Can Gamification Improve Higher Education Students' Learning Outcome, Flow, Confidence and Motivation in Inquiry Learning Systems?

Bachelor Thesis

by Nico Reisch
University of Twente
Faculty of Behavioural, Management and Social Sciences
Department of Psychology
Learning Sciences

6th July 2022

Supervised by:
Dr. Henny Leemkuil
Dr. Johannes Steinrücke

Abstract

Gamification, the use of game-based elements to motivate and engage learners, is a promising tool in inquiry-based learning. The strategy can be integrated in many ways to make the students perceive the content as more meaningful. Especially in online laboratory environments focusing on STEM subjects, gamification seems to facilitate students' immersion and, consequently, might influence students' learning outcome, motivation, confidence and flow perception. Previous research indicated gamification could be effective, but large portions of the research have been described as fragmented. An online laboratory environment with interactive elements, visuals and explanations introduced the participating higher education students to the efficiency of pulley systems. A narrative element in form of a storyline was integrated into the experimental condition, while the control condition solely focused on the content to isolate and measure the effect of gamification. No significant differences were found between the conditions concerning students' motivation, performance, confidence and flow. However, when considering the development within the conditions, students in the control condition perceived the learning environment to be significantly less of a challenge after finishing the study than before the start. Subjects in the experimental condition reported constant levels of challenge throughout the study. However, these results should be considered cautiously, as solely one form of gamification element was introduced. Further research will be required to observe the effects in a less isolated setting. Combining multiple gamification elements over a longer time could likely lead to a more thoroughly perceived learning experience and thereby to more significant differences between the experimental conditions.

Keywords: Computer-based learning, Inquiry-based Learning, Gamification, Motivation, Performance

Can Gamification Improve Higher Education Students' Learning Outcome, Flow, Confidence and Motivation in Inquiry Learning Systems?

Learning is an essential skill to allow progress and improvement over time. In recent years, new trends in learning and instruction can be observed. More and more attention has been spread to various instruction forms promising better results than traditional learning. In this regard, inquiry-based learning, computer-based learning and gamification are some of the most relevant developments (Tsai, 2017). Many propositions to possibly improve the learning outcome have been created and implemented through recent technical improvements. By implementing new strategies and evaluating their effects, it could be possible to facilitate the learning process in the future. This paper aims at examining the effects of gamification on higher education students' learning outcomes, level of confidence, flow and motivation in inquiry-based learning.

Inquiry-based Learning

Inquiry-based learning (IBL) has gained increased attention in the past few years. IBL is an instructional approach focussing on problem-solving and stepwise discovery learning. Pedaste et al. (2015, p. 48) state that IBL aims at providing an "authentic scientific discovery process". Learners have to participate actively and are responsible for directing their learning process. Additionally, learners who are effectively introduced to the inquiry process gain more knowledge throughout various scientific procedures than learners without proper introduction (Zacharia et al., 2015). Learners deal with hypothesis-testing and experiments and discover different causal relationships over time. Therefore, the learning material is divided into five inquiry phases: Orientation, Conceptualisation, Investigation, Conclusion, and Discussion. Moreover, IBL is described as a learning cycle, as learners can re-enter the cycle at any point if something is unclear (de Jong & Lazonder, 2014; Pedaste et al., 2015). This allows learners to adapt their learning rate and engage in a self-regulated learning process. By partaking in the investigative discovery process, learners understand the situation and often perceive the acquired content as meaningful (Zacharia et al., 2015). In IBL, learners can identify and examine complicated aspects step-by-step, reflect upon their inquiry learning experiences and check their level of comprehension. IBL has been found to result in better outcomes than traditional learning. However, this only is the case when the IBL is guided and learners are supported in their inquiry process (Abdi, 2014; Pedaste et al., 2015; Zacharia et al., 2015). Otherwise, learners are at risk of being confused by the task at hand. Prior studies indicate scientific discovery learning to be ineffective, when it is not guided (de Jong & Lazonder, 2014). The learning environment requires a clear structure, and learners must be introduced to tasks of increasing difficulty over time to allow successful IBL (Pedaste et al., 2015). When implemented correctly, IBL was found to positively affect inquiry skills such as identifying problems, formulating hypotheses, analysing data and drawing conclusions (Mäeots et al., 2008 as cited in Pedaste et al., 2015). With the help of IBL, learners are actively involved and engaged in the topic, making them consider the situation differently.

Computer-based Learning

Computer-based learning is a common approach for practical learning. Recent developments in the COVID-19 pandemic emphasised the importance of having access to computers and the internet in education (Walters, 2020). Various forms of computer-based learning exist and are often used to enable the learners to engage in self-regulated learning (Devolder et al., 2012). Computer-based learning allows learners to precisely indicate their needs and receive tailored feedback or recommendations immediately. Accurate, instant feedback is crucial to allow learners to adjust their performance. Therefore, computer-based learning has been found to positively impact motivation and learning effectiveness (Rico García & Vinagre Arias, 2010). As motivation is one of the critical indicators for an effective learning outcome, it will be explored what components are part of the concept. In recent years, combining computer-based and inquiry-based learning has gained more popularity, frequently in so-called online laboratory environments (Brewer et al., 2013; Govaerts et al., 2013; Zacharia et al., 2015). In these online laboratories, participants can regulate their learning process and simulate a scientific discovery process in the inquiry learning environment. The combination leads to deeper processing of the content and significant advantages for the learner. Platforms, such as graasp.eu and golabz.eu, can be accessed globally and provide teachers with the opportunity to create, share, adapt and evaluate their learning environments. The observations suggest that computer-based learning provides many new opportunities for learning procedures and often affects learning outcomes positively.

What is Gamification?

One of the most promising trends in computer-based inquiry learning is the integration of gamification. *Gamification* is defined as “using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems” (Kapp, 2012, p. 10). According to Landers et al. (2018), gamification can be integrated into existing learning environments to apply a gamified learning process. The existing learning environment must already be determined to be effective, as gamification aims at improving the instruction and does not replace it. Therefore, gamification can only succeed in a suitable functioning learning system.

A clear distinction must be made between the concepts of game-based learning and gamification, as both use game-based elements to keep learners engaged over time, but still follow different aims and methods (Al-Azawi et al., 2016; Kapp, 2012). Gamification implies using game-based elements to motivate the learner to achieve a specific learning goal, while game-based learning is integrated into a serious game. In most cases, computer-based learning is used to introduce game-based elements to the learners. When the elements are correctly presented, gamification often leads to intense emotional responses from the learners (Buckley & Doyle, 2014; Chapman & Rich, 2018). The emotional responses improve motivational factors and affect the learners' performance. Gamification has been found to impact the learners' motivation among all age groups positively, but especially children, adolescents, and young adults are highly motivated (Brewer et al., 2013; Chapman & Rich, 2018; Kapp, 2012). By integrating game-based elements into an instructional program with debriefing, objective clarification and feedback, the elements can be introduced more effectively (Hays, 2005). When the program's main objective is clear to the learners, they can focus on the subject and do not need to be as attentive as otherwise. Well-introduced programs allow the learner to play to learn and not waste their capacities on learning how to play (Kapp, 2012). By thoroughly planning the creation of the learning environment and introducing its procedure, the designers aim at increasing the learners' motivation and making the learned content more meaningful to the participant. Brewer et al. (2013) found, that the implementation of gamified elements leads to a considerable increase in the participants' task completion rate. Depending on the utilised gamification elements, the planned intentions and actual consequences following can vary strongly.

Dimensions of Gamification Taxonomy

Many different forms of gamification can be integrated to improve the learning process. The game designers aim at establishing higher levels of motivation, which are directly linked to better learning outcomes in digital environments (Lin et al., 2017). When integrating the gamification element, it is crucial to recognise its unique effect on the participants. The various gamification forms strongly deviate in their effects on the learner (Chapman & Rich, 2018; Mekler et al., 2017). Therefore, examining and evaluating how the elements affect students' intrinsic and extrinsic motivation and learning outcomes is vital. According to Toda et al. (2019a), the broad range of gamification tools makes it challenging for the designer to correctly assess and predict the results of these tools. Prior research indicated the need for a taxonomy system to classify the gamification elements leading to the distinctions: performance, ecological, social, personal and fictional (Toda et al., 2019a). As each dimension follows different aims, it is crucial to consider their effect. The taxonomy can distinguish between various gamification elements and allows the successful planning and implementation to reach the initial aims in the learning environment. While elements in the dimension of performance focus on providing meaningful feedback to the learners, elements in the ecological category provide the learner with the feeling that individual choices in the learning environment matter. Social elements can enhance learning by improving performance through collaboration or comparison with others. Lastly, personal and fictional elements make the learner perceive the learning environment as being relevant because of interest through personal involvement and introduction of the context. The gamification taxonomy enables an objective assessment and comparison to determine how the individual elements from different dimensions can affect the learners' motivation and performance.

Effects of Individual Gamification Elements

By carefully assessing the participants' needs, the gamification elements in the learning environment can be tailored to allow better performance. Some of the most used gamification elements are reward structures, levels and time restrictions (Kapp, 2012; Mekler et al., 2017; Ortiz et al., 2016). The use of reward structures can be integrated, for example, as points, leaderboards or awards. By facing an obstacle to overcome, for example, a certain number of points or a rank on the leaderboard, learners are encouraged to continue engaging with the learning system (Mekler et al., 2017). With the aid of Todas' taxonomy system (2019a), these elements can be categorised as performance-based, allowing participants to evaluate and reflect upon their performances. Another prominent strategy in this category is scaffolding, which offers choices in difficulty and keeps learners from various knowledge levels engaged over time (Devolder et al., 2012; Kapp, 2012). Scaffolding also enables learners to reflect upon their comprehension, self-regulate their learning process and gain motivation through the increased challenge difficulty (Gressick & Langston, 2017). Reward structures are probably the most prominent gamification elements (Kapp, 2012; Ortiz et al., 2016). Nevertheless, it is often doubted whether they are the most effective gamification elements. Reward structures are dependent on other factors such as active engagement, visualisation, storytelling and problem-solving processes in the learning environment to allow a meaningful learning process (Kapp, 2012). These factors are introduced to gamification elements to make the content more meaningful to the student. By using a personal and fictional dimension, students can put the situation into context, understand the tasks' importance and immerse themselves in the learning environment (Toda, 2019a). Thus, increased immersion can be established through different gamification elements on the fictional dimension, such as avatars or a storyline. Besides, the content often becomes clear and more meaningful to the learners by integrating a personal dimension, for example, upcoming objectives or puzzles. Conversely, when learners do not see the reason and context behind their task, they are often confused. Confusion can potentially harm the students' performance and motivation.

Therefore, it is crucial to understand that using gamification elements does not necessarily lead to improvement and the unique situation of the learner always has to be considered. When the initial aim and students' needs do not connect to the implemented gamification element, the learners are likely to be confused or stressed. Certain elements have also been found to harm the learners' performance. The use of time restrictions can be incorporated to increase the pressure that participants are facing. A running clock is related to disengagement for the participants through external pressure (Toda, 2019b). The creator must know the gamification elements' chances and pitfalls. By implementing suitable gamification elements into the materials and providing the participants with a challenge, learners' knowledge gain and motivation can likely be increased.

Gamification for Higher Education Students

Therefore, it is crucial that the creator knows their target group and tailors the game to future users' interests and preferences. By taking these factors into account, learners' motivation can be enhanced. The focus of this study will be tailored to the needs of higher education students. Most higher education students have access to technology, which can be utilised in education and integrated into the learning process (Lundin et al., 2010). Using the technology can help gain new chances to motivate the students and improve their learning outcomes. As many higher education students already have experience with gamification, the implementation is expected to be well accepted by the learners (Ortiz et al., 2016). Gamification was used in a study to teach medical students how to communicate in their job and resulted in almost 80% of the students reporting to feel more confident and knowledgeable than before (Duque et al., 2008). When the learning environment is a good fit for the students, they will likely engage with the materials and increase their knowledge.

The Theory of Gamified Learning

The theory of gamified learning concerns the effects of the gamification elements on the learner. Examining and evaluating the connections between the game characteristics, instructional content, behaviour/attitude, and learning outcomes can predict whether the gamified learning will be successful (Landers, 2014). Instructional content and game characteristics influence the learning outcome, but in both cases, the effects are mediated by the behaviour/attitude of the learner. Furthermore, behaviour/attitude has been found to moderate the relationship between instructional content and learning outcome. In 2014, Landers stated that it is vital for all of the factors mentioned before to be included in the effects, as gamification only works when the game elements/instruction leads to a particular behaviour, which leads to a specific learning outcome. Therefore, observing possible mediation and moderation that affect the learning outcome is crucial. The theory of gamified learning is employed to observe the effect of the instructions and the game characteristics on the behaviour/attitude and then on the learning outcome. By considering these effects, it can be predicted whether the gamification element leads to the desired learning outcome.

Intrinsic and Extrinsic Motivation

Through gamification, the learners' motivation can be influenced intrinsically and extrinsically. It is crucial for game designers to thoroughly decide which game-based elements are the most beneficial for the application. Extrinsic motivation leads to action through external factors influencing motivation, like an achievement or score to beat. Intrinsic motivation leads to action through internal motivation factors, such as the striving for self-fulfilment (Kapp, 2012). Both extrinsic and intrinsic motivation are valuable additions to the learning process. However, intrinsic motivation is at risk of being undermined by extrinsic motivation, as intrinsic objectives are often less tangible, leading to less engagement over time (Kim, 2015). In many cases, learners are extrinsically motivated to improve their performance by receiving an indication of their traced performance and comparing it to others. However, reward structures can also lead to intrinsic motivation when the determination to improve is facilitated by internal factors, such as the own performance expectations (Kapp, 2012). The complicated relationship between intrinsic and extrinsic motivation is vital to estimating gamification elements' effects on the resulting motivation and learning outcome. Various studies state that especially intrinsic motivation is crucial for a successful learning process, as higher levels of intrinsically motivated students were connected with more success when using gamified learning (Buckley & Doyle, 2014; Kapp, 2012). Malone and Lepper found several factors that strongly influenced learners' intrinsic motivation. Their six central factors in raising intrinsic motivation regarding learning are challenge, curiosity, fantasy, cooperation, competition and recognition (Malone & Lepper, 2021). The more these six factors are present in a learning environment, the more likely participants will be intrinsically motivated. By observing the effect of intrinsic and extrinsic motivational factors in gamification, it is possible to recognise how the use of game-based elements affects the learners' performance.

Motivation

Motivation is a concept that is not only considered through the intrinsic and extrinsic dimensions, as there are various definitions and components of the concept of motivation. According to Rheinberg et al. (2001), to assess motivation in learning with the questionnaire on current motivation (QCM), there are four relevant factors for initial motivation: probability of success, anxiety, interest and challenge. The probability of success relates to the learner's confidence in their skills and ability to succeed. Anxiety in this context can be described as the fear of failure. Interest relates to the personal preferences of the learner. Challenge means that the learners recognise that they want to achieve something and are motivated to succeed by facing a challenge. The relevant motivational factors determine how learners react to gamification (Vollmeyer & Rheinberg, 2006). The average scores in the QCM can predict the students' performance based on their motivation. By respecting the four relevant factors to keep the learners' motivation high, the QCM can aid in detecting possible flaws in the online learning environment.

Flow

Gamification has much potential to keep learners engaged in the relevant material. If the learning environment is a good fit for the learners, they can find their flow state, which is defined as "a mental state of operation in which a person is fully immersed and focused in what he or she is doing" (Kapp, 2012, p.71). When gamification is applied, certain conditions are essential for learners to achieve their flow state. These conditions are related to facing achievable tasks, maintaining concentration, having clear goals, receiving instant feedback, controlling the action, getting involved effortlessly, not facing self-concern and losing the sense of time (Csikszentmihalyi, 1990; Sillaots, 2014). Game designers often aim for learners to find their flow state and engage with the learning material, but predicting the effects during the design phase is challenging. The learning environment must be designed carefully, and participants must use the program as intended (Hays, 2005). Game designers must predict how the game-based elements affect the learners' engagement. Due to the increased importance of deep work, measurements such as the Flow-Short-Scale (FSS) have been created to establish flawless learning environments for the participants (Vollmeyer & Rheinberg, 2006). Low reported flow levels can indicate potential flaws in the learning environment, hindering the learners from engaging with the material.

Aiming at Improved Learning Outcome & Higher Confidence Levels

The online-learning environment aims at enabling learners to actively learn about a relevant topic of interest. Gamification can be used to motivate and guide learners to increased knowledge. Different types of knowledge can be approached. Positive learning effects through gamification are related to procedural, declarative, conceptual and rules-based knowledge (Hays, 2005; Kapp, 2012). According to Sitzmann (2011), computer-based learning does not only have a positive outcome on procedural and declarative knowledge. However, it is also closely related to confidence, as learners who utilised game-based elements to learn were around 20% more confident than those who participated in traditional learning without computers. Self-confidence in their abilities is a measurement that can be used to predict the learners' future performance. By considering the learners' perception, the online learning environment can be tailored to the learners' needs. Consequently, the game designers need to comprehend who will use their application and how it will be used. According to Hays (2005), the instructional game can only be effective if the instructional objectives are met and the game is played as planned. If the online learning environment is not utilised as intended, learners will probably not focus on the central topics of the program. When designing the program, knowing who will engage with the online learning environment is crucial to enable an effective learning process leading to satisfying learning outcomes and high levels of confidence.

The Self-Determination Theory

Increased engagement and motivational levels regarding learning have been explained through various theories. A central theory to describe specific effects of gamification is the self-determination theory (SDT) (Kapp, 2012). SDT also differentiates between intrinsic and extrinsic motivation and indicates that high intrinsic motivation is related to higher engagement (Kam, 2018). SDT defines various factors that influence the motivational levels of learners. The three relevant factors according to SDT are autonomy, competence and relatedness. These can be defined as having feelings of control in a situation, being confident in own skills to defeat a challenge and connecting to others. When the learner perceives autonomy, competence and relatedness as satisfactory, the intrinsic motivation is positively affected and suggests a better performance. The three central factors are directly related to predicted enjoyment and future gameplay (Kapp, 2012). The SDT suggests that learners who feel like they determine their learning process are more likely to engage with the material. Therefore, the components of SDT should be considered to observe the learning environment, evaluate the current state and establish high motivational levels among the learners through adjustments in the environment. By considering the SDT, the learning process can most likely be facilitated, leading to improved intrinsic motivation, which ultimately contributes to an increased knowledge gain.

Crucial Considerations for Implementing Gamification

Gamification has been found to impact various forms of knowledge positively. However, Kapp (2012) emphasises the importance of tailoring the learning procedures in the environment to the form of knowledge. Depending on the type of knowledge gained, it needs to be decided which gamification element can be introduced to improve the learning process. According to Mekler et al. (2017), when motivational factors are not affected by the gamification elements, in many cases, it can be attributed to mistakes or uncertainties in the implementation. In recent years, research on gamification has been done in various subjects, for example, computer science, social sciences, language arts, math, physics and biology (Chang et al., 2008; Kapp, 2012; Ortiz et al., 2016; Tsai, 2017). However, critics mentioned it is crucial to consider that some of the research on instructional programs is fragmented and should not be generalised (Hays, 2005). Prior research dealt with a broad spectrum of content, tasks and participants. There is much discussion about which game-based elements can be introduced to which situations and topics and when gamification should not be used (Toda, 2019a). By closely examining the gamification effects and indications from prior research, flaws can be prevented and tailored to the specific requirements of learners in this domain (Mekler et al., 2017). However, the complexity and variety of gamification elements present the researchers, designers and instructors with a significant challenge. Each element's unique criteria and the effects of its use are hard to predict. Frameworks, such as the gamification taxonomy, can introduce fitting elements more effectively (Toda, 2019a). The design process of the learning environment has to be carefully planned, revised and adapted to the needs of the learners to allow successful inquiry learning.

Outlook on Gamification Use and its Effects

Gamification can be implemented into many learning environments dealing with several topics. According to Abdi (2014), integrating gamification elements into science subjects can increase the learners' interest in the subject, improve problem-solving skills and raise the perceived importance of scientific evidence. As STEM topics are practical and well-suited for inquiry learning (Chang et al., 2008; Pedaste et al., 2015), this study will be based on a physics subject. The following paper will examine the effects of gamification on higher education students' motivation, confidence, flow and learning outcomes in an online inquiry learning system about pulley systems. Inquiry-based learning and computer-based learning provide the students with a chance at a deep understanding of the content. By combining the elements to introduce gamification, students can learn differently and more effectively than in traditional learning. As mentioned above, both, the theory of gamified learning and the self-determination theory suggest improved motivation through gamification. Nevertheless, evaluating the effects of several gamification elements at a time is challenging (Toda, 2019a). Therefore, students were only introduced to the gamification element of a storyline to make the content meaningful for the learners. Thus, the focus of this study will be on narrative gamification elements. This research aims at examining how fictional gamification elements can enhance students' learning process and eventually facilitate their performance in the future. The research question that will be approached in this study is whether gamification use can improve higher education students' learning outcomes, flow, confidence and motivation in inquiry learning systems. Consequently, by observing differences in these dimensions, challenges and problems in the learning environment can be discovered to facilitate the students' learning process in the future.

H1: Gamification positively affects the learning outcome in a computer-based online inquiry-learning system among higher education students.

H2: Gamification positively affects the level of motivation in a computer-based online inquiry-learning system among higher education students.

H3: Gamification positively affects perceived flow in a computer-based online inquiry-learning system among higher education students.

H4: Gamification positively affects the confidence level in a computer-based online inquiry-learning system among higher education students.

Methods

Design

A between-subject design was employed to compare the gamified and control conditions. The development within the groups was also considered for the dimensions with more than one measurement. The study received ethical approval from the Ethical Committee of the Behavioural, Management and Social Sciences at the University of Twente (UT).

Participants

The participants were mainly recruited through the test subject pool of the UT and partly through convenience sampling. Subjects were required to have sufficient skills in English to take part in the study. Students of the UT were rewarded with 0.5 test subject credits for their participation in the study. In total, 82 higher education students participated in the study.

After reviewing the initially collected data, 27 were excluded from the analysis because of not consenting to participate, skipping learning phases or not finishing the study. Consequently, 55 peoples' data was considered for the data analysis (35 male, 20 female), with participants' age ranging between 19 and 28 years old (M = 21.8, SD = 2.2). In total, 28 students were considered for the experimental condition and 27 for the control condition. Before participating, the respondents were given information about the study and had to agree to the consent form. Thus, subjects were informed about the handling of their data and the procedure and were told they could quit the study at any point if they chose to do so.

Materials

Initially, it was planned to conduct the data collection in-person to ensure guidance for the participants in case it was required. Due to unforeseen health circumstances, the data collection was instead conducted online, and participants had to use their computers. The study was performed through the learning platform graasp.eu/golabz.eu. The platform allows learners to engage in guided computer-based inquiry learning processes, and users can integrate various elements for measurement. An inquiry learning system was established to introduce participants to the efficiency of pulley systems. The inquiry learning phases were taken as a guideline but were partly renamed and split to make the learning system more appealing to the subjects. Furthermore, two phases were added before the learning process to include consent and measure prior knowledge and motivation. As gamification has been proven most effective when integrated into a functioning learning system, several existing learning elements such as a "Pulley Simulator", a video explanation and pictures were incorporated into the learning environment. Throughout the inquiry phases, several explanations were shown to prevent confusion about the procedure and focus on the content. These elements promise the participants a progressive comprehension of the functioning of pulley systems.

Students in the gamification condition were introduced to a storyline related to the problem where two pictures were different, and the text of each phase was slightly adapted to describe the situation of a fictive character (see Appendix A). The control condition displayed similar content without incorporating the storyline and instead focused more on the technical aspects (see Appendix B). In the first phase of the environment with gamification, participants were introduced to the storyline of Laura, a fictional higher education student to relate to. In the control condition, participants instead received a short introduction to pulley systems without a background story. The "Conceptualisation" phase was split into "Demonstration" and "Combination" to make the participants engage without losing concentration. Throughout the phases “Demonstration”, “Combination”, and “Investigation”, subjects from both conditions watched the same videos and engaged in the same pulley simulator. However, in the condition with gamification, Laura was shortly referred to in the text in all of those phases. In the “Demonstration” phase, learners are introduced to the subject in the video and have to state three hypotheses and indicate their confidence level. Next, the “Combination” phase contains the second part of the video introduction. After experimenting with the simulator and its different variables in the “Investigation” phase, learners were introduced to the same practical problem in both environments in the “Conclusion” phase. However, the problem was framed differently, as the condition with gamification again referred back to Laura’s situation. The control condition just introduced the same numbers relevant to solving the practical problem. In the “Discussion” phase of both conditions, learners had to indicate other purposes for which pulley systems can be applied.

Apps such as “Quest 2.0” and “Hypothesis Scratchpad” were introduced to have the participants answer questions about the content and predict how the variables of a pulley system are related. These answers and predictions were used to assess the participants’ understanding and confidence. Next, the FSS and the QCM were implemented to measure how the subjects perceive the learning process flow and whether they are motivated. While the FSS is based on 13 normative items focusing on the dimensions of flow and worry, the QCM is constructed out of 18 normative items measuring the facets of challenge, interest, probability of success and anxiety. These measurements are based on seven-point Likert scale items ranging from 1 = not agree at all to 7 = completely agree (Vollmeyer & Rheinberg, 2006). Furthermore, elements such as “Time Spent Summary”, “App Overview”, and “Quiz Overview” were integrated into the learning environment to measure the participants’ metadata as exclusion criteria to limit bias and evaluate their performance. SPSS was utilised for the data analysis.

Procedure

Once the participants signed up for the study, they were randomly allocated to one of the two conditions. They received a link to the learning environment, where they had to register with a username. Deception was used to ensure the results were not influenced by being aware of the conditions. The subjects were told that the focus of the study will solely be on inquiry learning in online systems. Furthermore, participants were informed that the experiment would take about 30 minutes, but as they could dictate their learning pace, it can take longer or shorter to finalise. After consenting to participate in the study, participants had to indicate their confidence level, answer five questions regarding prior knowledge and fill in the items for the QCM. Once the opening questionnaires were finished, the actual inquiry learning process started. After finishing the inquiry learning process, participants had to answer ten open questions about the content. Additionally, participants had to rate their confidence level in knowledge regarding the subject in three normative items on ten-point Likert scales. Lastly, the participants had to fill in the QCM post-test and the FSS. After participants had finalised the questionnaires, they were informed about the deception used, and their responses were recorded and assessed.

Data Analysis

After the data collection, the data was reviewed in graasp.eu for strong outliers. The review led to 27 participants being excluded from the research due to the metadata, incomplete results or missing consent. The remaining participants' data were manually transferred to SPSS and then analysed. The two conditions were assessed regarding their performance, confidence, flow and motivation difference. Before the main hypotheses were investigated, the data was examined regarding the statistical assumptions of fitting variables, independence of the observations, no strong outliers, normality and homogeneity of variances.

Performance

The answers to the practical problem, related areas and the knowledge post-test were evaluated for performance (see Appendix C). The questions regarding prior knowledge were not taken into consideration for the assessment. In total, ten open questions were assessed, and for each of the questions, the respondents' answer was either coded as incorrect (0), partially correct (0.5) or entirely correct (1). The score for each question was then combined to calculate an overall performance score. Each participant's answers were considered several times and re-evaluated to preserve reliability. Next, t-tests were run to establish norm values for the two conditions to compare if the gamification element significantly influenced the participants' performance.

Motivation

As motivation was measured before and after the study, the participants' tendencies and changes in motivation were observed (see Appendix D). Participants indicate their current level of motivation by answering the QCM items of the facets: probability of success, anxiety, interest and challenge. At first, the scales were examined for significant differences in the pre-test of the conditions. Based on the assumption that there is no significant difference before the start of the learning procedure, the differences between the two measurements were computed for each scale. Independent and paired sample t-tests were then used to observe potential significant discrepancies in the development between and within the conditions.

Flow

For flow, the participants' answers to the FSS were assessed (see Appendix E). As the scale was only measured once, the participants' answers to the facets flow and worry were recorded after the subjects finished the study. Flow and worry were considered separately to comprehend and follow the participants' perception of the learning process. The corresponding item scores for both elements were combined, and the mean value was calculated. By running independent sample t-tests, it was determined whether there was a significant difference in the perception of flow and worry in the two conditions, which could lead to variation in levels of immersion for the participants.

Confidence

For confidence, the participants had to indicate their confidence level in one normative pre-test item, three generated hypotheses and three normative post-test items (see Appendix F). All seven items ranging from scores between 1 and 10 were used to measure the trait confidence. The scores for the pre-test items and hypotheses were combined to gain a mean score before the study. The post-test items were also combined, and the mean was calculated for the score after the study. For the post-test, the third item's coding had to be reversed, as it measured low confidence levels. By running independent sample t-tests for the pre-test, it could be assured that no significant prior difference between the groups exists. Based on this assumption, t-tests were utilised to compare the developments in either condition. Furthermore, paired sample t-tests were used to observe the development within either condition.

Assumptions

The statistical assumptions to conduct an independent sample t-test were examined. Because the study is a between-subject design with deception use and participants only could take part in either condition, the assumption of independence of observations has been met.

As the independent categorical variable for the analysis is always the assigned condition and the dependent variables are the measured factors on continuous scales that are evaluated, the combination is suitable for an independent and paired sample t-test.

The data was scanned for extreme outliers to preserve reliability using boxplots. Several outliers have been found; for example, the pre-measurement in anxiety had five outliers deviating from the norm. For most other scales, there were no strong outliers.

Next, the data was investigated regarding the normal distribution of the dependent variables on the continuous independent variables. The Shapiro-Wilk test was considered to detect variables violating the assumption of normality. Overall, the test resulted in a significance below the critical alpha value of 0.05 in four cases, leading to the rejection of the null hypothesis that there is a normal distribution. The confidence post-test, $W(28) = 0.914$, $p = .025$, and the worry test, $W(28) = 0.908$, $p = .017$, violated the normality assumption for the experimental condition. For the control condition, the anxiety pre-test, $W(27) = 0.871$, $p = .003$, and the challenge pre-test, $W(27) = 0.923$, $p = .047$, violated the normality assumption. As a result, the Mann-Whitney U test was introduced as a non-parametric measure to compare these dimensions.

Results

Performance

The content of the open questions was recoded to scores and evaluated to compare the performance level between the two conditions. Participants could obtain a maximum of ten points. On average, participants in the gamification condition ($M = 6.43$, $SD = 1.76$) scored lower than subjects in the control condition ($M = 6.74$, $SD = 1.78$). However, the difference in performance between the gamified condition and the control conditions is not significant, $t(53) = -0.65$, $p = .516$.

Motivation

The QCM measurements before and after the experiment were considered for the scales: anxiety, challenge, interest and probability of success (Table 1).

Before participation, the average level of anxiety in the gamified condition was slightly higher than in the control condition. The pre-test differences regarding anxiety were non-significant, $U = 367$, $p = .852$. After participation, the average level of anxiety was merely lower in the gamification condition than in the control condition, but the difference between the developments in the conditions was non-significant, $t(53) = -0.47$, $p = .640$.

The challenge level before participation was very similar between experimental condition and control condition, suggesting no significant prior differences, $U = 388$, $p = .866$. After engaging in the study, students in the experimental condition reported higher challenge levels than in the control condition. The differences in development between how challenging the tasks were perceived are non-significant, $t(53) = 0.94$, $p = .353$. When considering the development within the groups, no significant differences were found for the experimental condition, $t(27) = 1.04$, $p = .307$, but the control condition reported significantly lower challenge levels after the participation, $t(26) = 2.58$, $p = .016$.

Before taking part in the study, subjects in the experimental condition reported lower interest levels than in the control condition. The discrepancy in the pre-measurement is not significant, $t(53) = -1.16$, $p = .251$. After the students took part in the study, this discrepancy in interest was still recognisable, as subjects in the experimental condition were, on average, less interested than in the control condition. Regardless of the discrepancy in values, the differences in the development of interest have been determined not to be significant, t (53) = 0.06, p = .950.

Lastly, participants in the control condition indicated their probability to succeed higher than in the experimental condition. The difference between the pre-measurements is not significant, t (53) = - 1.00, p = .323. After engaging with the learning environment, subjects in the control condition still reported a higher probability to succeed than subjects in the experimental condition. Nevertheless, when comparing the differences in the development between the two conditions, the data is non-significant, t (53) = -0.05, p = .960.

Table 1

Comparison of Mean Scores and Standard Deviations from QCM Measurement

Condition	Gamification		Control	
	Before	After	Before	After
Anxiety	2.71 (1.25)	2.60 (1.13)	2.63 (0.82)	2.63 (0.99)
Challenge	4.52 (1.10)	4.31 (1.07)	4.54 (0.84)	4.08 (0.74)
Interest	3.81 (1.13)	3.74 (1.51)	4.18 (1.24)	4.10 (1.15)
Probability to succeed	4.06 (1.17)	4.26 (1.38)	4.34 (0.89)	4.56 (1.21)

Flow & Worry

For the FSS, the facets flow and worry were considered separately based on the students' responses after engaging with the learning environment (Table 2). On average, subjects in the gamified condition perceived higher flow levels than in the control condition. The difference between the perceived flow has been determined not to be significant, t (53) = 0.52, p = .605.

Participants in the experimental condition have also indicated higher levels of worry than in the control condition. However, once again, the results suggest a non-significant relationship between the conditions, U = 350, p = .634.

Table 2

Comparison of Mean Scores and Standard Deviations from FSS Measurement

Condition	Gamification	Control
Flow	3.83 (1.08)	3.69 (0.91)
Worry	2.35 (0.97)	2.17 (0.71)

Confidence

The mean scores for confidence were compared before and after participants engaged in the learning environment to observe the development (Table 3). Subjects in the gamification condition were slightly more confident before participation than in the control condition. The deviation between the pre-scores was not significant, $t\,(53) = 0.11$, $p = .912$. After engaging with the learning environment, subjects in the gamified condition were still slightly less confident than in the control condition. However, when considering the development within each group, participants in the control condition, $t\,(26) = -4.40$, $p < 0.001$, and in the experimental condition, $t\,(27) = -3.63$, $p = .001$, were significantly more confident about their knowledge at the end than in the beginning. Nevertheless, the difference between the conditions regarding its development has been determined not to be significant, $U = 367.5$, $p = .980$.

Table 3

Comparison of Mean Scores and Standard Deviations Confidence

Condition	Gamification	Control
Before	5.86 (1.53)	5.81 (1.90)
After	7.11 (2.38)	7.28 (1.84)

Time

Further examinations were done with the metadata produced directly through the learning platform graasp.eu. Several applications were integrated into the learning environment to measure and compare the conditions. Participants' time spent on the application was recorded and indicated that students in the experimental condition ($M = 39.61$, $SD = 16.59$) spent more minutes on the task than students in the control condition ($M = 36.89$, $SD = 12.54$). The mean time difference between the condition is non-significant, $t\,(53) = 0.68$, $p = .497$.

Discussion

Implications

The primary aim of this study was to determine whether gamification in inquiry learning systems can positively affect higher education students' learning outcomes and motivation. Students' level of confidence and perception of flow was also considered.

Motivation & Learning Outcome

The first hypothesis regarding the learning outcome was rejected, as no significant difference has been found between the control and experimental conditions. In the second hypothesis considering the level of motivation, no significant differences have been found between the two groups, leading to the rejection of the hypothesis. However, one of the study's central findings is the development of the dimension of challenge within the conditions, as the control groups reported a significantly decreased level of challenge while the experimental group did not. The significant difference suggests that the integrated narrative elements in the gamified learning environment likely contributed to the perception of a challenge for the students and kept them engaged.

The QCM pre-tests and prior knowledge questions indicate that participants in the control group were slightly more interested and estimated their probability to succeed higher than the experimental group before participation. Furthermore, the answers that some students made before the study to indicate their prior knowledge showed that several students were aversive to the subject and had negative associations with physics. Previous studies have shown that physics and other STEM subjects are suitable for the use of gamification (Chang et al., 2008; Kapp, 2012). However, the negative attitude of the participants towards the subject could impact the results to some extent. For several reasons, the average learning outcome was slightly higher for the control group. As the SDT states, autonomy, competence and relatedness are crucial factors in improving intrinsic motivation and performance in gamification (Kapp, 2012; Kam, 2018). The facets of the SDT are, to some extent, comparable to the dimensions of the QCM. For example, the probability to succeed and competence are both closely connected to the trust in own abilities. Besides, autonomy and relatedness are both relevant for how well the learners can immerse in the learning environment and influence whether the students consider the task challenging and interesting. In this case, the QCM dimension of anxiety is not fully applicable, as the students did not consider the task too pressuring based on their indications. The violation of the assumptions of homogeneity of variables and normal distribution was caused by the notably low response scores. The considerable overlap in the QCM results between the conditions suggests that the SDT dimensions of autonomy, competence and relatedness relevant to intrinsic motivation and learning outcome were not fully present in the gamification condition.

Furthermore, the storyline was possibly not relatable for some students leading to problems immersing in the learning environment. Many other studies have integrated several gamification elements simultaneously (Kapp, 2012; Ortiz et al., 2016). This can lead to the participants perceiving the learning experience to be more thorough and better immersion. However, as Toda (2019a) states in the taxonomy, various gamification elements aim at different goals to achieve. When multiple elements are combined, it would be problematic to isolate the effects caused by the elements. Nevertheless, introducing multiple, well-fitting gamification elements at once makes the students perceive the inquiry learning process as more thorough and leads to better immersion in the learning environment, which positively impacts learning outcome and motivation.

Flow & Worry

The third hypothesis regarding the flow perception was rejected, as the reported numbers only indicate minor discrepancies between the conditions. The reported answers on the FSS indicated that many participants scored extremely low on worry and consequently violated the assumption of normal distribution. Similarly to anxiety in the QCM, the FSS dimension of worry is likely to violate the assumption of normal distribution because participants did not have to worry about consequences in case of bad performance. Initially, the data collection was planned to be conducted in person. As the study was conducted online, it is still in question whether participants would have reported higher levels of worry and anxiety on average if the study had been conducted in person. The subjects might perceive the data collection to be more obtrusive and consequential. Several elements are crucial to allow the participants' perception of flow. For example, instant feedback could have likely led to higher reported flow levels (Sillaots, 2014). However, it was not possible to effectively introduce the element because of the online data collection. By integrating elements adhering to Csikszentmihalyi's flow conditions into the environment, students will likely engage with the material and report higher levels of flow with the tasks.

Confidence

Lastly, the fourth hypothesis related to the confidence level has been rejected, as no significant difference has been found between the conditions. However, the reported confidence level significantly increased throughout the study when considering the development within the conditions. Prior studies suggest that gamification can increase the participants' level of confidence (Kapp, 2012; Wan Hamzah et al., 2015). In this study, the level of self-confidence increased throughout the experiment; however, this was not caused by using gamification elements. Instead, it is more probable that the increase in confidence was caused by the explanations, simulation and scaffolding elements to introduce the concepts to the learners, which led to higher confidence in knowledge about the topic.

Conclusion & Outlook

Ultimately, none of the four main hypotheses examined in this study was significant. Although the main hypotheses examined in the study have been rejected, it is crucial to consider the background of these results to draw conclusions. Despite the results, gamification is a promising strategy to introduce learners to diverse topics and have them engage in inquiry-based learning on the computer. Many students have reported increased motivation and better learning outcomes through gamification (Chapman & Rich, 2018; Kapp, 2012; Ortiz et al., 2016; Wan Hamzah et al., 2015). With further technological advances and additional research, gamification seems destined to facilitate learning processes in the future. Gamification in online inquiry learning systems might be time-consuming to create, design and implement (Kapp, 2012). However, after successful implementation, many learners can engage with the materials simultaneously regardless of location. By incorporating appropriate supporting elements into the online learning environment, students can be encouraged to self-regulate their learning process (Devolder et al., 2012). Computer-based inquiry learning provides an added value as participants can self-regulate the learning process to fulfil their needs and receive tailored support and feedback. The pursuit of better and more efficient learning processes through technology implementation has been researched for many decades. According to McDonald (2021), inventions such as Skinners' Teaching Machine in the 1950s can be considered precursors for computer-based learning with the aim to provide individual support for students. The rich history of developed learning tools displays the constant importance of innovation regarding efficient learning procedures. Still, there is a lot of optimisation and consideration needed to consistently implement effective gamification elements.

Limitations & Recommendations for Future Research

Gamification leads to many challenges in design, as the learning environment has to fulfil the needs of the participants and therefore, it has to be closely considered in the design and implementation process of the learning environment. In most cases, ineffective use of gamification can be attributed to the implementation of the elements (Kim, 2015; Mekler et al., 2017). According to Kapp (2012), a team working on gamification in learning and instruction usually contains at least five members, including a project manager, artist, game designer, programmer and expert in the domain. Thus, each expert can focus on their primary task and collaborate with the other experts to create a thorough learning environment. The research team in this study was relatively small in comparison and led to several challenges in the creation process.

Another vital consideration is the time that participants spent in the learning environment. Despite the timely difference between the conditions not being significant, on average, subjects in the gamified learning environment took more time to finish the study. When comparing the time spent in the environment with other studies related to gamification in STEM, it stands out that the most effective studies were longitudinal studies over multiple weeks and often up to one semester (Ortiz et al., 2016). Throughout this study, some significant differences between the conditions were already recognisable, for example, the QCM score development within the perception of challenge. If the scope of the study was more extensive and participants had to spend more time in the environment, other more significant differences between the conditions would likely emerge.

For the QCM, FSS and the level of confidence normative items were introduced. As the items can only indicate the participants' perception and cannot be measured objectively, the data might be at risk for social desirability. In many cases, when self-reporting is utilised, participants feel the need to present themselves in a favourable image to fit in (van de Mortel, 2008). Therefore, it is possible that the participants' actual perception of motivational factors, flow and confidence is, to some extent, skewed in the reported scales. As most participants completed the learning environment after using their real name as the nickname in the system, the subjects might have reported dishonest values for items they considered controversial. By controlling for social desirability in the questionnaires, risk for bias can be strongly reduced.

One factor to consider is the division of participants into the conditions. As the subjects were randomly allocated to their condition, proportionally more female participants were in the experimental condition than in the control condition. When evaluating the scores, it was recognised that many females scored lower than the average for confidence and performance. The unequal distribution among the conditions might lead to the displacement of the accurate population scores. Most previous research indicates that the effectiveness of gamification does not depend on age or gender (Chapman & Rich, 2018; Kapp, 2012). Due to the utilised non-probability sampling method in this study, the results must be considered cautiously and should not be generalised, as the population of higher education students is likely not sufficiently represented. For further research, it is advisable to introduce matching of participants to the conditions based on the study field and the prior knowledge of the participants to minimise the prior differences and their effects on the scores.

These considerations connect to the statement by Hays in 2005 that it is crucial not to directly generalise the findings based on one topic and learning group. There is a large number of different gamification elements that can be integrated into the learning environment aiming at various goals, topics and target groups. In this study, the focus of the gamification element was classified as a storyline, which is considered a fictional element in the taxonomy (Toda et al., 2019a). Therefore, the gamification use might have been more effective if an element from another category in the taxonomy had been introduced, for example, performance, personal or social. Thus, the different effects of gamification elements are still in question and should be considered in research in the future.

The focus of the learning environment was on physics, as it is well suited for inquiry-based learning. Recent studies compared in a literature review have examined various STEM subjects among higher education students and were mostly considered effective (Ortiz et al., 2016). Further research will be necessary to determine whether gamification is useful when integrated into other domains and which gamified elements are most effective in improving the students' learning process.

References

Abdi, A. (2014). The effect of inquiry-based learning method on students' academic achievement in science course. *Universal Journal of Educational Research*, *2*(1), 37–41. https://doi.org/10.13189/ujer.2014.020104

Al-Azawi, R., Al-Faliti, F., & Al-Blushi, M. (2016). Educational gamification vs. Game Based Learning: Comparative Study. *International Journal of Innovation, Management and Technology*, 131–136. https://doi.org/10.18178/ijimt.2016.7.4.659

Brewer, R., Anthony, L., Brown, Q., Irwin, G., Nias, J., & Tate, B. (2013). Using gamification to motivate children to complete empirical studies in lab environments. *Proceedings of the 12th International Conference on Interaction Design and Children*. https://doi.org/10.1145/2485760.2485816

Buckley, P., & Doyle, E. (2014). *Interactive Learning Environments*, *24*(6), 1162–1175. https://doi.org/10.1080/10494820.2014.964263

Chang, K.-E., Chen, Y.-L., Lin, H.-Y., & Sung, Y.-T. (2008). Effects of learning support in simulation-based physics learning. *Computers & Education*, *51*(4), 1486–1498. https://doi.org/10.1016/j.compedu.2008.01.007

Chapman, J. R., & Rich, P. J. (2018). Does educational gamification improve students' motivation? if so, which game elements work best? *Journal of Education for Business*, *93*(7), 315–322. https://doi.org/10.1080/08832323.2018.1490687

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper and Row.

de Jong, T., & Lazonder, A. W. (2014). The Guided Discovery Learning Principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning*, 371–390. https://doi.org/10.1017/cbo9781139547369.019

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, *28*(6), 557–573. https://doi.org/10.1111/j.1365-2729.2011.00476.x

Duque, G., Fung, S., Mallet, L., Posel, N., & Fleiszer, D. (2008). Learning while having fun: The use of video gaming to teach Geriatric House calls to medical students. *Journal of the American Geriatrics Society*, *56*(7), 1328–1332. https://doi.org/10.1111/j.1532-5415.2008.01759.x

Govaerts, S., Cao, Y., Vozniuk, A., Holzer, A., Zutin, D. G., Ruiz, E. S., Bollen, L., Manske, S., Faltin, N., Salzmann, C., Tsourlidaki, E., & Gillet, D. (2013). Towards an online lab portal for inquiry-based STEM Learning at school. *Advances in Web-Based Learning – ICWL 2013*, 244–253. https://doi.org/10.1007/978-3-642-41175-5_25

Gressick, J., & Langston, J. B. (2017). The guilded classroom: Using gamification to engage and motivate undergraduates. *Journal of the Scholarship of Teaching and Learning*, *17*(3), 109–123. https://doi.org/10.14434/v17i3.22119

Hays, R.T. (2005). The effectiveness of instructional games: A literature review and discussion. Naval Air Warfare Center Training Systems Division (No 2005–004).

Kam, A. H. T., & Umar, I. N. (2018). Fostering authentic learning motivations through gamification: A self-determination theory (SDT) approach. *Journal of Engineering Science and Technology*.

Kapp, K. M. (2012). *The gamification of learning and instruction*. Wiley.

Kim, B. (2015). Designing gamification in the right way. *Library technology reports*, *51*(2), 29-35.

Landers, R. N. (2014). Developing a theory of gamified learning. *Simulation & Gaming*, *45*(6), 752–768. https://doi.org/10.1177/1046878114563660

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming*, *49*(3), 315–337. https://doi.org/10.1177/1046878118774385

Lin, M.-H., Chen, H.-g, & Liu, Kuang-S. (2017). A study of the effects of digital learning on learning motivation and learning outcome. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(7). https://doi.org/10.12973/eurasia.2017.00744a

Lundin, J., Lymer, G., Holmquist, L. E., Brown, B., & Rost, M. (2010). Integrating students' mobile technology in Higher Education. *International Journal of Mobile Learning and Organisation*, *4*(1), 1. https://doi.org/10.1504/ijmlo.2010.029951

Malone, T. W., & Lepper, M. R. (2021). Making learning fun: A taxonomy of intrinsic motivations for learning. In *Aptitude, learning, and instruction* (pp. 223-254). Routledge.

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048

Ortiz, M., Chiluiza, K., & Valcke, M. (2016). Gamification in higher education and STEM: A systematic review of literature. *EDULEARN Proceedings*. https://doi.org/10.21125/edulearn.2016.0422

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). Questionnaire on current motivation. *PsycTESTS Dataset*. https://doi.org/10.1037/t04507-000

Rico García, M., & Vinagre Arias, F. (2000). A comparative study in motivation and learning through print-oriented and computer-oriented tests. *Computer Assisted Language Learning*, *13*(4-5), 457–465. https://doi.org/10.1076/0958-8221(200012)13:4-5;1-e;ft457

Sillaots, M. (2014). Achieving Flow through Gamification: A study on Re-designing Research Methods Courses. *8th European Conference on Games Based Learning*, *2*.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, *64*(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

Toda, A. M., Klock, A. C., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., Bittencourt, I., Gasparini, I., Isotani, S., & Cristea, A. I. (2019a). Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learning Environments*, *6*(1). https://doi.org/10.1186/s40561-019-0106-1

Toda, A. M., Oliveira, W., Shi, L., Bittencourt, I., Isotani, S., & Cristea, A. (2019b). Planning Gamification strategies based on user characteristics and DM: A gender-based case study. In Proceedings of the Educational Data Mining 2019 conference (pp. 438–443). Montréal. Retrieved from http://arxiv.org/abs/1905.09146

Tsai, F.-H. (2017). The development and evaluation of a computer-simulated science inquiry environment using gamified elements. *Journal of Educational Computing Research*, *56*(1), 3–22. https://doi.org/10.1177/0735633117705646

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), 40-48.

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review*, *18*(3), 239–253. https://doi.org/10.1007/s10648-006-9017-0

Walters, A. (2020). Inequities in access to education: Lessons from the Covid-19 pandemic. *The Brown University Child and Adolescent Behavior Letter*, *36*(8), 8–8. https://doi.org/10.1002/cbl.30483

Wan Hamzah, W. M., Haji Ali, N., Mohd Saman, M. Y., Yusoff, M. H., & Yacob, A. (2015). Influence of gamification on students' motivation in using e-learning applications based on the motivational design model. *International Journal of Emerging Technologies in Learning (IJET)*, *10*(2), 30. https://doi.org/10.3991/ijet.v10i2.4355

Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A., Kamp, E. T., Mäeots, M., Siiman, L., & Tsourlidaki, E. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in Science: A literature review. *Educational Technology Research and Development*, *63*(2), 257–302. https://doi.org/10.1007/s11423-015-9370-0

Appendix A

Screenshots from Experimental Condition

Figure 1

Orientation Phase in the Experimental Condition

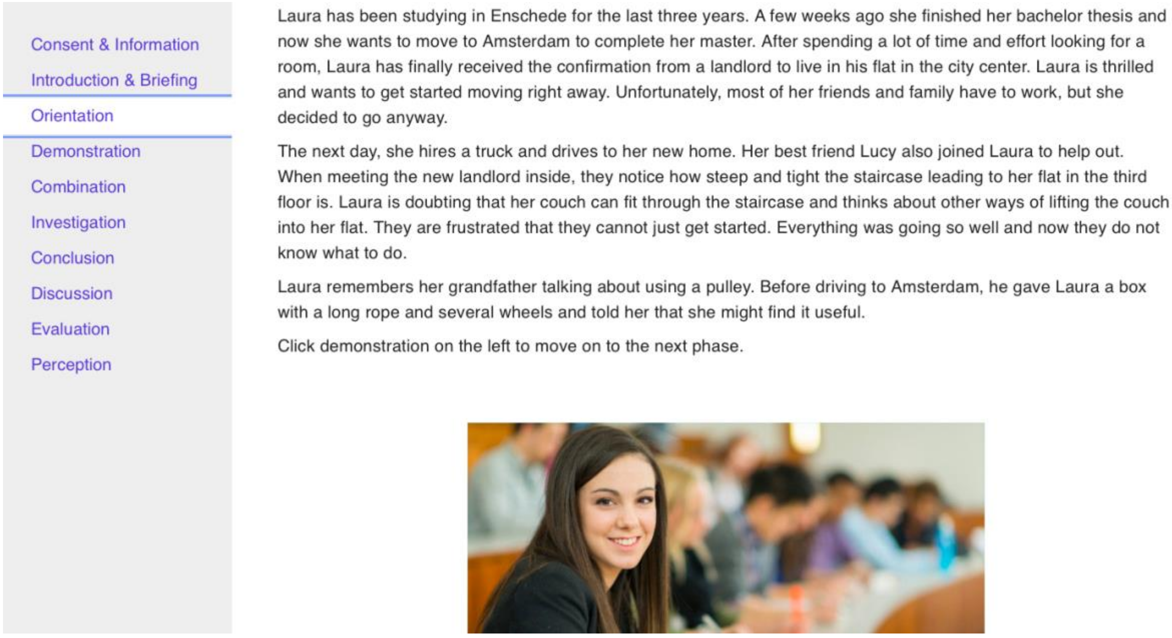


Figure 2

Demonstration Phase in the Experimental Condition

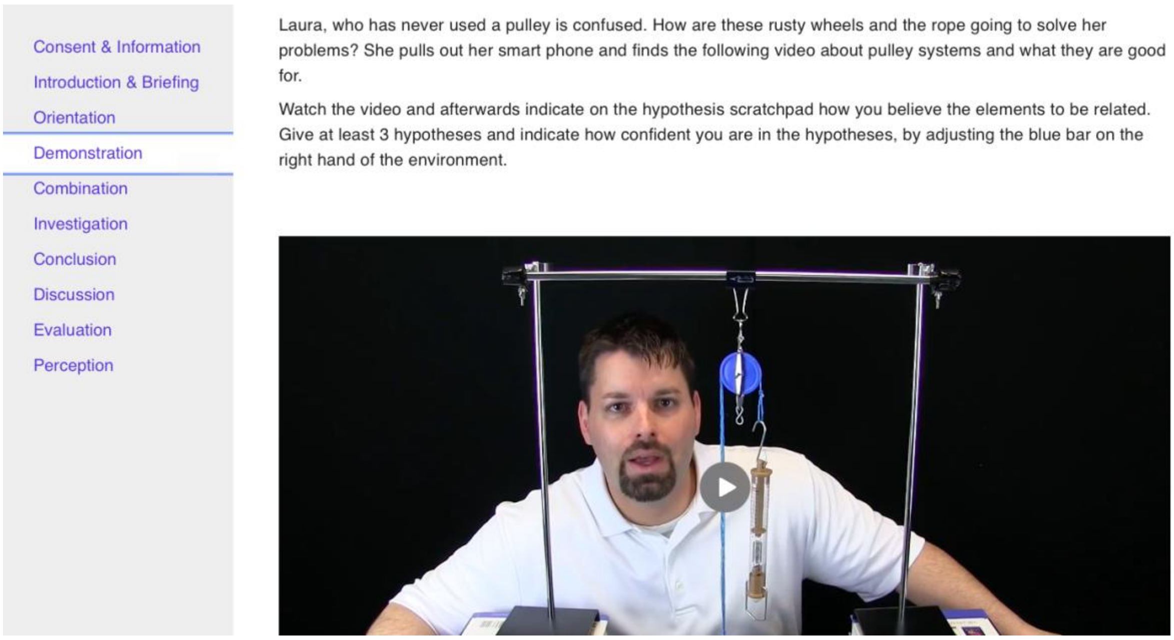


Figure 3

Combination Phase in the Experimental Condition

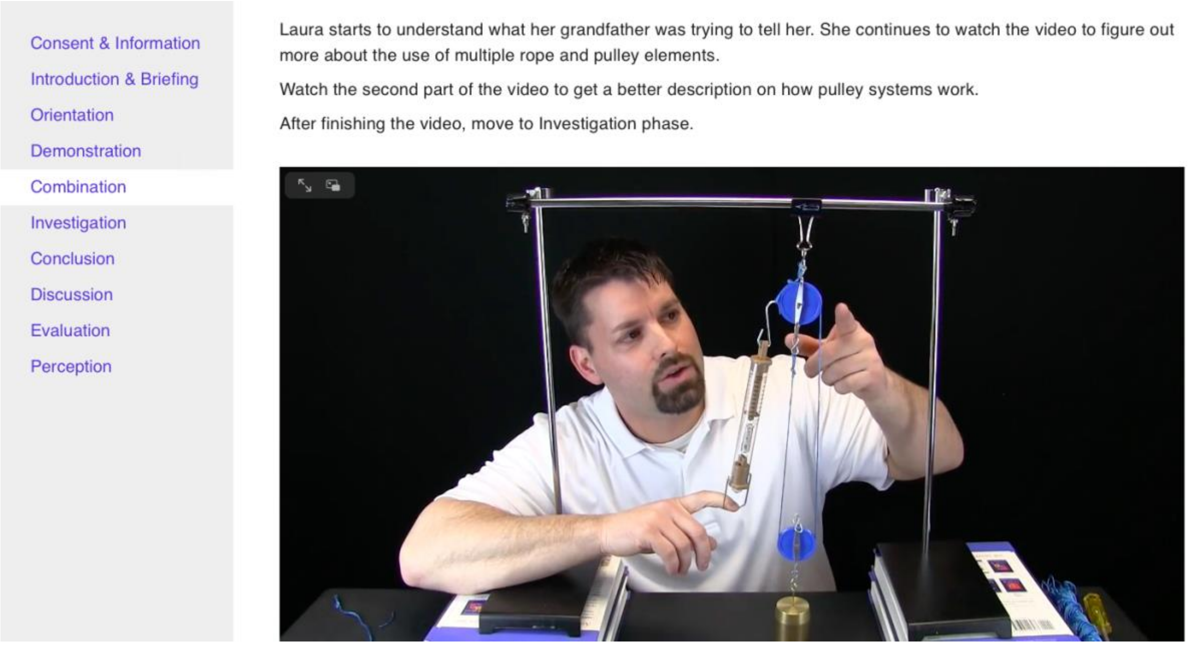


Figure 4

Investigation Phase in the Experimental Condition

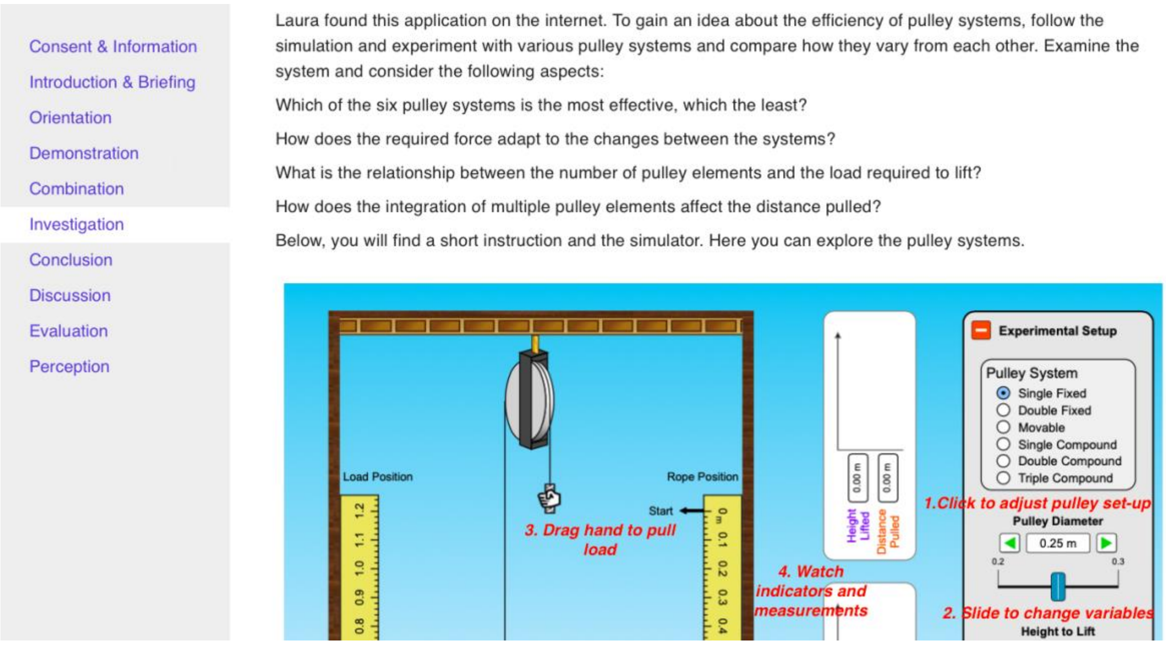


Figure 5

Simulator in Investigation Phase in the Experimental Condition

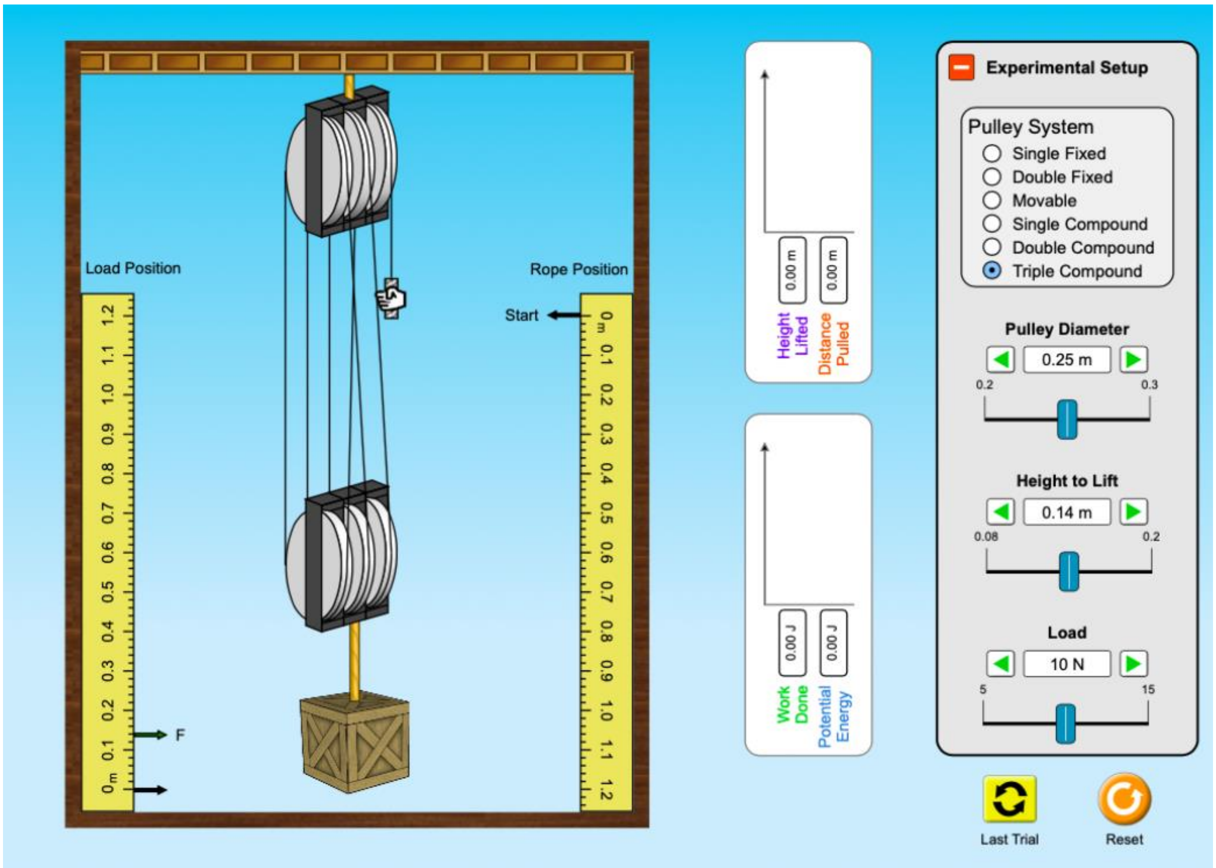


Figure 6

Conclusion Phase in the Experimental Condition

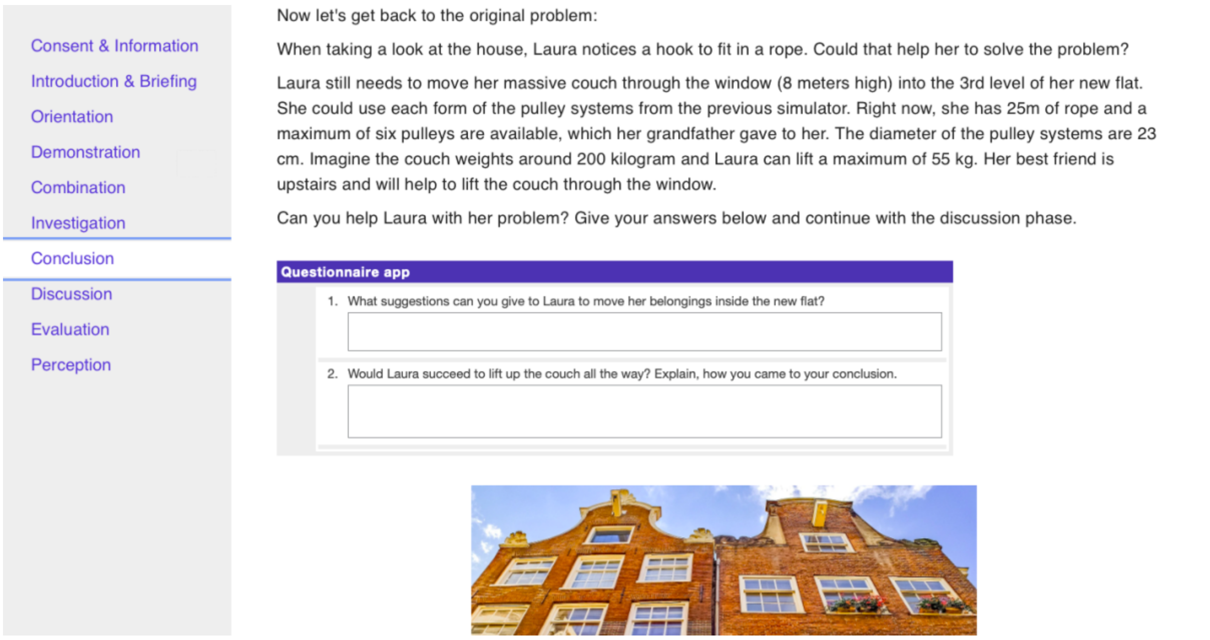
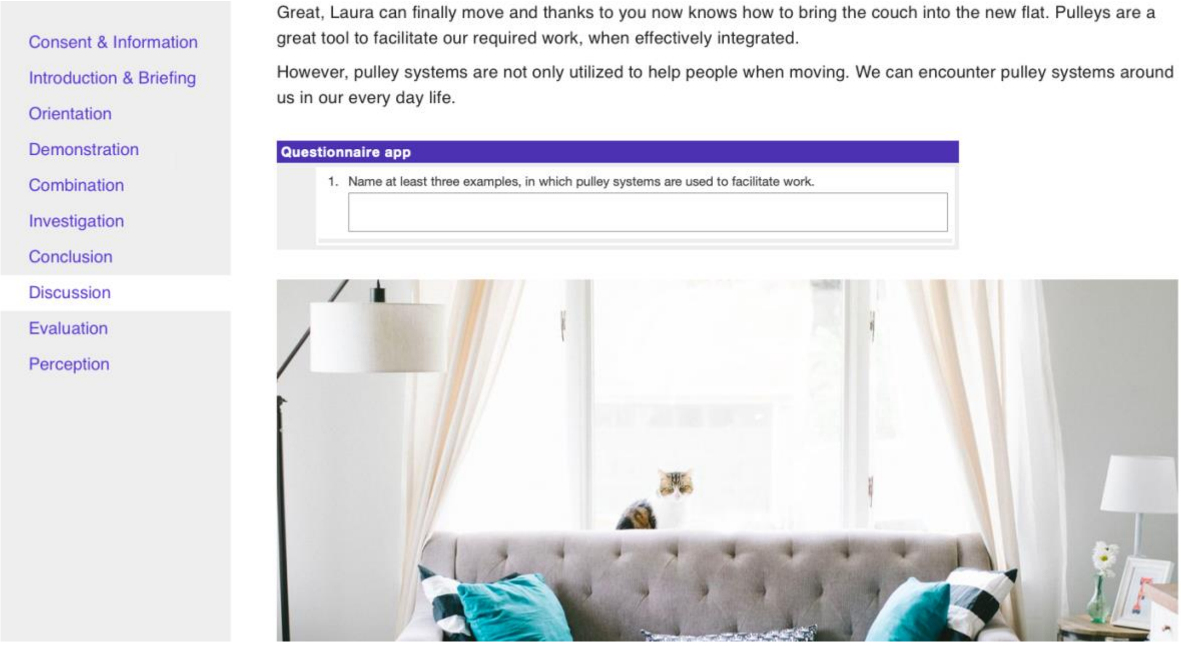


Figure 7

Discussion Phase in the Experimental Condition



Appendix B

Screenshots from Control Condition

Figure 8

Orientation Phase in the Control Condition

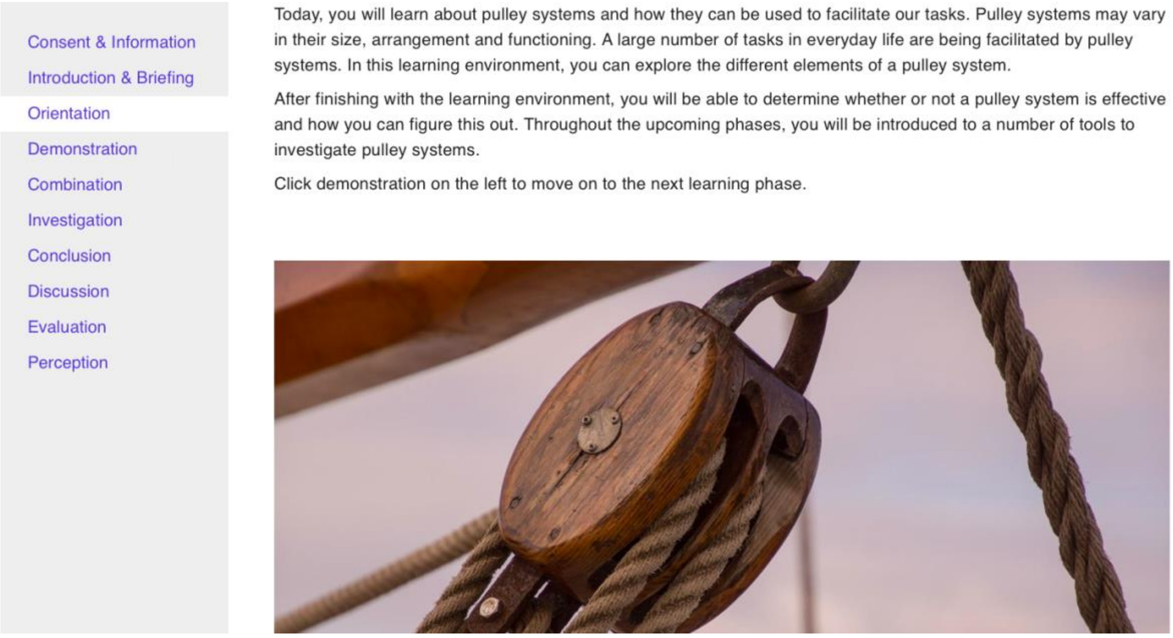


Figure 9

Demonstration Phase in the Control Condition

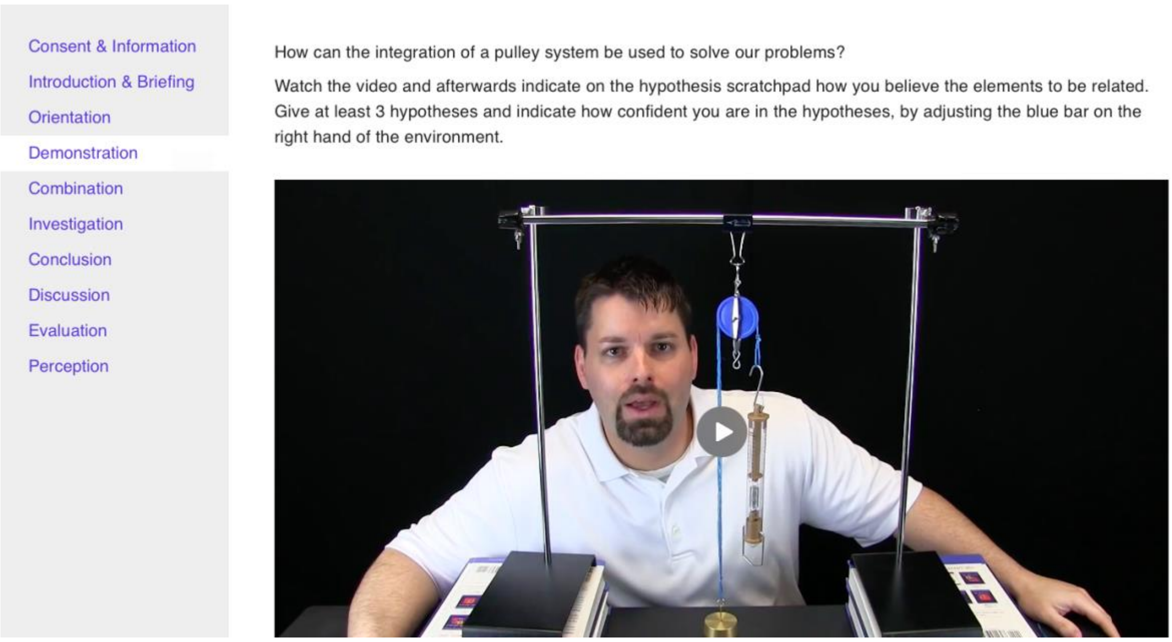


Figure 10

Combination Phase in the Control Condition

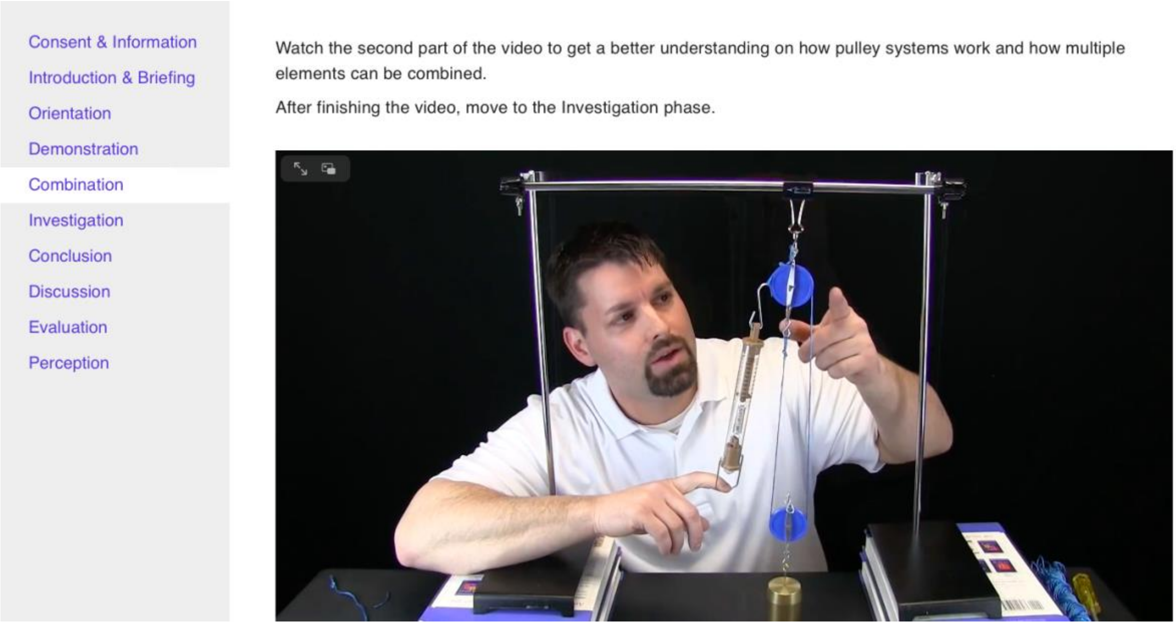


Figure 11

Investigation Phase in the Control Condition

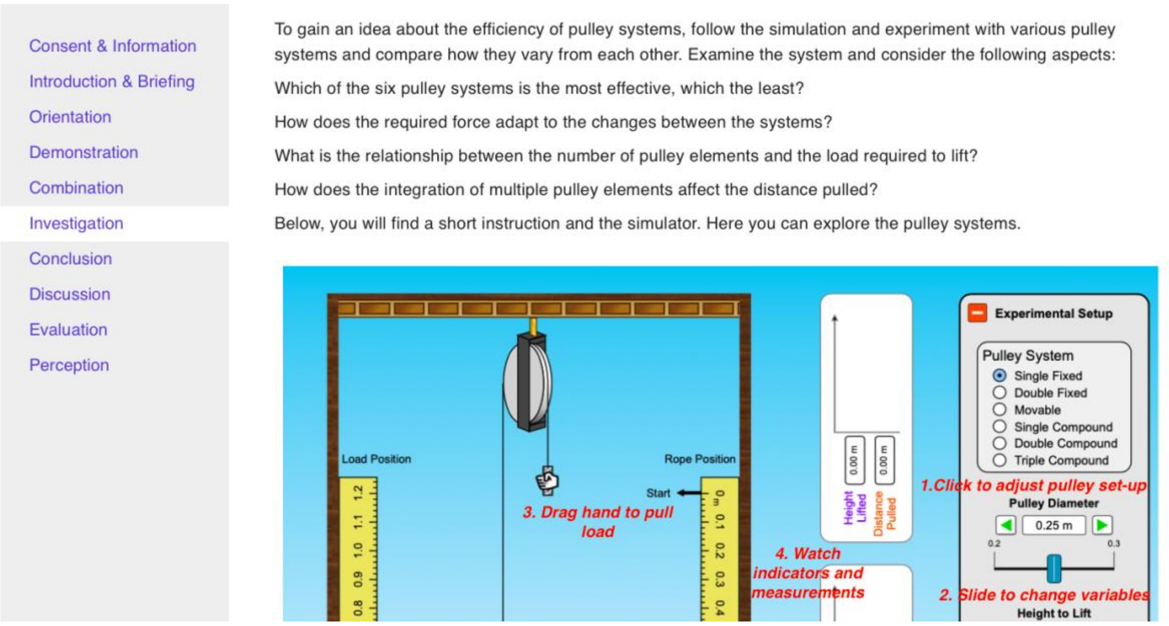


Figure 12

Simulator in Investigation Phase in the Control Condition

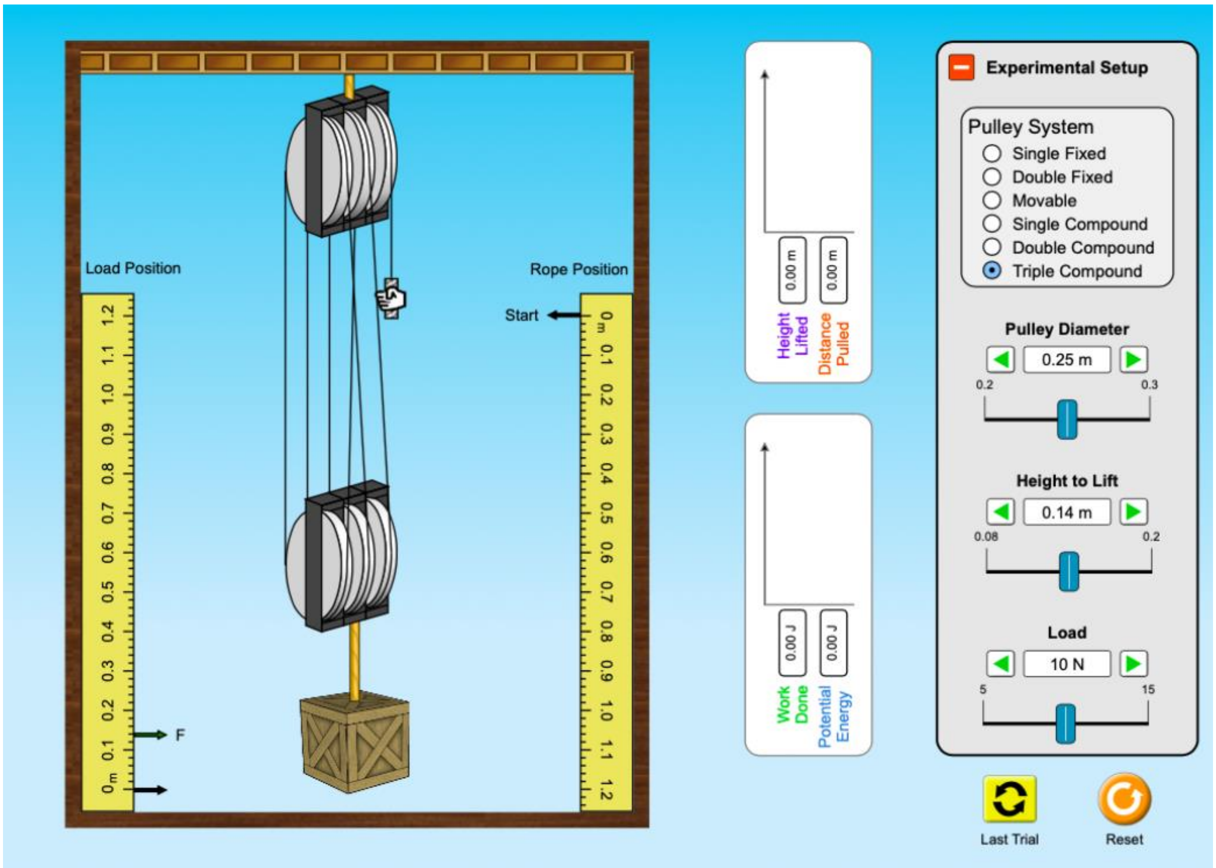


Figure 13

Conclusion Phase in the Control Condition

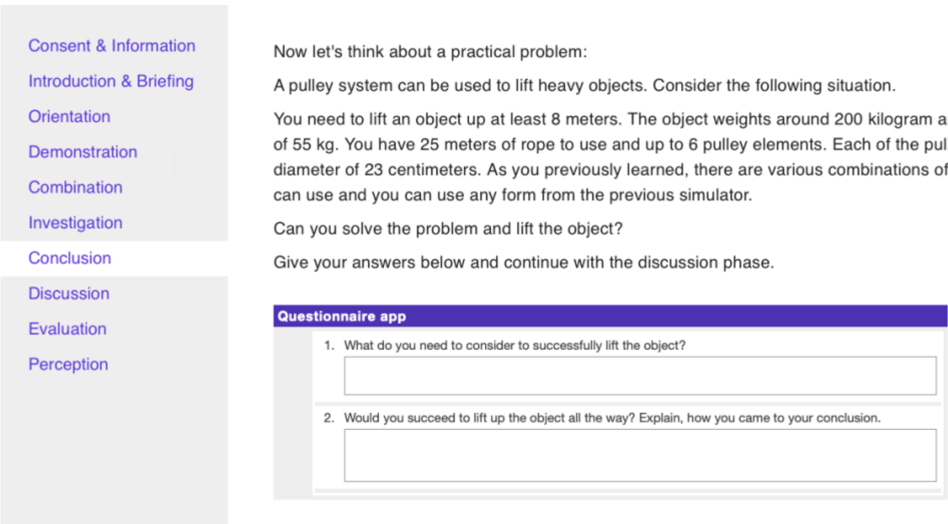
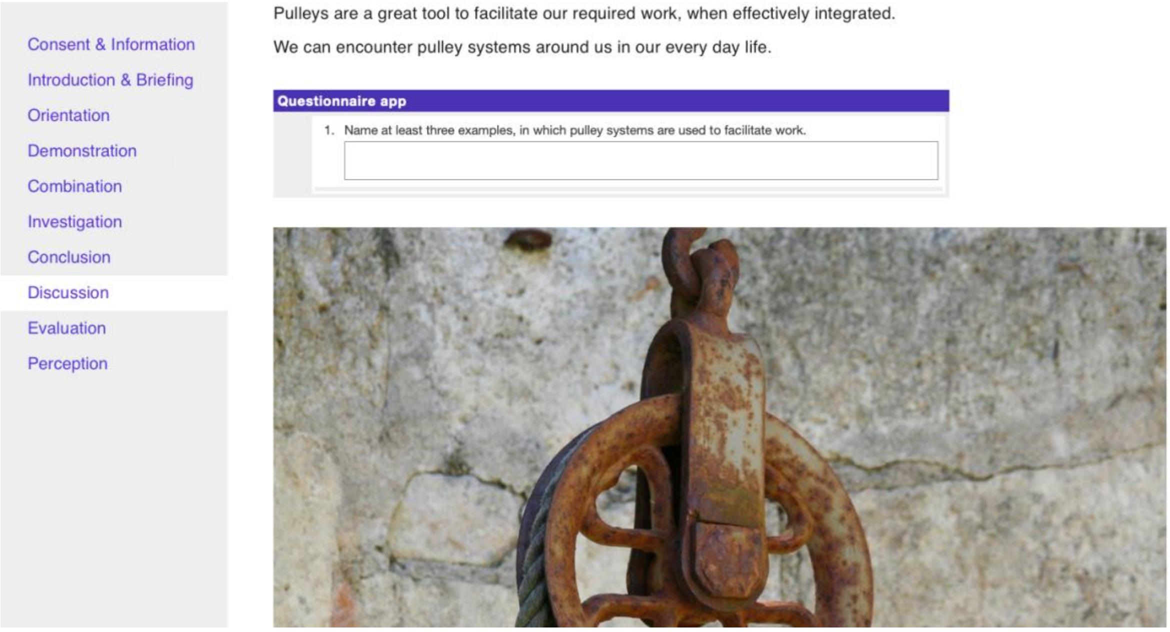


Figure 14

Discussion Phase in the Control Condition



Appendix C

Items used for Measurement of Learning Outcome

Item Number	Item	Measurement	Condition
1.	What are the first words that come to mind, when you think about physics?	Pretest	Both
2.	What do you know about pulley systems?	Pretest	Both
3.	Which unit is used to measure physical force?	Pretest	Both
4.	Can you name essential components for an effective pulley system?	Pretest	Both
5.	What suggestions can you give to Laura to move her belongings inside the new flat?	Posttest	Experimental
5.	What do you need to consider to successfully lift the object?	Posttest	Control
6.	Would Laura succeed to lift up the couch all the way? Explain, how you came to your conclusion.	Posttest	Experimental
6.	Would you succeed to lift up the object all the way? Explain, how you came to your conclusion.	Posttest	Control
7.	Name at least three examples, in which pulley systems are used to facilitate work.	Posttest	Both
8.	What do you know about pulley systems?	Posttest	Both
9.	Which unit is used to measure physical force?	Posttest	Both
10.	Can you name essential components for an effective pulley system?	Posttest	Both
11.	What is the effect of the pulley diameter size?	Posttest	Both
12.	How is the pulled rope length related to the height lifted?	Posttest	Both

13.	How can the required force be calculated?	Posttest	Both
14.	Which system is more effective? The double fixed or the double compound system and how big is the difference in efficiency? Why?	Posttest	Both

Note. Participants answered the open question spread across multiple phases.

Appendix D

Items used for the Questionnaire of Current Motivation (QCM)

Item number	Item	Dimension of Motivation
1.	I like to partake in experiments.	Interest
2.	I think I am up to the difficulty of this task.	Probability to succeed
3.	I probably won’t manage to do this task.	Probability to succeed, recoded negatively
4.	While doing this task I will enjoy playing the role of a scientist who is discovering relationships between things.	Interest
5.	I feel under pressure to do this task well.	Anxiety
6.	This task is a real challenge for me.	Challenge
7.	After having read the instruction, the task seems to be very interesting to me.	Interest
8.	I am eager to see how I will perform in the task.	Challenge
9.	I’m afraid I will make a fool out of myself.	Anxiety
10.	I’m really going to try as hard as I can on this task.	Challenge
11.	For tasks like this I don’t need a reward, they are lots of fun anyhow.	Interest
12.	It would be embarrassing to fail at this task.	Anxiety
13.	I think everyone could do well on this task.	Probability to succeed
14.	I think I won’t do well at the task.	Probability to succeed, recoded negatively

15.	If I can do this task, I will feel proud of myself.	Challenge
16.	When I think about the task, I feel somewhat concerned.	Anxiety
17.	I would work on this task even in my free time.	Interest
18.	I feel petrified by the demands of this task.	Anxiety

Note. Participants indicated their motivation on a seven-point Likert scale.

Appendix E

Items used for the Flow-Short-Scale (FSS)

Item number	Item	Dimension
1.	I feel just the right amount of challenge.	Flow
2.	My thoughts/activities run fluidly and smoothly.	Flow
3.	I don’t notice time passing.	Flow
4.	I have no difficulty concentrating.	Flow
5.	My mind is completely clear.	Flow
6.	I am totally absorbed in what I am doing.	Flow
7.	The right thoughts/movements occur of their own accord.	Flow
8.	I know what I have to do each step of the way.	Flow
9.	I feel that I have everything under control.	Flow
10.	I am completely lost in thought.	Flow
11.	Something important to me is at stake here.	Worry
12.	I won’t make any mistake here.	Worry
13.	I am worried about failing.	Worry

Note. Participants indicated their flow and worry perception on a seven-point Likert scale.

Appendix F

Items used for Measurement of Confidence

Item Number	Item	Measurement
1.	I am confident in my knowledge about pulleys and their effectiveness.	Pretest
2.	Students generate own hypothesis with scratchpad tool.	Pretest
3.	Students generate own hypothesis with scratchpad tool.	Pretest
4.	Students generate own hypothesis with scratchpad tool.	Pretest
5.	I know how to use a pulley system.	Posttest
6.	I am confident in my knowledge about pulleys and their effectiveness.	Posttest
7.	I am still in doubt, if I understood what pulleys are used for.	Posttest

Note. Participants indicated their confidence on a ten-point Likert scale.