What is the relationship between Storytelling and learning outcome?

Master of Educational Sciences & Technology

Duncan Stern

University of Twente

Supervisor: Henny Leemkuil & Natasha Dmoshinskaia

17-09-2022

Table of contents

Summary

Gamification has been increasingly popular over the last decade. It is increasingly used by educational designers, but not all aspects of storytelling have been researched or are still being researched for their effects on learning outcome. The current study aims at finding answers on a particular aspect of gamification, storytelling, on the learning outcome. Therefore, the following research questions were investigated: To what extent does storytelling contribute to the learning outcome of secondary education students in physics education? And: To what extend does storytelling contribute to the motivation perceived by the secondary education students in physics education? Participants had to study material which either included or excluded storytelling. This was investigated using 53 participants divided into two conditions, a control condition without storytelling and an experimental condition with storytelling. Participants had to answer questions on their motivation at the start and end of the research and a concluding test to assess their learning outcome. No significant difference was found between the two conditions in the current study. It can currently not be assessed whether storytelling affects students' learning outcome within this context. More research is needed to examine this effect thoroughly. For two aspects of motivation, significant differences were found within the control condition, a difference that was not present in the experimental condition. This suggests that storytelling contributed to this lack of decrease in motivation. Some limitations arise with the current results, and recommendations for future research are made. For example, more questions should be included to assess the effects of storytelling better, the sample size needs to be increased, and the storytelling operationalisation could be improved. The current research provides a solid foundation for future research, and it should be further investigated what the effects of storytelling are on students' learning outcome.

Keywords: Storytelling, Gamification, Learning Outcome, Motivation

Introduction

Gamification has been increasing in popularity over the past 20 years (Kapp, 2012). Kapp (2012) mentions both its necessity and effectivity have been recognised more by educational designers. As a result, gamification is increasingly being used in education (Kapp, 2012). Educational designers have adapted several gamification elements into different platforms to fit the needs of their users. However, not all gamification aspects have been validated for their educational use and effectiveness (Kapp, 2012). An example of such an aspect is storytelling. Storytelling is a method in which, by having a written narrative, the learner can read through the material as if they are reading a novel (Bechkoff, 2019). As Bechkoff (2019) has already shown, storytelling within gamification can increase the learning outcome and the efficiency of learning.

Nonetheless, more research needs to be conducted to assess the scale on which storytelling can enhance the learning performance of its users (Bechkoff, 2019). It is known that storytelling can positively influence the transfer of information, especially of complex information (Ibarra-Herrera et al., 2019). However, the effects of storytelling have yet to be discovered within an exact science, such as physics.

The increase in popularity of gamification requires the conduction of further research to assess all aspects of gamification. This research should aid in assessing what aspects of gamification we ought and ought not to use in our education. Storytelling is one of the aspects of gamification that is yet to be discovered for its practical use. The current research aims to find answers to how and to what extent storytelling contributes to a higher learning outcome. Therefore, a tool is be developed in which storytelling is used as the prime gamification aspect, compared to a similar environment in which storytelling is not adapted. According to earlier research, the tool with storytelling can and will yield a higher learning outcome (e.g. Bechkoff, 2019; Ibarra-Herrera et al., 2019). That can be beneficial for education, as storytelling can then be applied in different contexts to increase the learning outcome of its users. Bechkoff (2019) has made suggestions for multiple factors that explain the higher learning outcome caused by the addition of gamification. These suggestions form part of the foundation for the current research.

Furthermore, it is essential to note that the effect of storytelling is indirect. Storytelling does this through a mediating factor, being motivation (Kapp, 2012; Vollmeyer & Rheinberg, 2006). Motivation can be measured using four separate factors: self-efficacy, anxiety, interest and challenge. A more thorough explanation is provided in the theoretical framework to discuss all aspects of gamification, storytelling and motivation. In the current research the research questions are: To what extent does storytelling contribute to the learning outcome of secondary education students in physics education? And: To what extend does storytelling contribute to the motivation perceived by the secondary education students in physics education?

Theoretical Framework

The following section gives an overview of various research relevant to the research questions. As mentioned prior, parts of the framework provided by Bechkoff (2019) is used to examine the effects of storytelling on the learning outcome. However, first, gamification is discussed, as well as storytelling. Next, motivation is further investigated and the four factors explaining motivation are discussed.

Gamification

Gamification is the process of adding *game-based mechanics*, *aesthetics*, and *game-thinking* to a learning environment to engage people, motivate action, promote learning and solve problems (Kapp, 2012). Aspects of gamification can include badges, points and rewards. The addition of a scoreboard or storytelling and problem-solving can also be seen as gamification elements. This definition and elements of gamification will be used in the current research to describe gamification. There are several elements in this definition, which are discussed below to provide more information about this definition. 1) *Game-based*. Containing the elements of a game. 2) *Mechanics*. Using elements that are frequently used in games (e.g. levels, points, time-limits). 3) *Aesthetics*. Using the graphics and design of gaming interface. 4) *Game-thinking*. The idea of thinking about an everyday experience and converting it into an activity that has elements of competition, cooperation, exploration and storytelling. 5) *Engage*. Connecting players to the gaming process. 6) *People*. The consumer, learner or player. 7) *Motivate action*. The process that energises and gives direction, purpose and meaning to behaviour and actions. 8) *Promote learning*. All theory is based on educational psychology research and both motivates and educates learners. 9) *Solve problems*. By focussing on cooperation and competition, it allows players wanting to win and thus learn.

Operant conditioning

Operant conditioning is a method for learning in which consequences such as reinforcements and punishments are used in response to specific learner behaviour (Skinner, 1937). This causes the learner to form associations between the behaviour and the consequences, increasing or decreasing that behaviour (Skinner, 1937).

In gamification, operant conditioning is implemented through rewards (Bechkoff, 2019) by awarding people with in-game points on a scoreboard, trophies or in-game money (Landers et al., 2015). By awarding the players with in-game points, a scoreboard is created amongst players. The points act as a reinforcement for the player to be placed highest on the scoreboard (Landers et al., 2015). A higher rank on the scoreboard awards the player with a feeling of euphory (Landers et al., 2015). Moreover, a scoreboard allows for both reinforcements and punishments to be implemented. A punishment is implemented by deducting points from a player for undesired behaviour. The implementation of punishments comes with a danger. Players can get discouraged by the deduction of points and start to fear failure, resulting in a lack of answers (Landers et al., 2015). This occurs most often when a public scoreboard with multiple players is used and less when the scoreboard shows the individual player's score (Landers et al., 2015).

Another way of implementing operant conditioning is through the use of skill-based badges. The player is awarded a badge based on the mastery of a mechanic. This reinforces the player's behaviour to master mechanics. The reinforcement increases the player's desire to obtain badges (Landers et al., 2015). Thus, operant conditioning has a beneficial effect on the learning outcome (Bechkoff, 2019).

Storytelling

This research, however, will focus on storytelling as an aspect of gamification. As mentioned before, storytelling is a method in which a written narrative guides the reader through the material as if they were reading a novel (Bechkoff, 2019). Adding a story to the material provides relevance and meaning to the experience (Kapp, 2012). In early games, where storytelling originates from, the element of storytelling was more implied than implemented. A thin story layer made the game more exciting and engaging. Often, the name of the game and the addition of crude graphics were seen as enough to create a compelling story in the player's mind.

In contrast, storytelling in education requires a compelling story with interactive game elements to help the player learn the desired outcome (Kapp, 2012). Different elements provide this effect: characters, plot, tension, and resolution. By adding these elements together, a compelling story is created. The addition of this story is an aspect of gamification and is the aspect that is researched in the current research.

Conveying events in words and images is also considered storytelling (Giakalaras, 2016). Due to the nature of physics, being an exact science, this form of storytelling is more reachable within the current context. The effects of storytelling follow Bechkoff's (2019) framework, further discussed below.

The effects of storytelling specifically are, however, not necessarily explained through operant conditioning (Kapp, 2012). Gamification in general benefits most from the effects of operant conditioning. As mentioned priorly, the effects of storytelling are mostly explained through motivation (Kapp, 2012; Landers et al., 2015), changing attitudes rather than behaviour. This effect is further mentioned and discussed in sections below.

Motivation

Research has shown that gamification increases the player's motivation (Kapp, 2012). The level of effort players give to a particular activity might also increase, possibly leading to an increased learning outcome (Bechkoff, 2019). Therefore, it is beneficial to use gamification elements that enhance the player's motivation.

Two types of motivation are distinguished: intrinsic and extrinsic motivation (Ryan & Deci, 2000). Intrinsic motivation is the motivation that occurs by the enjoyment the activity causes. This means no apparent rewards are present outside of the activity that increase the intrinsic motivation (Ryan & Deci, 2000). Extrinsic motivation is caused by the desire to obtain a reward or avoid a punishment. This shows the connection between operant conditioning and motivation. Extrinsic motivation can be influenced by operant conditioning (Landers et al., 2015; Ryan & Deci, 2000).

Intrinsic motivation contributes more to the learning outcome than extrinsic motivation (Landers et al., 2015). If no intrinsic motivation is present, providing players with extrinsic motivation has little to no effect on their motivation (Ryan & Deci, 2000). Adding rewards to a learning activity that the learner does not find interesting is not beneficial to the learning outcome and, therefore, the player must perceive both types of motivation (Landers et al., 2015).

Gamification can also cause an increase in intrinsic motivation (Malone, 1981). Malone (1981) has developed a theory for intrinsically motivating instruction. Their theory distinguishes three key elements to make gamification motivational: challenge, fantasy and curiosity. Adding a meaningful goal or difficulty level to the instruction creates an uncertain outcome. This uncertainty intrinsically motivates the player to put more effort into their performance. Malone (1981) describes this process as challenge to increase intrinsic motivation. To include fantasy, an environment in which it "evokes mental images of things not present to the senses or within the actual experience of the person involved" (Malone, 1981, p. 360) is necessary. It enables the player to use existing knowledge to understand new phenomena. Storytelling could play a role in creating fantasy (Bechkoff, 2019). Curiosity is evoked by providing an optimal level of informational complexity and an exciting environment. This can be both of sensory and cognitive nature. All three elements should be included in the instruction to increase the player's intrinsic motivation (Malone, 1981).

Lepper (1988) suggests another theory to motivate learners. Their theory allows for extrinsic motivational techniques to be absent. The theory consists of four principles. The principles control, challenge, curiosity and contextualisation ought to be built into the instructional activities (Lepper, 1988). The player needs to be provided with decisions independent of outside influences to include control and appropriate goals must be provided to challenge the player. Moreover, the learner must be appealed by a sense of curiosity. This can be achieved by employing activities involving content or problems of inherent interest to the learner. Last, an authentic context or environment is required. This stresses to the learner how the instruction can be utilised. A combination of both Lepper's (1988) and Malone's (1981) theories for motivation will be used in this research.

To assess the motivation of learners, the questionnaire on current motivation (QCM) can be used (Vollmeyer & Rheinberg, 2006). This questionnaire assess motivation through four factors: *probability of success*, *anxiety*, *interest* and *challenge*. *Probability of* success is further discussed as self-efficacy, as this is a more precisely specified version of probability of success (Vollmeyer & Rheinberg, 2006). All factors are further discussed below.

Self-efficacy

Furthermore, gamification affects the learning outcome through *self-efficacy* (Bechkoff, 2019). Self-efficacy is defined as one's belief in their capacity to succeed in specific situations to accomplish a task (Bandura, 1977; Vollmeyer & Rheinberg, 2006). It shows a person's interpretation and confidence in their abilities. Bandura (1997) has shown that a higher self-efficacy affects the learning outcome positively. Therefore, the gamification implemented ought to positively affect the perceived self-efficacy. A player with high self-efficacy is competitive because of a perceived higher chance of success.

In contrast, a player with low self-efficacy is likely to avoid all gameplay due to fear of failure (Bandura, 1997; Landers et al., 2015). Punishments tend to decrease self-efficacy (Bandura, 1997) and self-efficacy should not be decreased. That would demotivate the player and lower the learning outcome. To motivate the player and therefore increase the learning outcome, they must experience high self-efficacy. Thus, to positively impact self-efficacy through operant conditioning, punishments should be avoided (Landers et al., 2015) Gamification could ensure this by utilising a scoreboard or awarded points (Banfield & Wilkerson, 2014), because the scoreboard creates challenge (Lepper, 1988; Malone, 1981). This can ultimately increase self-efficacy and, therefore, the learning outcome (Banfield & Wilkerson, 2014).

Anxiety

People also experience *anxiety* (Vollmeyer & Rheinberg, 2006). This can be interpreted as a fear of failure in a specific situation. Anxiety is, however, not the opposite of a learner with a high self-efficacy. A learner with high anxiety can still have a high self-efficacy, as they do not want to fail, but still expect to succeed, thus having a high self-efficacy (Vollmeyer & Rheinberg, 2006). It is important to minimise the anxiety perceived by learners, as this has a negative effect on the learning outcome (Kapp, 2012).

Interest

Gamification adds an enjoyable dimension to learning (Kapp, 2012). The enjoyable dimension is further stimulated by the *interest* of the learner (Vollmeyer & Rheinberg, 2006). The topic of the learning material is important to spark interest by the learner. This is closely tied to the intrinsic motivation perceived by the learner (Kapp, 2012; Vollmeyer & Rheinberg, 2006). If the learner is interested in the topic, that creates a positive effect on the learning outcome.

Challenge

The last factor mentioned by Vollmeyer & Rheinberg (2006) is challenge. This factor assess the extent to which learners feel the situation is an achievement if they succeed in it. It is comparable to the challenge mentioned by Lepper (1988). Including challenge ultimately increases the intrinsic motivation.

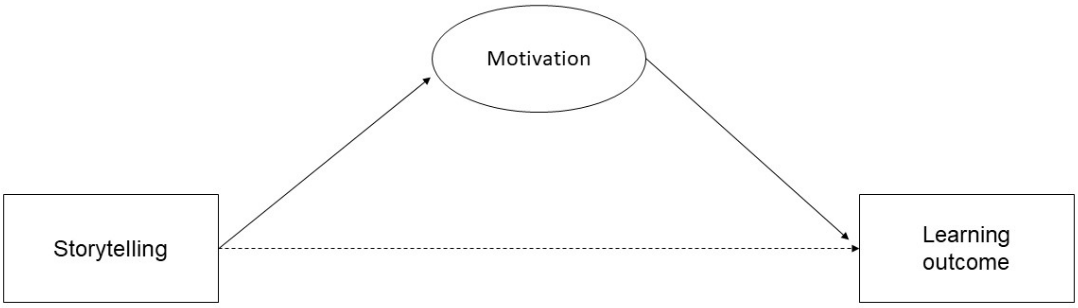
Research questions

To summarise, the proposed research aims to find answers to the extent to which gamification, and in particular storytelling, affects the learning outcome. Already conducted research (e.g. Kendall & Kendall, 2017) does not provide answers to all contexts in which storytelling can be used. Therefore, the proposed research aims to fill in part of this gap and research the effects of storytelling on the learning outcome in physics education. It is expected that the transfer of information within physics education is similar to that of other beta subjects (Silva et al., 2019). Thus, the questions aimed to answer are: To what extent does storytelling contribute to the learning outcome of secondary education students in physics education? And: To what extend does storytelling contribute to the motivation perceived by the secondary education students in physics education? The framework mentioned above will be used to conduct the research and answer the research questions. Motivation, through the four factors self-efficacy, anxiety, interest and challenge will be examined for their effects on the learning outcome, and together these effects can explain the effects of storytelling on the learning outcome.

Research design and methods

The current research had a quantitative, between-subjects design, and included one experiment. The experiment was held amongst two groups, an experimental and a control group. The experimental group was exposed to storytelling, whereas the control group was not. The learning outcome was measured as a dependent variable whilst taking motivation into account. The independent variable in this research was the exposure to storytelling. A graphic depiction of the assumed model is shown in figure 1. A post-test was used to determine the learning outcome and a second post-test was used to measure the mediators. The two tests together have answered the research questions. Both tests were performed using Qualtrics (Qualtrics, 2022). It was assumed that the participants had little to no prior knowledge of the physics topic. In the Dutch system, this topic would have been taught in the science track, a track which typically psychology students were not likely to have chosen. Therefore, they would all participate with the same prior knowledge. Moreover, the material contains a short lesson. Including a pre-test on this material would provide focus for the participants on what to study to gain the highest measured learning outcome. Without a pre-test, participants are not provided with the organisation advantage in advance.

Figure 1



The assumed model.

Respondents

In this research, convenience opportunity sampling was used. The participants who took part were psychology students from the University of Twente. Participants were selected through the SONA platform. SONA is a platform used by the University of Twente for Psychology student to sign up to take part in research. Students were awarded credit scores for their participation. This type of sampling was deemed appropriate for this research. Psychology students are not expected to have a background in physics. Therefore, it was expected that no differences in prior knowledge exist amongst participants. Furthermore, a part of the sample consisted of another convenience sample within the network of the researcher, consisting of Educational Sciences students from Utrecht University. The same assumption was made as with the Psychology students, expecting no difference in prior knowledge. All participants were randomly assigned to a condition, based on their time of signing up.

As Malone (1981) suggested, there are ways to create an intrinsically motivating instruction, even when the user does not find the topic intrinsically motivating. Therefore, it was assumed that a convenience sample would suffice for the current research. A sample size of 53 participants was reached (n = 53), 28 in the experimental and 25 in the control, as this would result in sufficient power (Tomczak et al., 2014). 52,8% were female, with an average age of M_{age} = 22.36 (SD = 2.89).

Instrumentation

This research used an adapted version of Pulsar's already existing physics lesson (te Brinke et al., 2019). This is a method widely used in the Netherlands to teach physics to secondary education learners. The lesson used is about electrical circuits. It teaches how to read circuits and do calculations with the information provided. The lesson was adapted to include storytelling for the experimental group and exclude any storytelling for the control group. Moreover, the lesson was adapted to fit inside the online environment of EdApp. EdApp is an online learning environment and is typically used to play SCORM content. SCORM content is a file format used to save designed content, often used by learning developers. EdApp could, however, also be used in the current context. It is user-friendly and free to use for students. EdApp saves data from the participants and tracks how long they spend on each part of the environment. EdApp provided all participants with a unique ID to distinguish participants, whilst also guaranteeing their anonymity.

The dependent variable learning outcome was measured at a ordinal measurement level. Students were asked several multiple choice questions which have correct answers. Students needed to find the correct answer and were awarded points based on their performance. If the answer was incorrect, no points were awarded. A questionnaire made in Qualtrics, implemented in the EdApp environment, was used as the post-test to determine the learning outcome. The questions asked were adapted versions of the Pulsar method (te Brinke et al., 2019). This method is widely used in secondary education in the Netherlands and therefore the questions were deemed of sufficient quality.

The different aspects of motivation (Vollmeyer & Rheinberg, 2006) were measured using separate pre- and post-tests in Qualtrics. All are measured at an interval measurement level. An adaptation of an existing questionnaire by Vollmeyer & Rheinberg (2006) was used. The questionnaire contained 18 items to measure the four constructs (Appendix C). All questions were rated on a seven-point Likert scale, ranging from "Totally disagree" to "Totally agree". This scale is considered suitable for the context in which attitudes are measured (Allen & Seaman, 2007). The questionnaire has shown to have high reliability and validity (Vollmeyer & Rheinberg, 2006). Therefore, it was assumed that this questionnaire is of sufficient quality. Some questions were adapted to fit the current context, in accordance to the suggestions made by Vollmeyer & Rheinberg (2006).

Procedure

All participants took part in the research on a laptop or PC. They were guided to the EdApp environment through an URL accessible within SONA. Participants first had to fill out an informed consent form (see Appendix A) containing basic information about the study and where to file any complaints. All participants had to agree to this form to take part in the research. If participants disagreed to this, they were redirected to a separate page, to thank the participant for their efforts, but that they could not participate. No data was stored of participants who disagreed.

Participants then continued with the introduction and instruction on using the environment. Next, participants were asked to fill out the pre-test to determine their motivation. The environment of the experimental group contained storytelling, and the control group's did not. Within the environment, instructions and exercises were presented (see Appendix D and E) Participants received 30 minutes to study the material and work through the environment. The post-tests followed this up for both the dependent variable and the mediators. The tests were conducted inside Qualtrics. After participants had filled out the forms, the experiment was concluded.

A request was made to the ethics committee of the University of Twente to conduct the research within the sample. No ethical concerns were addressed by the committee. Participants agreed to the informed consent, eliminating any involuntary participation.

Data analysis

The proposed research yielded fully anonymised, quantitative data. All participants produced data on their learning outcomes and the pre- and post-test questionnaire. The post-test questionnaire on learning outcome resulted in either a correct or incorrect answer. This questionnaire produced results for each participant on an interval scale. The pre-test and post-test questionnaire to assess participants' motivation was rated on a seven-point Likert scale, producing results on an interval scale. A high average score on each separate factor resulted in a high factor. This meant a high average score on self-efficacy meant the participant rated their self-efficacy high.

Two questions were recoded to fit the model. These questions were reversed, meaning a high score on self-efficacy meant their self-efficacy was low. To fit the rest of the items, item 3 and 14 were recoded (Appendix C).

Data was analysed using SPSS (version 28.0.1). A paired samples *t*-test was used to compare the average scores on the pre- and post-test between and within the two conditions. An independent samples *t*-test was used to test the difference in learning outcome between the experimental and control group. The mean and standard deviation were calculated for the learning outcome and the questionnaire. Separate independent samples *t*-tests were used to compare the score on each separate question in the test, to distinguish what questions could be considered easier or more difficult. Also, questions are formulated on different levels of Bloom's taxonomy (Bloom et al., 1956), and therefore it is looked into each question separately to look for distinguishes.

Results

To compare the learning outcome of the participants, an independent samples *t*-test was used to compare the sum score on the test by participants in the control group (n = 25) to the average score on the test by participants in the experimental condition (n = 28). Neither Shapiro-Wilk statistic was significant, indicating that the assumption of normality was not violated. Levene's test was also non-significant, thus equal variances can be assumed. Participants in the control group on average scored M = 7.00, SD = 2.14 and participants in the experimental group on average scored M = 7.61, SD = 2.30. The *t*-test showed this difference was not statistically significant, $t(52) = 1.01$, $p = .32$, two-tailed, $d = 0.27$.

Separate independent samples *t*-tests were used to compare the score on each separate question in the test by participants in the control group (n = 25) to the score on each separate question in the test by participants in the experimental group (n = 28). No Shapiro-Wilk statistic was significant, indicating that the assumptions of normality were not violated. Levene's tests were also non-significant, thus equal variances were assumed. On average, 52% of participants in the control condition scored correctly on test question 6 and on average, 82% of participants in the experimental condition scored correctly on test question 6. The *t*-test showed this difference was statistically significant, $t(51) = 2.43$ $p = .02$, two-tailed, $d = .67$. Moreover, on average, 84% of participants in the control condition scored correctly on test question 8, and on average, 46% of participants in the experimental condition scored correctly on test question 8. The *t*-test showed this difference was statistically significant, $t(51) = 3.04$, $p < .01$, two-tailed, $d = .84$. See Appendix B for the specific test questions.

A paired samples *t*-test with an α of .05 was used to compare mean scores on each factor of motivation to participants in either the control group or the experimental group. Only the significant differences are reported below. Only significant differences were found in the control condition, none in the experimental condition. On average, participants scored 0.62 points, 95% CI [0.12, 1.12], higher on self-efficacy in the control condition on their pre-test than on their post-test. This difference was statistically significant, $t(24) = 2.57$, $p = .02$, and medium, $d = 0.51$. On average, participants scored 0.86 points, 95% CI [0.41, 1.30], higher on interest in the control condition on their pre-test than on their post-test. This difference was statistically significant, $t(24) = 3.97$, $p < .001$, and medium to large, $d = 0.79$.

It was concluded that the assumptions of normality and normality of difference scores were not violated after outputting and visually inspecting the relevant histograms for all analyses.

Furthermore, separate independent samples *t*-test were used to compare the average scores on each factor of motivation in the test by participants in the control group ($n = 25$) to the score on each separate question in the test by participants in the experimental group ($n = 28$). No Shapiro-Wilk statistic was significant, indicating that the assumptions of normality were not violated. Levene's tests were also non-significant, thus equal variances were assumed. The results are depicted in table 1. See appendix C for the specific questions. On average, participants in the control condition scored significantly lower on self-efficacy in their post-test than on their pre-test, $t(51) = 2.34$, $p = .02$, $d = 0.64$. Moreover, on average, participants in the control condition scored significantly lower on interest in their post-test than on their pre-test, $t(51) = 2.09$, $p = .04$, $d = 0.58$.

Table 1

Average scores per factor of motivation in each condition

Variable	Group	n	M_{pre}	SD_{pre}	M_{post}	SD_{post}
Self-efficacy	Control	25	4.36	0.89	3.74	1.26
	Experimental	28	3.85	0.70	3.91	1.17
Pre-test anxiety	Control	25	2.96	0.94	3.91	1.17
	Experimental	28	3.12	1.12	3.18	1.17
Pre-test interest	Control	25	4.06	1.03	3.20	1.33
	Experimental	28	3.50	0.91	3.46	1.35
Pre-test challenge	Control	25	4.87	0.56	4.74	0.93
	Experimental	28	4.68	0.75	4.44	0.74

Discussion

In the current research, it was examined whether storytelling has a positive effect on the learning outcome. It was aimed to answer the following research questions: To what extent does storytelling contribute to the learning outcome of secondary education students in physics education? And: To what extend does storytelling contribute to the motivation perceived by the secondary education students in physics education? Given the results, no significant difference was found between the control and experimental condition and their corresponding test scores. The data thus suggests that there is no significant difference present in learning outcome when storytelling is used compared to not using storytelling. Given the results, a significant decrease was found within the control condition and their corresponding scores on two aspects of motivation, self-efficacy and interest, in the post-test compared to the pre-test. This significant difference was not present within the experimental condition. This suggests that storytelling contributes to the removal of this effect, making storytelling preferable over no storytelling to maintain the motivation of participants.

It was first hypothesised that incorporating storytelling as a gamification element would increase the learning outcome. With the current results, it may be said that the data do not support this hypothesis, as no statistical significant difference has been found between the learning outcome of the two conditions. However, there appear to be differences in average test scores per question within the two conditions. On average, the experimental group scored higher than the control group on questions 1 through 6 but scored lower, on questions 7 through 11. For two questions, questions 6 and 8, this difference is significant. For question 6, participants in the experimental condition scored significantly higher than those in the control condition. For question 8, participants in the control condition scored significantly higher than those in the experimental condition. This statistical difference in both directions suggests that there is an effect present, a medium effort for question 6 and a large effect for question 8.

These findings might be explained by the taxonomy by Bloom et al., (1956). This taxonomy consists of six levels to distinguish the level of thinking necessary for educational objectives (Bloom et al., 1956). It can be argued that questions 1 through 6 require a higher level of thinking than questions 7 through 11. Questions 1 through 6 require at least *understanding*, whilst questions 7 through 11 require at most *remembering* (see appendix B). With there being one significant result amongst both groups, this might suggest that storytelling is more effective when used with questions that require a higher level of thinking than *understanding*. This is in line with the theoretical framework, which argues that storytelling creates a compelling story, which in turn could lead to higher-order thinking (Kapp, 2012). However, this effect needs further research to produce conclusive results, as there might exist an effect of storytelling on higher-order thinking skills that is not yet uncovered.

The results of the independent samples *t*-test suggest that their self-efficacy and interest in the pre-test show a significant difference in the control group compared to the experimental group. Participants thus perceived their chances of success and the interest in physics as a topic higher in the control condition than in the experimental condition. To avoid statistical significant difference in the pre-test, more questions ought to be used within a larger sample. The currently used questionnaire contained 4 to 5 questions per topic, and the sample consisted of a total of 53 participants, with 28 in the experimental group and 25 in the control group. Further calculations show that to achieve a power of .8 with the current results, a sample size of 217 is necessary to generalise the results. This further strengthens the claim that a larger sample size is necessary to research this topic thoroughly. Nevertheless, adding more questions per topic might lower the standard deviation, increasing the power with a smaller sample (Pandis, 2016).

Moreover, the current research only encompasses a post-test to determine the learning outcome. A pre-test to determine participants' prior knowledge would have been desirable to strengthen the results. However, a significant difference exists between the self-efficacy perceived in the pre-test among the control and experimental group. This might suggest that, as students in the control group rated their odds of succeeding higher, their prior knowledge of physics was also higher. This, in turn, might explain why there is a lack of statistical results between the two groups on their average learning outcome. Additionally, the lack of a pre-test to determine the prior knowledge might raise the discussion that the learning outcome measured was not the learning outcome. It could merely be a test score and not a knowledge increase. Therefore, it is recommended to include a pre-test to determine the participants' prior knowledge.

Furthermore, the operationalisation of storytelling could be improved in future research. The current research aims at using the definition of storytelling by Kapp (2012). However, the material used in this research could be improved on the interpretation of this definition. The nature of the material used, being a abstract physics lesson about invisible forces and phenomena, makes it harder to incorporate storytelling. Therefore, the current research focuses more on storytelling as a way to provide more context to the presented information (Giakalaras, 2016). This too could provide the relevance and meaning to the experience that storytelling ought to provide (Kapp, 2012). However, the absence of characters, plot, tension and resolution make the operationalisation of storytelling less conveying. These aspects were excluded because of the nature of the topic. They allow less inclusion of these aspects, as the nature of the topic is on an atomic level. If these aspects had been present, perhaps within a different subject, a more distinct result might have been achievable.

For future research, it is desirable to include the recommendations made above. This would strengthen the results and overall provide better generalisable results. To further strengthen the results, some more recommendations are made. It would be advisable to have participants participate in a lab setting with supervision. The current research let participants participate without supervision, making an entry in this research less of a demand from participants. However, this also leaves room for error within the results, as participants could have scrolled through the research instead of studying the material. This makes the current results less reliable. Another way of incorporating the guarantee that participants studied the material is through tracking the time participants took to complete the research and comparing this to a minimum time necessary to study the materials. This minimum would be based on a sample before starting the actual research. This option is less desirable but requires less effort from participants.

Furthermore, the studied material itself could be improved. The material in the current research was adapted from an existing method in the Netherlands (te Brinke et al., 2019). Both the studied material as well as the questions to assess the learning outcome were adapted from this method. This method, however, assumes a specific prior knowledge amongst its users. This prior knowledge was not guaranteed in the current sample, making the effectivity of the method lower than when used with its intended users. The material could be improved by using a topic and questions that require little to no prior knowledge from its users to fully comprehend and understand the topic and questions. One way of accomplishing this could be through the use of the same method (te Brinke et al., 2019), but for a lower grade.

All in all, the current research does not provide evidence that storytelling has a significant effect on the learning outcome of students studying physics. However, as previously mentioned, there are multiple reasons to believe this effect does exist and further research is necessary to fully examine this effect. Multiple recommendations have been made to gain new insights into this field of research and should be taken into careful consideration in future research

For now, it can be said that there is no significant difference between participants who have experienced storytelling in their material and participants who have not experiences storytelling in their material. Thus, until future research has been conducted, there is no reason to assume storytelling leads to a significant increase in learning outcome within a similar sample with the same topic. However, the current research provides a solid foundation for future research with clear recommendations to build on and a lot more is to be discovered about storytelling.

References

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, *40*(7), 64–65. https://doi.org/10.12691/ajnr-1-1-1

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co.

Banfield, J., & Wilkerson, B. (2014). Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. *Contemporary Issues in Education Research (CIER)*, *7*(4), 291–298. https://doi.org/10.19030/cier.v7i4.8843

Bechkoff, J. (2019). Gamification Using a Choose-Your-Own-Adventure Type Platform to Augment Learning and Facilitate Student Engagement in Marketing Education. *Journal for Advancement of Marketing Education*, *27*(1).

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Handbook I: Cognitive domain. *New York: David McKay*.

Giakalaras, M. M. (2016). Gamification and storytelling. *Univ. Aegean*, *8*, 1–7.

Ibarra-Herrera, C. C., Carrizosa, A., Yunes-Rojas, J. A., & Mata-Gómez, M. A. (2019). Design of an app based on gamification and storytelling as a tool for biology courses. *International Journal on Interactive Design and Manufacturing*, *13*(4), 1271–1282. https://doi.org/10.1007/s12008-019-00600-8

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

Kendall, J. E., & Kendall, K. E. (2017). Enhancing online executive education using storytelling: An approach to strengthening online social presence. *Decision Sciences Journal of Innovative Education*, *15*(1), 62–81. https://doi.org/10.1111/dsji.12121

Landers, R. N., Bauer, K. N., Callan, R. C., & Armstrong, M. B. (2015). Psychological theory and the gamification of learning. In *Gamification in education and business* (pp. 165–186). Springer.

Lepper, M. R. (1988). Motivational considerations in the study of instruction. *Cognition and Instruction*, *5*(4), 289–309. https://doi.org/10.1207/s1532690xci0504_3

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369. https://doi.org/10.1016/S0364-0213(81)80017-1

Pandis, N. (2016). The chi-square test. *American Journal of Orthodontics and Dentofacial Orthopedics*, *150*(5), 898–899. https://doi.org/10.1016/j.ajodo.2016.08.009

Qualtrics. (2022). *Qualtrics*. https://www.qualtrics.com

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68. https://doi.org/10.1037/0003-066X.55.1.68

Silva, J. B. da, Sales, G. L., & Castro, J. B. de. (2019). Gamificação como estratégia de aprendizagem ativa no ensino de Física. *Revista Brasileira de Ensino de Física*, *41*. https://doi.org/10.1590/1806-9126-rbef-2018-0309

Skinner, B. F. (1937). Two types of conditioned reflex: A reply to Konorski and Miller. *The Journal of General Psychology*, *16*(1), 272–279.

te Brinke, L., van den Broeck, T., Buil, S., Hoekzema, D., de Jong, G., Massolt, J., van der Veen, R., Verbeek, J., & Zuurbier, P. (2019). *Natuurkunde 4 VWO Leerboek* (3rd ed.). Noordhoff Uitgevers bv.

Tomczak, M., Tomczak, E., Kleka, P., & Lew, R. (2014). Using power analysis to estimate appropriate sample size. *Trends in Sport Sciences*, *21*(4), 195–206.

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

Appendix A

Informed Consent Form

Taking part in the study

I have read and understood the study information dated [10/06/2022], or it has been read to me. I have been able to ask questions about the study through e-mail to d.s.stern@student.utwente.nl and my questions have been answered to my satisfaction.

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. I am aware I can do this through the closing of the environment I have download this form from.

I understand that taking part in the study involves two online questionnaires and an online test about the material I have studied during the study. The total time investment is approximately 30 minutes.

Use of the information in the study

I understand that the information I provide will be used solely for the writing of a master's thesis. The data collected will not be used for other purposes.

I understand that any personal information collected about me that can identify me, such as my name or where I live, will not be shared beyond the study team. I am aware no such information is asked for, but if provided, will not be shared.

Future use and reuse of the information by others

I give permission for the data that I provide (participant-ID, age, gender, the results to both questionnaires and the answers provided to the test) to be archived in Storytelling & Learning Outcome, so it can be used for future research and learning. If any data is shared, the data will be completely anonymously.

Study contact details for further information:

Duncan Stern, d.s.stern@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities &

Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl

Appendix B

Test questions

When you comb your hair, hairs sometimes get attracted to the comb.

1. What happens to the comb and to the hairs?
    a. Nothing.
    b. The comb gets charged.
    c. The hairs get charged.
    d. **Both get charged.**
2. If you hold the comb so close to your hair that only a few hairs touch the comb, you will see the hairs shoot away from the comb. Explain why.
    a. **The charge from the comb first jumps to the single hairs, which then get the same charge as the comb. And equal charges repel each other.**
    b. The charge from the single hairs jump to the comb, and both get charged. And equal charges repel each other.
    c. The charge from the air jumps to the hair, which then get the same charge as the hairs. And equal charges repel each other.
    d. The charge from the air jumps to the comb, which then get the same charge as the hairs. And equal charges repel each other.
3. If you then hold the comb in your hair for a while, the attraction is gone again. Explain why.
    a. The comb receives charge from the hairs, making the comb more negative than the hairs. The repelling then disappears.
    b. The hairs receive charge from the comb, making the hairs more negative than the comb. The repelling then disappears.
    c. **The charge from the comb bounces back to the hairs, making both neutral. The repelling then disappears.**
    d. The charge from the hairs bounce back to the comb, making both positive. The repelling then disappears.

4. Explain why an atom becomes negatively charged when an electron is added.
    a. Electrons are positive. You thus get a negative charge within the atom.
    b. **Electrons are negative. You thus get a negative charge within the atom.**
    c. The electron forms a neutron. This creates a negative charge within the atom.
    d. The electron forms a proton. This creates a negative charge within the atom.
5. Explain what happens when an electron leaves an atom.
    a. You get a negative atom.
    b. **You get a positive atom.**
    c. You get a neutral atom.
    d. This has no effect on the atom.
6. Does the positive pole of a battery have an excess or shortage of electrons? How do you know that?
    a. It has an excess of electrons. Because electrons are positively charged and the pole is positively charged.

   b. It has an excess of electrons. Because electrons are negatively charged and the pole is positively charged.
   c. It has a shortage of electrons. Because electrons are positively charged and the pole is positively charged.
   d. **It has a shortage of electrons. Because electrons are negatively charged and the pole is positively charged.**
7. The unit for the electric charge is …
   a. Amperage
   b. **Coulomb**
   c. Volts
   d. Joules
8. The unit for the electric current is …
   a. **Amperage**
   b. Coulomb
   c. Volts
   d. Joules
9. The unit for electric voltage is …
   a. Amperage
   b. Coulomb
   c. **Volts**
   d. Joules
10. The formula for electric voltage is …
   a. $U = Q / t$
   b. **$U = \Delta E/Q$**
   c. $U = J / s$
   d. $U = \Delta E + Q$
11. The formula for electric current is …
   a. **$I = Q / t$**
   b. $I = J / s$
   c. $I = C / s$
   d. $I = Q / \Delta E$

Appendix C

Questionnaire questions

1. I like physics. (I)
2. I think I am up to the difficulty of this task. (P)
3. I probably won't manage to do this task. (P-)
4. While doing this task I will enjoy playing the role of a physicist who is discovering relationships between things. (I)
5. I feel under pressure to do this task well. (A)
6. This task is a real challenge for me. (C)
7. After having read the instruction, the task seems to be very interesting to me. (I)
8. I am eager to see how I will perform in the task. (C)
9. I'm afraid I will make a fool out of myself. (A)
10. I'm really going to try as hard as I can on this task. (C)
11. For tasks like this I don't need a reward, they are lots of fun anyhow. (I)
12. It would be embarrassing to fail at this task. (A)
13. I think everyone could do well on this task. (P)
14. I think I won't do well at the task. (P-)
15. If I can do this task, I will feel proud of myself. (C)
16. When I think about the task, I feel somewhat concerned. (A)
17. I would work on this task even in my free time. (I)
18. I feel petrified by the demands of this task. (A)

(C): Challenge

(I): Interest

(P): Probability of success (Self-efficacy)

(A): Anxiety

(P-): Recoded Probability of success (Self-efficacy)

Appendix D

Screenshots of the experimental condition's environment

Electric charge

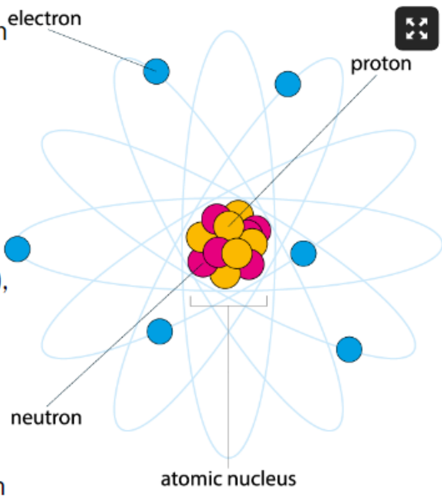
When you wipe a DVD, it will start attracting dust. A rubbed balloon can hang from the ceiling, and a rubbed plastic rod can bend water. **Electric charge** causes these phenomena.

Through friction, objects can charge electrically. Ancient Greeks already knew about this phenomenon. A piece of amber rubbed against a rough fur would attract hairs and dust.

Hanging two pieces of electrically charged amber next to each other would repel one another. Rubbed glass spheres also repel one another. However, a rubbed glass sphere and a rubbed piece of amber attract.

Atoms

In 1897, an English physicist named Sir Joseph John Thomson (1856-1940) discovered that atoms contain very tiny negatively charged parts. He called these parts electrons.



However, atoms are neutral. This means that the atom must also contain a positive charge. With the discovery of the electron, other physicists started exploring what an atom should thus look like from the inside.

Eventually, a British chemist, Ernest Rutherford (1871-1937), discovered in 1911 that atoms consist primarily of empty space. The centre of an atom contains a positive core. Electrons are placed outside this core.

Later, researchers discovered that the core, also called the atomic nucleus, included positive and neutral parts called protons and neutrons. Take a look at the figure to see what an atom looks like.

Static electricity in practice

If you walk on carpet or pet the cat, you can also become statically charged. You are under electrical tension.

You only notice this when you touch the door handle, for example. A spark will jump and you will feel a painful shock.

The static charge then flows away to the earth in a short time; you discharge.

Appendix E

Screenshots of the control condition's environment

Electric charge

Two rubbed objects can repel or attract each other. Their electric charge causes this.

The charge can either be positive or negative.

Equal electric charges repel each other.

Opposite charges attract each other.

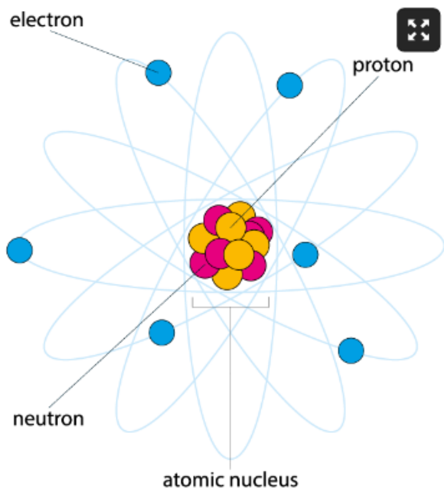
Atoms

An atom consists of a positive core, surrounded by negative electrons.

The core consists of both protons and neutrons.

As a whole, an atom is neutral.

It contains equal amounts of positively and negatively charged particles.



Static electricity

A statically charged object has an excess or deficiency of negative charge.

It receives more electrons. It is thus under tension.

The charge only flows away to earth when you yourself touch the object. This is called discharging.