other two conditions. However, it needs to be mentioned that the number of groups in the gamified expectation

management condition was quite small for analysing the effects on the group collaboration product. Therefore, the results of the group collaboration product need to be interpreted with caution, but considering the high congruence with the results of the individual collaboration product, the findings can give an indication of the effect.

Difference (in change) over time

Concerning the effects of time, contrary to the expectations, the data indicates that students did not increase in terms of their collaboration products over time, as all conditions performed similarly in *project one* compared to *project two*. Moreover, the students in the gamified expectation management condition did not improve more than students in the non-gamified expectation management condition and the control condition, as changes in the collaboration products over time were not significantly different between the three conditions.

Regarding the unexpected finding that students in the gamified expectation management condition did not improve more than the students in the other two conditions, it seems that the badges did not give a boost to the gamification effect for *project two*, which was expected since the badges were only introduced to the students after finishing and submitting the collaboration products for *project one.* The reason why the badges might have been not effective for a boost could be that too much was added to one task and thus too many cognitive resources were demanded from the students. To illustrate, an additional element, namely the badges, was added to the RIDE assessment, which is already a highly cognitive task as it requires students to reflect on their collaboration. Reflection on the collaboration involves becoming aware of how one performed, comparing it to benchmarks and subsequently drawing conclusions by identifying possible gaps and setting goals to close those gaps, which is also reflected in the RIDE assessment tool used in this study (Eshuis et al., 2019). These steps require a lot of cognitive resources, especially when students have little experience in reflecting on their collaboration (Kalyuga, Ayres, Chandler, & Sweller, 2003). Furthermore, Turan, Avinc, Kara, and Goktas (2016) demonstrate that additional cognitive resources are needed when using gamification, as game design elements such as badges are added to a task, which increases the cognitive load of students. As a result, the RIDE assessment tool in combination with the badges might have caused a cognitive overload for the students. Cognitive overload means that the task requires more cognitive resources than the students have available (American Psychological Association, 2020), resulting in reduced processing and performance in the task (Adcock, 2000; Kalyuga et al., 2003). Under this condition, students in the gamified expectation management condition might have not benefited as much from the task (the badges assignment and the RIDE assessment tool) in *project one* as expected and thus might not have strengthened the effect for *project two*.

A possible explanation why students in all three conditions did not improve in their collaboration products over time (between the two projects) could be due to the difficulty of the learning content of the projects. The content of *project two* can be considered more complex than the one from *project one, as* creating something completely new (*project two*) requires more mental effort and resources than replicating findings (*project one*). Thus, it could be that the benefits of knowing each other working together repeatedly were offset by the greater complexity of the content. Because high-complex tasks demand more from the group compared to tasks with lower complexity, such as a higher level and quality of coordination and interaction (Hagemann & Kluge, 2017; Pe, Mothe, & Brion, 2011), which students might have lacked. Another possible explanation for the absent improvement is that one repetition of the collaborative situation was not sufficient to show increases in the collaboration, as previous research by Kollar et al. (2007) argues that it takes multiple repetitions over a longer time.

Effects on the collaboration behaviour

Difference between conditions

Looking at the effects on the collaboration behaviour, the results suggest, contrary to the expectations, that neither expectation management nor the combination of expectation management and gamification seems to have influenced students' collaboration behaviour, as all three conditions had similar scores when looking at the self-and peer assessment together as well as separately.

A possible explanation could be that students might not have evaluated each other truthfully. All scores of students' self-and peer evaluation were quite high. Usually, students overestimate themselves (Dunning et al., 2004) and therefore self-assessment is often combined with peer-assessment. However, in this study, students evaluated others higher than themselves in terms of their behaviour which might indicate that they did not evaluate each other truthfully and were hesitant to criticise others. This is consistent with previous research showing that students are reluctant and rather avoid giving critical feedback to their peers even when the peer-assessment is anonymous (Lin, Liu, & Yuan, 2001). From casual observations it was also apparent that many students gave the same rating to all their group members and that only lower scores were given for themselves and group members who exhibited behaviour of free-riding and social loafing, so contributing no or little to the collaboration. This suggests students only expressed criticism to other students when the behaviour was extreme. The reason why students might not have evaluated each other truthfully could be that students lacked the skills necessary to evaluate each other and that they need additional support in this process. This is also in line with previous literature

demonstrating that providing peer feedback is a complex skill and students need training to be able to do it reliably (Cho & Schunn, 2007; Hanrahan & Isaacs, 2001; Li, Xiong, Hunter, Guo, & Tywoniw, 2019). Cho and Schunn (2007) argue that this is especially the case for students, as they typically lack experience in conducting peer assessments. Since the sample in the study was first-year bachelor students, it can be expected that they do not have much experience in giving feedback to each other regarding their collaboration behaviour. Therefore, it seems that students might need training in how to assess each other and provide feedback to peers.

Another possible reason why students might not have evaluated each other truthfully could also be that students did not feel safe enough to do so. The current study intended to overcome this problem with the use of avatars, as previous research suggests that they can create a safe setting (Falloon, 2010; Sillaots, 2014) and this was also indicated by the results of the collaboration products in this study. However, the design of the avatar might have been not ideal to create the safe space also for the evaluation (not only for discussing preferences and expectations), as avatars were only indirectly linked to the RIDE assessment tool (through the badges). A study by Hooi and Cho (2014) demonstrate that avatars have positive effects on self-disclosure if people have the feeling that the avatars in the online environment are their actual selves, but it can also have negative effects and that is if people feel identifiable. As in this study the avatars were indirectly linked to the RIDE assessment tool and both were labelled with the students' first names, some might have felt more identifiable than self-present during the evaluation. Thus, the benefits of avatars for a safe setting to communicate about personal thoughts, i.e. the evaluation of the collaboration, might have been lost on those students. It is recommended by Hooi and Cho (2014) to integrate non-identifiability measures. One possible measure in this study could be to let students choose the names of their avatars themselves instead of pre-labelling it for them and integrating those in the RIDE assessment tools instead of their first names. This might help the students in feeling a bit less identifiable when doing the evaluation, which in turn would benefit self-disclosure and a safe setting. However, it needs to be mentioned that full non-identifiability or anonymity could be difficult in this context and would not have been possible in the current study as the students were working together on assignments that were part of their regular coursework at university and thus the students were identifiable by their group members.

Difference (in change) over time

Regarding the effects of time, the data suggest that students' self-and peer-assessed collaboration behaviour increased over time. However, contrary to our expectations, this could be fully attributed to the students in the non-gamified expectation management condition and specifically to their peer-assessment of the collaboration behaviour. Students in the gamified

expectation management and control condition demonstrated no significant differences between *project one* and *project two*. Furthermore, in contrast to what we expected, the students in the gamified expectation management condition did not improve more than students in the non-gamified expectation management condition and the control condition, as changes in the collaboration products over time were not significantly different between the three conditions.

This absent improvement in the gamified expectation management condition could be explained by the distraction that gamification might have caused among the students in this condition. In general, instructional support should be clean, clear and to the point. In this study, however, seductive content in form of the game design elements was added, which might have caused students to lose track of what was important. This is in line with previous research mentioning distraction as a drawback of gamification (e.g. Andrade, Mizoguchi, & Isotani, 2016; Hakulinen, 2015; Yanes & Bououd, 2019). According to Bender, Renkl, and Eitel (2021), these seductive elements have (due to the distraction) a negative impact on students' performance if they are not informed about the purpose of these elements, which was not the case in this study. From casual observations regarding the dropout reasons (which are discussed later) it became apparent that students in the current study were more focused on the badges assignment than on the RIDE assessment tool. So it might be that students put more importance on the badges task and as a consequence wanted to finish the RIDE assessment tool fast or in less detail. Thus, this might have caused the data to show no improvement from *project one* to *project two* for students in the gamified expectation management condition, but an improvement for students in the non-gamified expectation management condition (as they were not distracted by the badges assignment).

Since this effect was entirely attributed to the peer-assessment, an alternative explanation could be considered. Specifically, the data suggested that it was the peers in the non-gamified expectation management condition that gave their group members higher scores in *project two* compared to *project one*. A possible explanation for this is that expectation management (without distraction) might have helped the students in the group to get to know each other better and develop some kind of friendship over the duration of the collaboration. A closer bond might have made it more difficult for students to criticise their group members. This is also consistent with previous studies mentioning that students are more hesitant to criticise friends (Gamlem & Smith, 2013; Harris & Brown, 2013). Previous research by Panadero, Romero, and Strijbos (2013) also show that a higher level of friendship can lead to higher peer-assessment scores. Thus, it might be that in *project two* the group members (that received the expectation management only) have developed a closer bond and were therefore even more reluctant to evaluate each other truthfully compared to *project one*, and thus gave even higher scores.

However, it needs to be considered that this effect of time in the non-gamified expectation management condition was small and there was also no difference in change over time between the conditions, which does not support this effect. In addition, this small effect was fully attributable to the peer-assessment, for which it is already argued that students still need experience and training. Therefore, it is questionable whether this is an effect that is actually expected.

By questioning this effect and the lack of improvement in the other two conditions (gamified expectation management condition and control condition), another possible explanation could be mentioned. Namely, that one repetition of the collaborative situation might not have been sufficient to show increases in the collaboration behaviour. This is consistent with previous research arguing that it takes multiple repetitions over a longer period of time to show increases (Kollar et al., 2007).

Another observation of the study worth discussing is the somewhat high dropout rate in the study and its implications for the design of the online collaboration support. The highest dropouts were due to not completing all steps in the evaluation in Task 3 and Task 4. Most student groups started the tasks but did not finish all of the subordinate steps. In particular, the pattern emerged that in Task 3, student groups in the gamified expectation management condition mostly dropped out due to not performing all steps of the RIDE assessment tool. And in Task 4, dropouts grew primarily in the control condition by not completing the third step of the RIDE assessment tool. However, the dropout within those conditions (gamified expectation management condition and control condition) seemed to be random, meaning that not specifically lower scores or higher scores were removed from the data, so it is assumed that the dropout had no significant impact on the average collaboration behaviour scores and thus no impact on the results.

A possible explanation for the dropout rates in the gamified expectation management condition in Task 3 could be that students in this condition had difficulties deciding what was important, as the students were more focused on the badges assignment and paid less attention to the RIDE assessment tool. It might be that the gamification was too distracting and added complexity to the already highly cognitive task of the RIDE assessment, as already mentioned before. This is consistent with previous research demonstrating that game-based elements such as badges can increase cognitive load and hinder focus on relevant elements (Turan et al., 2016). Other studies also mention distraction as a possible drawback of gamification in education (e.g. Hakulinen, 2015; Yanes & Bououd, 2019) and show that some students felt distracted by badges (Hakulinen, 2015). Andrade et al. (2016) add that distraction can be especially the case when the gamification is detached from the task, in this study the RIDE assessment tool. Thus, a possible improvement to the design would be to better integrate the tasks, i.e., to design them in such a way that the RIDE assessment tool and the badge assignment are combined under one seamless task. Another improvement would also be

to make the RIDE assessment tool less complex by removing steps or enhancing the user experience of the tool. For example, the user experience could be improved by displaying a progress bar that shows what step students are in, what they have completed, and what is still ahead of them. This is considered particularly useful for this tool, as it was clear from casual observations that some student groups were not even aware that they had not finished the tool. A final possible improvement is suggested by a study of Bender et al. (2021), which indicates that it might be helpful for students' cognitive load to inform them about the role of the game design elements in the collaboration support environment, which was not done in this study. In summary, these improvements could reduce complexity and cognitive load as well as distraction for students using the online collaboration support.

Looking at the dropout rates in Task 4, they could be explained by a possible lack of perceived usefulness of students in relation to the final step of the RIDE assessment tool at the end of the study and their module. Previous research has identified perceived usefulness as the strongest predictor of behavioural intention (Davis, 1989), which according to Ajzen (2006) leads to actual behaviour, i.e., to perform the task. In this study, the final step of the tool was to discuss what went well, what could be improved, and to set goals for future collaboration. Since the end of the study was also the end of the module for the students, and student groups are unlikely to remain for future collaborations, students might have thought that setting goals of what could be improved would not be useful to them and therefore may not have completed this step. This explanation was supported not only by the observed dropout pattern, but also by a student's casual remark that setting future goals at the end of the module did not make much sense. However, since this step can help students in future collaborations with other students, it is therefore recommended to make it clear to them why and how they can benefit from this step, even at the end of the collaboration with this specific group. By doing so, students could perceive the step as more useful, leading to more student groups completing the task.

Overall, it can be concluded that the results of the collaboration products suggest that students benefit from instructional collaboration support that integrates expectation management and gamification. Even though this effect was not found in the data of the collaboration behaviour, we carefully state that the combination of expectation management and gamification can positively influence university students' collaboration outcome, as this was indicated by the collaboration products, and students might need training to evaluate each other's collaboration behaviour as they might have been hesitant to criticise others. Moreover, students in the gamified expectation management condition did not improve (more) than students in the other two conditions. This could be attributed to the greater cognitive demands and possible distractions of gamification, and thus

improvements to the gamification design in the instructional collaboration support are recommended.

Limitations, implications and future research

In regard to the research findings, there are also some limitations of the current study that should be considered. First, even though a pilot was conducted, the expectation management questionnaire used as a treatment in this study showed a lack of reliability and validity, which might have influenced the results. Specifically, the internal consistency and factor analysis revealed that out of the seven dimensions in the questionnaire only two dimensions, namely evaluating and scheduling, showed a satisfactory level of reliability and validity. Second, the data were collected based on a convenience sample of all first-year psychology bachelor students of one university attending a specific module. Therefore, the generalisability of the results is limited as it is not clear whether the results can be generalised to a larger student population. In particular, the university in this sample focused on improving students' collaboration and collaboration skills throughout their first study year, therefore the results might be different for students at other universities with less focus on collaboration than at the sample's university. Moreover, the sample in this study was psychology students. But, according to a study of Vedel, Thomsen, and Larsen (2015), students who choose to study psychology tend to have personalities with higher levels of openness and agreeableness compared to students who choose other degree programmes, such as law or economics. Both personality traits have been shown to be positive predictors of students' collaboration performance (Stadler, Herborn, Mustafic, & Greiff, 2019). Therefore, it is not clear if the results can hold for university students in other degree programmes. Another limitation regarding the sample is that the vast majority of students in this study identified themselves as either Dutch or German (84.9%), which can be considered as having fairly similar cultural backgrounds (Meyer, 2014). However, the dimensions used in this study to manage students' expectations stem from cross-cultural research and thus a more culturally diverse sample might lead to different results. Especially because previous research suggests that culturally diverse students are more prone to differ in their communication and behaviour patterns (e.g. Bird & Osland, 2005; Cagiltay et al., 2015; Popov, 2013), which could make expectation management support more necessary with culturally diverse groups. Finally, students were randomly assigned to groups by the teachers of the module; therefore there might have been students in the groups who had already collaborated and others who had not, which could have affected the results since Løvold et al. (2020) and Zambrano et al. (2019) note that previous experience in working together could benefit the collaboration.

Despite the limitations, this study is a starting point to understand the effects of expectation management and gamification on the collaboration outcome in the context of university students. This insight fills a gap in research since it is a unique combination, which to the best of our knowledge has not been studied yet. The current study also contributes to practice by providing practitioners with insights into how to design or improve instructional support for student collaboration in university education. Further research is needed to confirm the results before this promising combination should be applied more widely.

First of all, it would be advisable that follow-up studies improve the treatment with respect to the expectation management items. The questionnaire used in this study could serve as a starting point for further refinements targeting the low internal consistency of the dimensions communicating, power distancing, deciding, trusting, and disagreeing as well as the unexpected factor loadings of six items within these five dimensions. In addition, it was beyond the scope of this study to investigate the group process in more detail. Future research should therefore look more closely at this process to see if groups differ in the quality of their discussions or in finding a common ground within the group. For this purpose, qualitative studies with interviews or observations can be conducted, which also include the aspect of a safe space to see whether students felt safe to discuss preferences and expectations and to evaluate each other to develop. The results of such studies could explain the mechanism behind effects and would deepen the understanding and explanations. Moreover, further research is needed to gain more insight into whether group composition or group size makes a difference and affects the influence of expectation management and gamification on the collaboration outcome. This information would allow for more precise explanations and conclusions about the effects of expectation management and gamification on university students' collaboration outcome.

References

Adcock, A. (2000). Effects of cognitive load on processing and performance. Retrieved from https://www.researchgate.net/publication/242394074_Effects_of_Cognitive_Load_on_Processing_and_Performance

Ajzen, I. (2006). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32*(4), 665-683. doi:10.1111/j.1559-1816.2002.tb00236.x

American Psychological Association. (2020). Cognitive overload. *In APA dictionary of psychology*. Retrieved from https://dictionary.apa.org/cognitive-overload

Andrade, F. R. H., Mizoguchi, R., & Isotani, S. (2016). The Bright and Dark Sides of Gamification. In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *Lecture Notes in Computer Science. Intelligent Tutoring Systems* (Vol. 9684, pp. 176-186). doi:10.1007/978-3-319-39583-8_17

Antin, J., & Churchill, E. F. (2011). Badges in Social Media: A Social Psychological Perspective. *Gamification Workshop Proceedings of the ACM Conference on Human Factors in Computing Systems*. Retrieved from https://www.researchgate.net/publication/264799686_Badges_in_Social_Media_A_Social_Psychological_Perspective

Azmi, S., Iahad, N. A., & Ahmad, N. (2015). Gamification in online collaborative learning for programming courses: A literature review. *ARPN Journal of Engineering and Applied Sciences, 10*(23), 18087-18094. Retrieved from https://www.semanticscholar.org/paper/Gamification-in-online-collaborative-learning-for-A-Azmi-Iahad/dad60b289236b4b65bf5253c2c2c731d13459284

Baccon, L. A., Chiarovano, E., & MacDougall, H. G. (2019). Virtual reality for teletherapy: Avatars may combine the benefits of face-to-face communication with the anonymity of online text-based communication. *Cyberpsychology, Behavior, and Social Networking, 22*(2), 158-165. doi:10.1089/cyber.2018.0247

Bakhanova, E., Garcia, J. A., Raffe, W. L., & Voinov, A. (2020). Targeting social learning and engagement: What serious games and gamification can offer to participatory modeling. *Environmental Modelling & Software, 134*. doi:10.1016/j.envsoft.2020.104846

Bender, L., Renkl, A., & Eitel, A. (2021). Seductive details do their damage also in longer learning sessions – When the details are perceived as relevant. *Journal of Computer Assisted Learning*, 1-15. doi:10.1111/jcal.12560

Bird, A., & Osland, J. S. (2005). Making Sense of Intercultural Collaboration. *International Studies of Management & Organization, 35*(4), 115-132. Retrieved from http://www.jstor.org/stable/40397648

Blumenfeld, P. C., Marx, R. W., Soloway, E., & Krajcik, J. (1996). Learning with peers: From small group cooperation to collaborative communities. *Educational Researcher, 25*(8), 37-40. doi:10.3102/0013189X025008037

Cagiltay, K., Bichelmeyer, B., & Kaplan Akilli, G. (2015). Working with multicultural virtual teams: Critical factors for facilitation, satisfaction and success. *Smart Learning Environments, 2*(11), 1-16. doi:10.1186/s40561-015-0018-7

Challco, G. C., Mizoguchi, R., & Isotani, S. (2016). An Ontology Framework to Apply Gamification in CSCL Scenarios as Persuasive Technology. *Revista Brasileira de Informática na Educação, 24*(02), 67-76. doi:10.5753/rbie.2016.24.02.67

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409-426. doi:10.1016/j.compedu.2005.02.004

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. doi:10.2307/249008

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining Gamification. *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments, 11*, 9-15. doi:10.1145/2181037.2181040

Dicheva, D., & Dichev, C. (2015). Gamification in Education: Where Are We in 2015? *Proceedings of E-Learn: World conference on E-learning in corporate, government, healthcare, and higher education*, 1445-1454. Retrieved from https://www.researchgate.net/publication/281649410_Gamification_in_Education_Where_Are_We_in_2015

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69-106. doi:10.1111/j.1529-1006.2004.00018.x

Eshuis, E. H., ter Vrugte, J., Anjewierden, A., Bollen, L., Sikken, J., & de Jong, T. (2019). Improving the quality of vocational students' collaboration and knowledge acquisition through instruction and joint reflection. *International Journal of Computer-Supported Collaborative Learning, 14*(1), 53-76. doi:10.1007/s11412-019-09296-0

Falloon, G. (2010). Using avatars and virtual environments in learning: What do they have to offer? *British Journal of Educational Technology, 41*(1), 108-122. doi:10.1111/j.1467-8535.2009.00991.x

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice, 20*(2), 150-169. doi:10.1080/0969594X.2012.749212

Giannetto, D., Chao, J., & Fontana, A. (2013). Gamification in a Social Learning Environment. *Issues in Informing Science and Information Technology, 10*, 195-207. doi:10.28945/1806

Hagemann, V., & Kluge, A. (2017). Complex Problem Solving in Teams: The Impact of Collective Orientation on Team Process Demands. *Frontiers in Psychology, 8*(1730). doi:10.3389/fpsyg.2017.01730

Hakulinen, L. (2015). *Gameful approaches for computer science education: From gamification to alternate reality games.* (Doctoral dissertation). Aalto University, Helsinki.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing Self- and Peer-assessment: The students' views. *Higher Education Research & Development, 20*(1), 53-70. doi:10.1080/07294360123776

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101-111. doi:10.1016/j.tate.2013.07.008

Hooi, R., & Cho, H. (2014). Avatar-driven self-disclosure: The virtual me is the actual me. *Computers in Human Behavior, 39*, 20-28. doi:10.1016/j.chb.2014.06.019

Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 263-280. doi:10.1007/s11412-016-9238-2

Jones, D. W. (1996). Empowered teams in the classroom can work. *The Journal for Quality and Participation, 19*(1), 80-86.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*(1), 23-31. doi:10.1207/S15326985EP3801_4

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721. doi:10.1016/j.learninstruc.2007.09.021

Kromand, D. (2007). Avatar Categorization. *Proceedings of the Digital Games Research Association International Conference, 4*. Retrieved from http://www.digra.org/digital-library/publications/avatar-categorization/

Lajoie, S. P. (2005). Multimedia Learning of Cognitive Processes. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 623-646). Cambridge: Cambridge University Press.

Landers, R. (2015). Developing a Theory of Gamified Learning. *Simulation & Gaming, 45*(6), 752-768. doi:10.1177/1046878114563660

Landers, R., & Landers, A. (2015). An Empirical Test of the Theory of Gamified Learning. *Simulation & Gaming, 45*. doi:10.1177/1046878114563662

Lawrence, G. (2013). A working model for intercultural learning and engagement in collaborative online language learning environments. *Intercultural Education, 24*(4), 303-314. doi:10.1080/14675986.2013.809247

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2019). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193-211. doi:10.1080/02602938.2019.1620679

Lin, S. S., Liu, E. Z.-F., & Yuan, S.-M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning, 17*(4), 420-432. doi:10.1046/j.0266-4909.2001.00198.x

Løvold, H. H., Lindsjørn, Y., & Stray, V. (2020). Forming and Assessing Student Teams in Software Engineering Courses. In M. Paasivaara & P. Kruchten (Eds.), *Lecture Notes in Business Information Processing. Agile Processes in Software Engineering and Extreme Programming – Workshops* (Vol. 396, pp. 298-306). doi:10.1007/978-3-030-58858-8_31

Meyer, E. (2014). *The culture map: Breaking through the invisible boundaries of global business*. New York: Public Affairs.

Mickan, S., & Rodger, S. (2000). Characteristics of effective teams: A literature review. *Australian Health Review, 23*(3), 201-208. doi:10.1071/AH000201

Oxford University Press. (2020a). Expectation. *In Lexico.com*. Retrieved from https://www.lexico.com/definition/expectation

Oxford University Press. (2020b). Manage expectations. *In Lexico.com*. Retrieved from https://www.lexico.com/definition/manage_expectations

Pan, Y., & Steed, A. (2017). The impact of self-avatars on trust and collaboration in shared virtual environments. *PLoS One, 12*(12). doi:10.1371/journal.pone.0189078

Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*(4), 195-203. doi:10.1016/j.stueduc.2013.10.005

Pe, C., Mothe, C., & Brion, S. (2011). The Impact of Team Virtuality and Task Complexity on NPD Coordination Modes. *Proceedings of the 44th Hawaii International Conference on System Sciences*, 1-10. doi:10.1109/hicss.2011.417

Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior, 65*, 23-30. doi:10.1016/j.chb.2016.08.004

Popov, V. (2013). *Scripting Intercultural Computer-Supported Collaborative Learning in Higher Education.* (Doctoral dissertation). Wageningen University, Wageningen.

Ravi Shankar, P. (2017). *Mapping Cultural Differences: An empirical study to understand cultural differences in an organizational setting.* (Master's thesis). Tilburg University, Tilburg.

Razmerita, L., & Brun, A. (2011). Collaborative Learning in Heterogeneous Classes: Towards a Group Formation Methodology. *Proceedings of the 3rd International Conference on Computer Supported Education, 2*, 189-194. doi: 10.5220/0003338901890194

Saab, N., van Joolingen, W. R., & van Hout-Wolters, B. H. A. M. (2007). Supporting Communication in a Collaborative Discovery Learning Environment: The Effect of Instruction. *Instructional Science, 35*(1), 73-98. doi:10.1007/s11251-006-9003-4

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371-380. doi:10.1016/j.chb.2016.12.033

Sailer, M., & Homner, L. (2019). The Gamification of Learning: A Meta-analysis. *Educational Psychology Review, 32*(1), 77-112. doi:10.1007/s10648-019-09498-w

Sillaots, M. (2014). Gamification of higher education by the example of course of research methods. *Proceedings of the 13th International Conference on Web-Based Learning, 1*, 106-115. doi:10.1007/978-3-319-09635-3_11

Stadler, M., Herborn, K., Mustafic, M., & Greiff, S. (2019). Computer-Based Collaborative Problem Solving in PISA 2015 and the Role of Personality. *Journal of Intelligence, 7*(3). doi:10.3390/jintelligence7030015

Tobias, S., Fletcher, J. D., Bediou, B., Wind, A. P., & Chen, F. (2005). Multimedia Learning with Computer Games. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 762-884). Cambridge: Cambridge University Press.

Turan, Z., Avinc, Z., Kara, K., & Goktas, Y. (2016). Gamification and Education: Achievements, Cognitive Loads, and Views of Students. *International Journal of Emerging Technologies in Learning, 11*(07), 64-69. doi:10.3991/ijet.v11i07.5455

van Roy, R., & Zaman, B. (2017). Why Gamification Fails in Education and How to Make It Successful: Introducing Nine Gamification Heuristics Based on Self-Determination Theory. In M. Ma & A. Oikonomou (Eds.), *Serious Games and Edutainment Applications* (pp. 485-509). doi:10.1007/978-3-319-51645-5_22

Vedel, A., Thomsen, D. K., & Larsen, L. (2015). Personality, academic majors and performance: Revealing complex patterns. *Personality and Individual Differences, 85*, 69-76. doi:10.1016/j.paid.2015.04.030

Werbach, K. (2014). (Re)Defining Gamification: A Process Approach. In A. Spagnolli, L. Chittaro, & L. Gamberini (Eds.), *Lecture Notes in Computer Science. Persuasive Technology* (Vol. 8462, pp. 266-272). doi:10.1007/978-3-319-07127-5_23

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

Wick, J. (2013). Managing expectations: "What do you mean?". *The Consultant Pharmacist, 28*(1), 58-62. doi:10.4140/TCP.n.2013.58

Yanes, N., & Bououd, I. (2019). Using gamification and serious games for English language learning. *Proceedings of the International Conference on Computer and Information Sciences*, 1-6. doi:10.1109/ICCISci.2019.8716451

Zambrano, J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2019). Effects of group experience and information distribution on collaborative learning. *Instructional Science, 47*(5), 531-550. doi:10.1007/s11251-019-09495-0

Appendices

Appendix A. Assessment forms and grading rubric for the collaboration products

Assessment form used for the group collaboration product in project one:

	Points (0-10)
1. Introduction (15%)	
• Introduce the scope in the first paragraph. • Provide a clear definition and / or explanation of concepts, terminologies and theories that are relevant to the hypothesis. Use the original article, Module book Brain and Cognition and the micro lecture for inspiration. • Introduce previous research, and the hypothesis and results of the original study. • Conclude the introduction by clarifying the type of replication you are aiming to perform and if there is any major difference in terms of procedure and controlled variables in your replication.	
2. Method (20%)	
• Describe the research methods in a correct way with the appropriate sections (Participants, Materials, Design & Procedure, Data Analysis). You can complement your section with the Appendices (not included in the word count). • Data analysis for the replication is appropriately reported • Make sure your Method section would allow another researcher to replicate the experiment.	
3. Results (20%)	
• Write down the results in a logical order. • Add descriptive statistics to provide insight into data patterns. Add inferential statistics to provide statistical grounding for answering the research question.	
4. Discussion (25%)	
• Paraphrase the hypothesis and answer it. This answer is explained and supported by referring to the described results. Report if results are in line or not with the results of the original study. • Discuss and support the results of this experiment (both if replication succeeds or fails) in the light of the theories mentioned in the introduction, and findings and theories from at least 3 other relevant scientific articles you find your-self. • Discuss and support the limitations (when relevant) shortly (< 200 words) insofar they limit the ability of the current findings to address the hypothesis. And (only if replication fails) add a brief explanation of the probable reasons, consequences and future work needed. • Reflect on and give support for possible implications of the results for practical domains or research. • End the discussion with two to three sentences providing a clear conclusion to the article.	
5. Work attitude (10%)	
• Communication with the tutor has been good; clear agreements have been made and have been met. • The group has shown an active and professional attitude (the group: spoke English during project meetings, has attended the majority of the project meetings, was on time and has shown initiative and constructive project management skills). • Feedback and suggestions for improvement have been handled professionally and implemented properly.	
6. Quality of reporting (10%)	
• There is a clear integration and coherence between the different elements of the report. • The report has a logic and consistent structure • The language used is scientific and spelling and grammar are correct • All references in the text and the reference list are in accordance with the APA guidelines.	
Final Grade (1-10)	

Assessment form used for the group collaboration product in project two:

	Points (0-10)
1. Heuristic assessment (25%)	

• Provide a clear description of the product; i.e., what is it? How it is used? Which are the main functions?
• Introduce and explain the importance of Heuristic evaluation and the 11 design principles to inform a redesign
  ○ A short background on Heuristic Analysis and Wickens principles
  ○ brief connection of the Wickens principles to the cognitive factors of attention, memory and perception
• Perform an overall inspection of the product. Describe weaknesses (potential issues) and strengths of the product you identified in the overall inspection
• Discuss at least 5 principles you selected that you consider important for the redesign, connect these to the cognitive factors (attention, memory and perception) i.e., here you can extend the analysis you made before about the 11 principles and the theory handbook.
• Explain why you selected these principles and excluded the others for the evaluation of the product	

2. Redesign and Rapid Prototyping (15%)	
• Justify and explain at least three solutions you want to implement under the light of the cognitive factors (attention, memory and perception) i.e., Which interactive aspects you aim to improve? How will the changes affect the interaction you aim to achieve? • Ensure cohesion between the solutions and problems identified in HE • Describe your prototype • Explain and justify the design choices in light of your HE, cognition, and your target group. • Report the link to the prototype.	
3. Usability Protocol and advice (30%)	

• Describe the methods of the usability test.
• Report and describe the context of use and the findings of the usability test in terms of effectiveness (major and minor issues), and satisfaction.
• Analyse the data and discuss results by answering the following questions:
  ○ How many and which problems were experienced by participants
  ○ Does the new design solve the issues you identified in the HE? Are there new/unexpected issues identified by the end-users? Is the prototype usable?
  ○ To what extent is the prototype satisfactory?
• Draw a general conclusion about the usability of the prototype.
• Based on your conclusion on the prototype's usability, advice future steps that should be taken in the iterative design process to further progress the design of your prototype.	

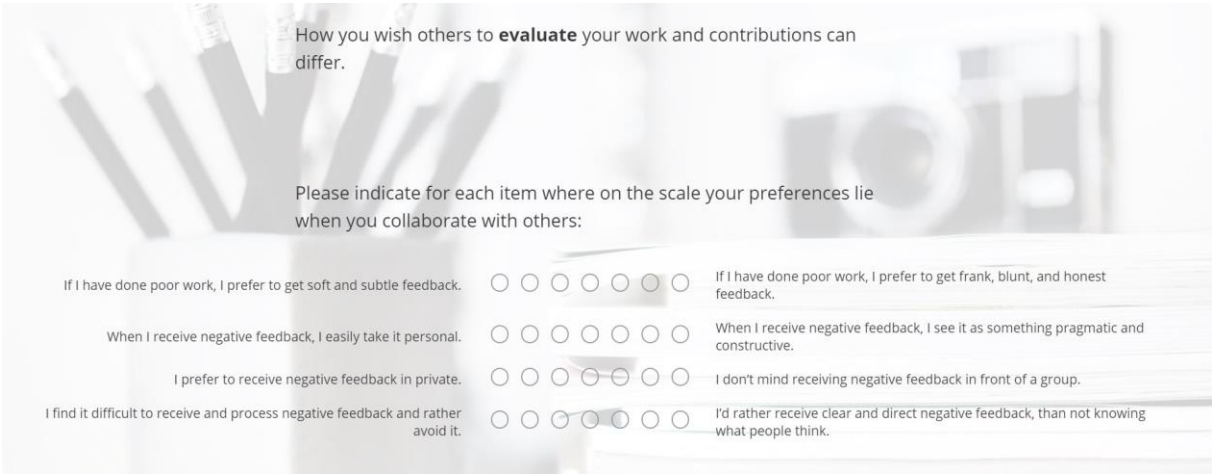
4. Work attitude (15%)	
• Communication with the tutor has been good; clear agreements have been made and have been met. • The group has shown an active and professional attitude (the group: spoke English during project meetings, has attended the majority of the project meetings, was on time and has shown initiative and constructive project management skills). • Feedback and suggestions for improvement have been handled professionally and implemented properly.	
5. Quality of reporting (15%)	
• There are clear integration and coherence between the different elements of the report. • The report has a logic and consistent structure. • The language used is scientific and spelling and grammar are correct. • All references in the text and the reference list are in accordance with the APA guidelines.	
Final Grade (1-10)	

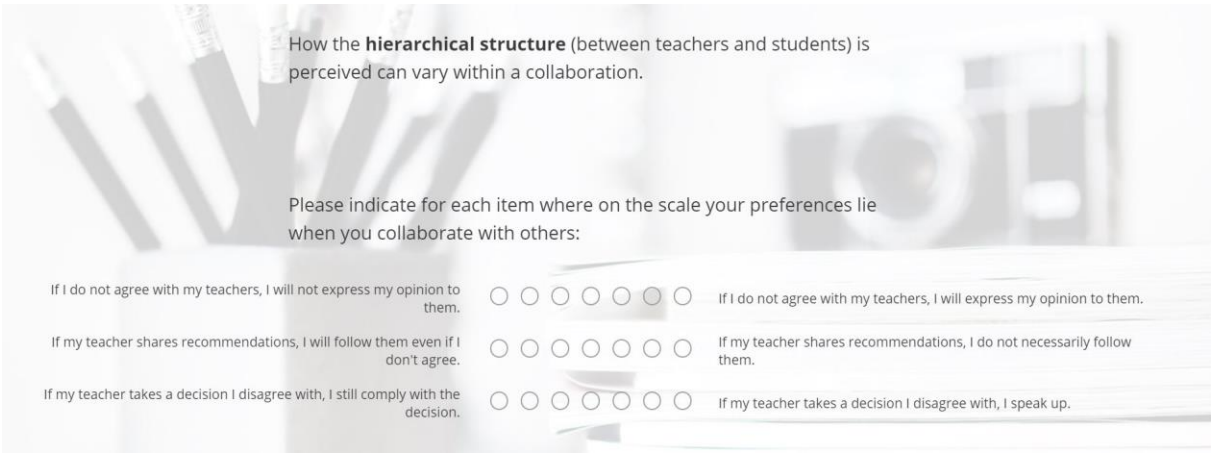
Grading rubric used for the individual collaboration products in project one and project two:

Criteria	*Requirements*	*Max. points*
Evidence	The answer covers all the main aspects and provides necessary details so that a person who is not aware of the project goal and its context could understand the answer	4
Accurateness	Use of accurate terminology/jargon and statements	3
Coherence	Logical organization of the content	2

Appendix B. Expectation management questionnaire items and presented results

Items:





How you feel a group should ideally come to a **decision** can differ.

Please indicate for each item where on the scale your preferences lie when you collaborate with others:

I think decisions can be made by an individual in the group.	○ ○ ○ ○ ○ ○ ○	I think decisions should involve everybody in the group.
I believe that decisions can be made even though not every group member agrees.	○ ○ ○ ○ ○ ○ ○	I believe that decisions should only be taken when everyone agrees.
It is more important to be time-efficient than to reach full consensus.	○ ○ ○ ○ ○ ○ ○	Reaching full consensus is more important than being time-efficient.

Trust plays an important role in collaborations. However, on which basis you trust someone might differ.

Please indicate for each item where on the scale your preferences lie when you collaborate with others:

In my opinion, a personal connection to my fellow group members is needed to build a collaborative relationship.	○ ○ ○ ○ ○ ○ ○	In my opinion, it is not necessary to personally connect to my fellow group members to build a collaborative relationship.
I prefer to invest time just getting to know my fellow group members — without discussing group work much.	○ ○ ○ ○ ○ ○ ○	I prefer to talk only about group work with my fellow group members.
I trust my fellow group members only after I spend time getting to know them personally.	○ ○ ○ ○ ○ ○ ○	I trust my fellow group members based on their task competence.

In a collaboration, you might occasionally **disagree**. How you view disagreement and how you make your group members aware of disagreement can vary.

Please indicate for each item where on the scale your preferences lie when you collaborate with others:

I believe that open debate and discussion is likely to ruin relationships.	○ ○ ○ ○ ○ ○ ○	I believe that open debate and discussion is an indicator of a healthy group.
I do not express my point of view when I disagree with one of my fellow group members.	○ ○ ○ ○ ○ ○ ○	I openly express my point of view when I disagree with one of my fellow group members.
I believe that frequently expressing disagreement should be avoided.	○ ○ ○ ○ ○ ○ ○	I believe that frequently expressing disagreement should not be avoided because it can benefit the group's success

Making appointments is essential for a good collaboration, however what you expect in regards to appointments and how you prefer to **schedule** your tasks and time can differ.

Please indicate for each item where on the scale your preferences lie when you collaborate with others:

If I have a group meeting at 9:00 a.m., It is acceptable to arrive 5, 10, or 15 minutes later.	○ ○ ○ ○ ○ ○ ○	If I have a group meeting at 9:00 a.m., I do not arrive any minute late.
I value flexibility over organization and structure.	○ ○ ○ ○ ○ ○ ○	I value organization and structure over flexibility.
I believe that a plan or schedule is a broad guideline that can be changed.	○ ○ ○ ○ ○ ○ ○	I believe that a plan or schedule should be followed closely once it is agreed upon.

Presented results (with example scores):

Communicating

The way how you communicate can vary from implicit (min -3) to explicit (max 3):
Implicit (meaning that your communication entails: sophisticated, nuanced, and layered communication; it might require people to read between the lines; messages can be implied instead of explicitly expressed; people might need to pay attention to your nonverbal communication)
Explicit (meaning that your communication entails: precise, simple, and clear communication; messages that can be understood at face value; repetition that fosters clarification; verbal communication is the main form of communicating and non-verbal communication is less relevant)

You scored **-3** on communicating.

This means that you prefer **implicit** over explicit communication, and might expect that from others with whom you collaborate.

Evaluating

How you evaluate somebody's work or contribution and how you wish others to evaluate your work and contributions can range from indirect (min -3) to direct (max 3):
Indirect (meaning that you feel negative feedback, when provided to you, should be subtle, in private, and with care; and that people's feelings should be taken into account; receiving negative feedback in front of a group might be problematic for you)
Direct (meaning that you feel negative feedback, when provided to you, is an evaluation, which is something pragmatic; to benefit most from it negative feedback should be provided as soon as possible in a clear rather than subtle manner)

You scored **-2** on evaluating.

This means that you prefer **indirect** over direct negative evaluations, and might expect that from others with whom you collaborate.

Power distancing

How the hierarchical structure (between teachers and students) is perceived can range from a hierarchical structure (min -3) to an egalitarian structure (max 3):
Hierarchical (meaning that you feel the distance between a teacher and a student is high and status is important; you see teachers as strong authority figures and consider it impertinent to express disagreement with them)
Egalitarian (meaning that you feel the distance between a teacher and a student is low and status is less important; you see teachers more as facilitators and feel more comfortable in expressing disagreement with them)

You scored **-1** on power distancing.

This means that you have no outspoken preference in terms of the hierarchical structure, but be aware that others might have them.

Deciding

How you feel a group should ideally come to a decision can vary from individual (min -3) to consensual (max 3):
Individual (meaning that you feel decisions can be made by individuals, this can either be a group leader or any other group member working on and responsible for a specific task)
Consensual (meaning that you feel decisions should ideally be made through unanimous agreement of all group members, even if it takes a long time)

You scored **0** on deciding.

This means that you have no outspoken preference in the way you come to a decision in a group, but be aware that others might have them.

Trusting

The basis on which you trust someone can vary from personal, social, relationship-based (min -3) to more practical, pragmatic, task-based (max 3):
Relationship-based (meaning that you feel an emotional connection should be a part of a good collaboration; trust is built through sharing social time together and is based on knowing the other members personally; relationships build up slowly over time but the investment is worth it)
Task-based (meaning that you feel trust in group work is built through task-related activities and is based on the reliability and usefulness of a person; relationships are built and dropped easily when group work ends; emotional connection is not necessary to perform well as a group)

You scored **1** on trusting.

This means that you have no outspoken preference on the basis on which you trust someone, but be aware that others might have them.

Disagreeing

How you view disagreement and how you make your group members aware of disagreement can vary from avoiding confrontation (min -3) to a more confrontational approach (max 3):
Avoiding confrontation (meaning that you feel that debate and disagreement within a group is something negative and will harm the harmony or relationships in the group; and that confrontation is inappropriate and should be avoided)
Confrontational (meaning that you feel that debate and disagreement within a group can yield to improved results and is therefore something positive; and that confrontation is appropriate and will not impede relationships or group harmony)

You scored **2** on disagreeing.

This means that you prefer a **confrontational approach** over avoiding confrontation for disagreements in a group, and might expect that from others with whom you collaborate.

Scheduling

What you expect in regards to appointments and how you prefer to schedule your tasks and time can differ from flexible (min -3) to linear (max 3):
Flexible (meaning that you value flexibility over organization; adaptable schedules and flexible time management are important to you; many tasks can occur simultaneously and interruptions are accepted)
Linear (meaning that you value punctuality and organization over flexibility; adhering to schedules and respecting deadlines are important to you; you want to complete one activity after another without interruption)

You scored **3** on scheduling.

This means that you prefer **linear** over flexible scheduling, and might expect that from others with whom you collaborate.

Appendix C. Factor analysis of expectation management questionnaire items

Factor Loadings Resulting from a Principal Axis Factoring Using Oblique Rotation (N = 197).

	Factor Loadings						
Items	EVA	SCH	TRU	DEC	POW	COM	DIS
When I receive negative feedback, I easily take it personal. - When I receive negative feedback, I see it as something pragmatic and constructive.	**.72**	-.13	.14	.19	-.05	-.12	.10
If I have done poor work, I prefer to get soft and subtle feedback. - If I have done poor work, I prefer to get frank, blunt, and honest feedback.	**.61**	.07	-.08	.03	.11	.12	.05
I prefer to receive negative feedback in private. - I don't mind receiving negative feedback in front of a group.	**.60**	.03	-.01	-.03	.09	-.03	-.05
I find it difficult to receive and process negative feedback and rather avoid it. - I'd rather receive clear and direct negative feedback, than not knowing what people think.	**.59**	.02	.06	-.06	.02	.09	.20
I value flexibility over organization and structure. - I value organization and structure over flexibility.	-.09	**.70**	.07	-.04	.00	-.02	-.00
I believe that a plan or schedule is a broad guideline that can be changed. - I believe that a plan or schedule should be followed closely once it is agreed upon.	-.00	**.67**	-.02	.10	.02	-.05	-.03
If I have a group meeting at 9:00 a.m., It is acceptable to arrive 5, 10, or 15 minutes later. - If I have a group meeting at 9:00 a.m., I do not arrive any minute late.	.15	**.63**	.04	.04	-.11	.08	.09
In my opinion, a personal connection to my fellow group members is needed to build a collaborative relationship. - In my opinion, it is not necessary to personally connect to my fellow group members to build a collaborative relationship.	.00	-.12	**.70**	-.05	-.01	-.02	-.10
I prefer to invest time just getting to know my fellow group members — without discussing group work much. - I prefer to talk only about group work with my fellow group members.	-.05	.30	**.46**	-.09	.05	-.05	-.00
I trust my fellow group members only after I spend time getting to know them personally. - I trust my fellow group members based on their task competence.	*.08*	*.10*	*.26*	*-.03*	*-.05*	*-.03*	*.12*
It is more important to be time-efficient than to reach full consensus. - Reaching full consensus is more important than being time-efficient.	-.03	-.06	-.02	**.70**	.09	-.02	.08
I believe that decisions can be made even though not every group member agrees. - I believe that decisions should only be taken when everyone agrees.	.09	.23	-.17	**.49**	-.06	-.06	-.07
I think decisions can be made by an individual in the group. - I think decisions should involve everybody in the group.	*-.12*	*.24*	*-.08*	*.36*	*.03*	*-.14*	*.39*

If my teacher takes a decision I disagree with, I still comply with the decision. - If my teacher takes a decision I disagree with, I speak up.	.08	-.05	-.04	.06	**.82**	-.06	-.03
If I do not agree with my teachers, I will not express my opinion to them.- If I do not agree with my teachers, I will express my opinion to them.	.24	.01	-.07	-.07	**.52**	-.20	-.04
If my teacher shares recommendations, I will follow them even if I don't agree. - If my teacher shares recommendations, I do not necessarily follow them.	*-.05*	*-.03*	*.19*	*.09*	*.37*	*.16*	*.01*
When I communicate, I expect people to sometimes read between the lines. - When I communicate, I make sure that people don't have to read between the lines.	.16	-.10	.08	.20	-.07	**-.79**	-.04
My communication is sophisticated, layered, and nuanced. - My communication is clear, concise, and direct.	.14	.16	.17	.07	.13	**-.43**	-.02
When I've discussed something, I assume the discussion is enough to create a shared understanding. - When I've discussed something, I like to finish with a clear concise statement to make sure everybody has the same understanding.	*-.15*	*.07*	*-.07*	*-.12*	*.07*	*-.32*	*.11*
I strive to communicate in an implicit manner. - I strive to communicate in an explicit manner.	*.23*	*.21*	*-.08*	*-.12*	*.07*	*-.19*	*.12*
I believe that open debate and discussion is likely to ruin relationships. - I believe that open debate and discussion is an indicator of a healthy group.	.13	-.07	-.02	-.09	-.11	-.15	**.75**
I believe that frequently expressing disagreement should be avoided. - I believe that frequently expressing disagreement should not be avoided because it can benefit the group's success.	.05	.09	-.00	.17	.05	.15	**.53**
I do not express my point of view when I disagree with one of my fellow group members. - I openly express my point of view when I disagree with one of my fellow group members.	*.21*	*-.08*	*.03*	*-.23*	***.41***	*-.07*	***.42***

Note. EVA = Evaluating. SCH = Scheduling. TRU = Trusting. DEC = Deciding. POW = Power Distancing. COM = Communicating. DIS = Disagreeing. Factor loadings over .40 appear in bold. Unexpected factor loadings appear in italic.