*Corresponding Author's Email: koepke@hs-mittweida.de
Proceedings of the International Conference on Teaching, Learning and Education
Vol. 2, Issue. 1, 2025, pp. 13-25
DOI: https://doi.org/10.33422/ictle.v2i1.1111

ISSN: 3030-122X online

Evaluating a Gamified Learning App: Usability, User Experience and Learning in Higher Education

Josefine Welk*, Christian Roschke, Dominik Breck, Max Schlosser, Marc Ritter
University of Applied Sciences Mittweida

Abstract

A variety of methods can be used to increase student motivation, for example different types of gamification strategies like leaderboards, badges or experience points. Among such methods, the StudiSQ app incorporates various gamification elements. In the app, students can independently repeat content from courses in the form of quizzes. Gamification elements such as badges, ranking lists and duels are used to increase motivation. Approximately 60 students took part in an evaluation of the app with a focus on usability, user experience and learning effect. The latter was measured using two tests, with some of the participants using StudiSQ to prepare for the second test and others using conventional learning methods. In terms of usability (65.9/100) and user experience (0.61/-3 to 3), results were medium to slightly positive. On average, the test results of all participants deteriorated; students who used StudiSQ performed slightly better (-0.36 vs. -0.63 points). No statistically significant difference was found between the two test groups during the evaluation. Appropriate measures for the design and implementation of follow-up studies were discussed. To promote use by students and teachers, it is also necessary to fix bugs and improve usability. Nevertheless, around 90% of the students would use the app again.

Keywords: Asynchronous learning, Educational Technology, E-Learning, Evaluation, Learning Outcomes

1 Introduction

The use of gamification can have a positive effect on the motivation and learning effect of students. This effect was demonstrated by Alonso-Sánchez et al. (Alonso-Sánchez et al., 2025) as part of their study on the influence of gamification on students' learning, academic performance, cooperation and motivation. The study showed that the use of gamification had a positive impact on all four key areas analysed. The more than 350 students surveyed also reported that they prefer learning using gamification instead of conventional teaching methods. Mazarakis and Bräuer (Mazarakis & Bräuer, 2023) also investigated the influence of gamification on student motivation. As part of their study, over 500 students were asked to

answer multiple-choice questions until they lose motivation. A maximum of 190 questions could be answered. To investigate the influence of gamification on motivation, some of the students answered the questions without the use of gamified elements, while other students used individual elements (feedback, progress bars, badges, narratives) or combinations of these. The study showed a positive influence of gamification on student motivation. The progress bar and badges, in combination with the feedback, motivated students to answer all the questions. Groening and Binnewies (Groening & Binnewies, 2019) were also able to prove the positive influence of a reward system. However, they also identified various factors that influence the positive effect, including the appropriate level of difficulty to achieve the reward and the number of rewards themselves. Kherazi and Bourray (Kherazi & Bourray, 2024) were able to achieve comparable results and, as part of their study, attributed a decisive role to gamification in creating an effective and motivating learning environment.

A more critical view of the gamification approach was published by Sailer et al. (Sailer et al., 2017). In their study, they confirmed their hypothesis that gamification alone does not necessarily have a positive effect on motivation; the targeted integration of game design elements, on the other hand, was recognised as having a psychological effect that can have a positive impact on motivation. Bai et al. (Bai et al., 2020) identified two possible negative effects of gamification in their analysis of 24 quantitative studies: on the one hand, test subjects stated that gamification had no additional benefit, but on the other hand could trigger anxiety and jealousy. At the same time, the negative effects were also contrasted with four positive effects. Gamification can arouse enthusiasm, provide performance-related feedback, satisfy the need for attention and support defining personal goals.

In addition to the influence of gamification on the motivation of students, the influence on learning success also plays an important role. Nurtanto et al. (Nurtanto et al., 2021) demonstrated a positive effect of gamification on learning success through a comprehensive literature review. Sailer and Homner (Sailer & Homner, 2020) also investigated the influence of gamification on cognitive, motivational and behavioral learning outcomes. A significant, small effect was found for all three aspects examined. However, the paper also emphasizes that the factors that enable successful gamification are diverse and unresolved.

In addition to gamified learning applications, mobile applications are also becoming increasingly popular. Especially interesting are the flexible application possibilities, which can be easily integrated into the everyday lives of students as well as into lectures. Pedraja-Rejas et al. (Pedraja-Rejas et al., 2024) were able to demonstrate a positive effect of mobile learning applications on students' learning success and critical thinking skills. Criollo-C et al. (Criollo-C et al., 2021) also confirm the thesis that mobile learning applications can have a positive effect on learning. They emphasise the possibility of integrating numerous learning methods into the application, which allows learning to be individualised.

Although the integration of gamification can have a positive effect on motivation and learning success as described above, this is generally not enough to achieve this effect. It is also essential to consider usability and user experience. Ishaq et al. (Ishaq et al., 2020) examined a learning application with regard to user experience and usability and found serious shortcomings. In their conclusions, they describe that improvements to these two aspects are necessary to realize regular use of the learning application. Mohtar et al. (Mohtar et al., 2023) also formulated that not only the content and its preparation (e.g. through gamification) are decisive for the frequency of use, motivation and ultimately the learning success, but also the usability of the application.

The app StudiSQ which was developed by Schlosser et al. (Schlosser et al., 2023) fits into the context of gamification and mobile learning described above. It incorporates various

gamification elements (feedback, leaderboards, badges, experience points, progress) with the aim of creating a learning environment that promotes and motivates learning. During an evaluation, the effect of gamification on user motivation was examined. The results showed an increase in motivation after using the app. Building on these results, this article focuses on the systematic analysis of the StudiSQ app in terms of its usability, user experience, and contribution to learning success. The aim is to derive well-founded recommendations for further development of the app and thus create the basis for its successful use in higher education teaching. To this end, a multi-stage evaluation procedure was developed, implemented, and evaluated. The method is described in the following section. The sections 3 and 4 list and discuss the results. Section 0 summarises.

2 Methods

The methodology used to evaluate the StudiSQ app is explained below. Therefore, the general conditions are outlined first, followed by a description of the phases of the evaluation and how they were analysed.

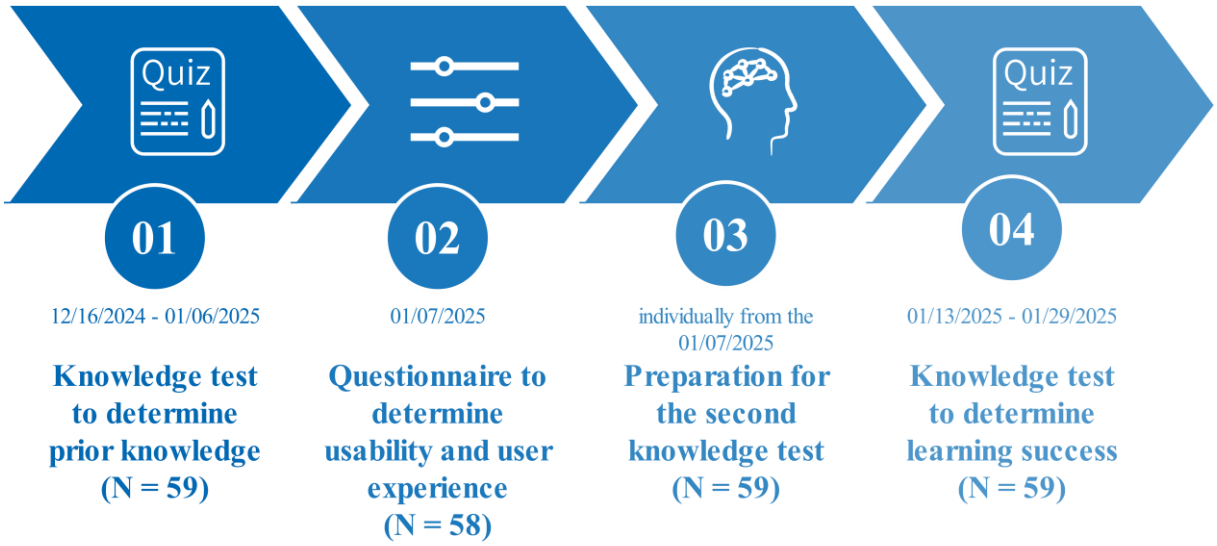
2.1 General conditions

The target group of the evaluation were students of the bachelor's degree program *Media Informatics and Interactive Entertainment*, especially participants of the course *Human Machine Interaction*. In the curriculum, this course is part of the first semester, so the focus of the evaluation was on first-year students; however, students in higher semesters who were repeating the course were also able to take part in the evaluation. The evaluation was carried out over a period of 1.5 months and included both asynchronous phases and a face-to-face phase. A total of 58 students completed the evaluation, although the number of participants varied depending on the phases described in section 2.2.

2.2 Evaluation procedure

The evaluation comprised four different, consecutive phases, which are shown in Figure 1.

Figure 1: Evaluation process consisting of three phases to determine learning success (knowledge test - preparation - knowledge test) and the phase to determine usability and user experience (questionnaire). The implementation period and the number of participants are specified for each phase.



The four phases of the evaluation are presented in detail below.

Determining prior knowledge through the first knowledge test To determine prior knowledge, all students completed a knowledge test focussing on basic programming using the Java programming language. This programming language is used in the course *Introduction to Computer Science* of the bachelor's degree program *Media Informatics and Interactive Entertainment*, which is why all participants have a basic knowledge of the language. The students were given three weeks to complete the test, during which they were able to work flexibly. The test comprised 13 tasks; a maximum of 25 points could be achieved. The types of tasks chosen were selection tasks, assignment tasks, sequence tasks and cloze tasks. Table 1 provides an overview of the topics surveyed and lists the number of tasks in each category, a sample task and the task type.

Table 1: Content of the first knowledge test divided into categories. For each category, the number of questions and a corresponding example task including its task type are listed.

Category	N	Example task	Type of task
Number systems	1	Assign the given numbers to the corresponding number systems.	Assignment task
Variables	2	Sort the following data types for the representation of integers in ascending order according to their value range.	Sequence task
Programming structures	5	Which of the following statements about constructors are true?	Selection task
Memory structure array	3	Given an array with the name array and a length of 10, specify the index with which the last field of the array can be addressed.	Cloze task
Object orientation	2	Select the essential properties of object orientation from the following options.	Selection task

Evaluation of the usability and user experience The usability and user experience of the app StudiSQ were evaluated as part of a lecture of the course *Human Machine Interaction*. Therefore, the students were provided with a questionnaire. The questionnaire included a request for general personal information (age group, gender) and information on the test environment used (type of device, operating system). The questionnaire also contained tasks to be completed, for which the students were asked to rate how intuitively they were able to

solve them (Likert scale from 1 to 5), as well as statements to determine usability based on the System Usability Scale (Brooke, 1995). For each statement, students were asked to indicate the extent to which they agreed with the statement (1 - strongly disagree, 5 - strongly agree). To determine the user experience, the students were given word pairs based on Laugwitz et al. (Laugwitz et al., 2008), whereby the scales novelty and stimulation were omitted. In the last section of the questionnaire, the students answered questions about their overall impression (rating in stars, interest in using the app, general feedback). The students were given 180 minutes to complete the task and the questionnaire as part of the course. In addition, students were asked to note any bugs they encountered while working on the tasks.

Preparation for the second knowledge test Before taking the second knowledge test, the students prepared individually; the minimum preparation time for the test was 6 days, the maximum preparation time was 22 days, limited by the examination in the *Introduction to Computer Science* course. To prepare for the second knowledge test, the students were divided into two groups. For this purpose, a semi-randomised group allocation was carried out by randomly assigning the students to the two groups, provided that two groups of equal strength were formed. In this context, equal strength means that the groups are similar in size on the one hand, and on the other hand also cover the entire performance spectrum of the students. The semi-randomised allocation made it possible to create two groups in which both very high-performing and lower-performing students were represented. Each student was informed individually whether the app StudiSQ could be used to prepare for the second test. Students who were allowed to use StudiSQ were given a quiz on Java programming in the app. The quiz contained 22 questions, whereby the app StudiSQ only provides questions of the types selection question and assignment question. The focus of the content was based on the subject categories listed in Table 1. The test included tasks from the first knowledge test as well as additional tasks that expanded the range of topics and thus enabled broader preparation for the second knowledge test. Students who were not allowed to use the app were asked to prepare themselves using conventional methods without naming them specifically.

Determining learning success through the second knowledge test Similar to the knowledge test carried out in the first phase to assess prior knowledge, a second knowledge test was carried out in the fourth phase to assess learning success. In this phase, students were also able to complete the test flexibly within a one-week period. The thematic focus of the tasks was also on the basics of Java programming language. Like the first test, the second knowledge test also comprised 13 tasks with a total of 25 points. The task types also corresponded to those of the first knowledge test, and the content focus of the tasks remained the same as well; the tasks themselves were partially modified or completely replaced in their answer options. The aim of this procedure was to check the students' acquisition of skills by ruling out the possibility that specific questions in the first knowledge test have been memorized. Table 2 provides an overview of the topics surveyed and lists the number of tasks in each category, a sample task and the task type.

Table 2: Content of the second knowledge test divided into categories. For each category, the number of questions and a corresponding example task including its task type are listed.

Category	N	Example task	Type of task
Number systems	1	Assign the given numbers to the corresponding number systems	Assignment task
Variables	3	How many bytes does a variable of type int in memory?	Cloze task
Programming structures	5	Which statements about the do-while loop are true?	Selection task
Memory structure array	1	Which statements about arrays are true?	Selection task
Object orientation	3	What is the key word for inheritance?	Cloze task

2.3 Assessment of the evaluation

To determine the learning success, the points achieved in the first and second knowledge tests were analysed; a distinction was made regarding the use of the app StudiSQ. Usability was analysed based on the method described by Brooke (Brooke, 1995) for evaluating the System Usability Scale. The values collected to determine the user experience were standardised based on Laugwitz et al. (Laugwitz et al., 2008). Statistical measures were determined for both usability and user experience, and these were compared regarding the operating system used by the students. In addition, the intuitiveness of the tasks and the overall impression of the app StudiSQ were analysed.

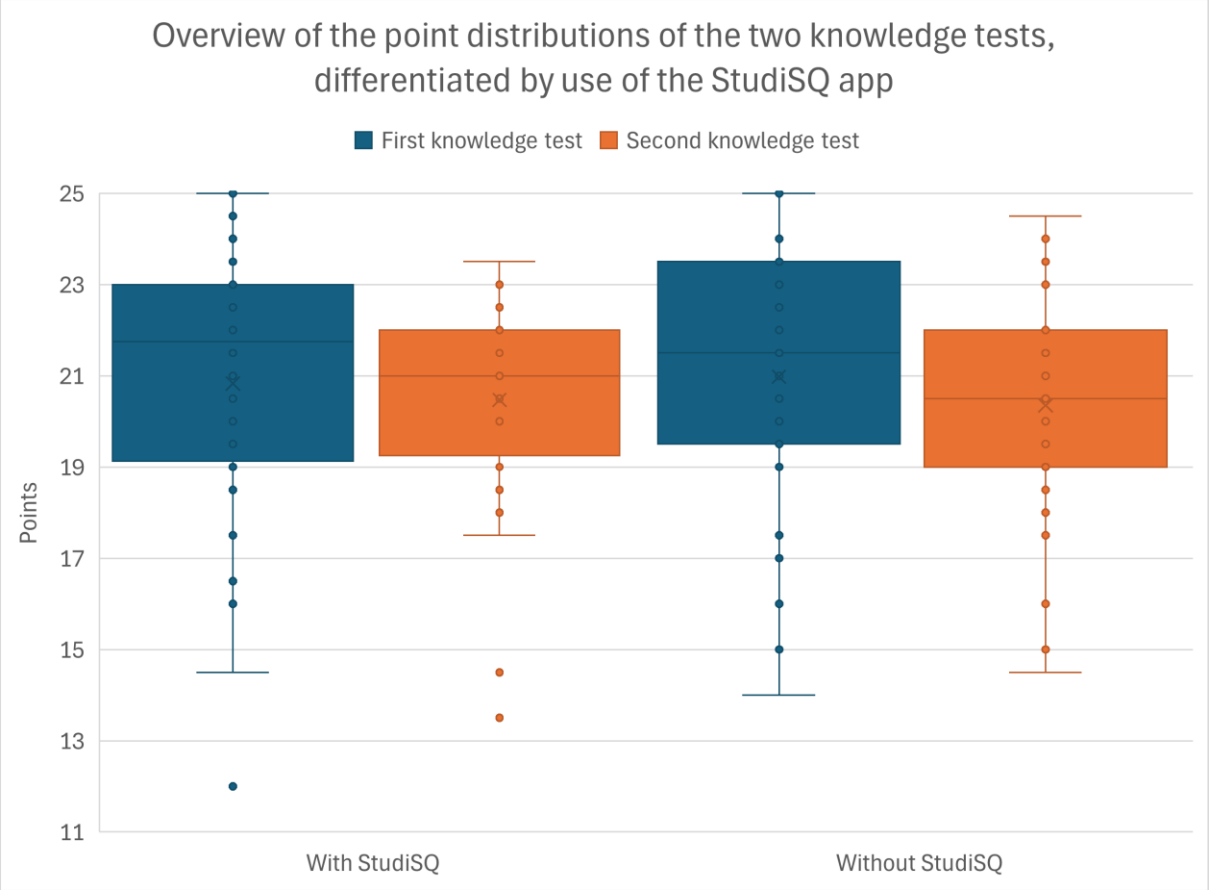
3 Results

The results of the evaluation are presented below. A distinction is made between the results regarding learning success, usability and user experience as well as the general evaluation of the app StudiSQ.

3.1 Learning success

In terms of learning success, all students deteriorated minimally on average (20.9 vs. 20.4 points, $\sigma = 3.04$ or $\sigma = 2.41$). The change in points varied between the extreme values of -10 and 6. Students who used StudiSQ to prepare for the second test deteriorated by -0.36 points ($\sigma = 3.28$) on average. The group that prepared with conventional teaching materials deteriorated by -0.63 points ($\sigma = 3.22$). The extreme values, between which the point changes varied, also differed regarding app use. The score changes of students who used StudiSQ varied between the extreme values of -6.5 and 6, while those of students who did not use the app fluctuated between the extreme values of -10 and 4.5. Figure 2 shows the results of the first and second knowledge test, differentiated according to the use of the StudiSQ app. The spread of the results of the second knowledge test is smaller when using the app than when not using the app. The lower limit of the results using StudiSQ also shifted more strongly upwards compared to the lower limit of the results without using the app.

Figure 2: Results of the two knowledge tests depending on the use of the StudiSQ app



Based on these results, an evaluation was carried out depending on the performance of the students as shown in Table 3. Based on the lower (19.5 points) and upper quartile (23 points) of the first knowledge test, the students were divided into lower performers (points < 19.5), average performers (19.5 ≤ points ≤ 23) and higher performers (points > 23). On average, the lower-performing students improved by 2.77 points (σ = 2.15); students with average performance deteriorated by -0.5 points (σ = 2.37) and high-performing students deteriorated by -3.88 points (σ = 2.68). While average and higher-performing students who used the StudiSQ app deteriorated more on average (-0.58 vs. -0.39 and -4.33 vs. -3.5 respectively), the lower-performing students who used the app scored more points on average (3.13 vs. 2.17).

Table 3: Results of the knowledge tests differentiated according to the performance level of the students and the use of the app StudiSQ

Performance Level	All Students		Students using StudiSQ		Students not using StudiSQ	
	Average Point Difference	Standard deviation	Average Point Difference	Standard deviation	Average Point Difference	Standard deviation
Low	2.77	2.15	3.13	2.12	2.17	1.91
Average	-0.5	2.37	-0.58	2.18	-0.39	2.58
High	-3.88	2.68	-4.33	2.25	-3.5	2.94

To further investigate the results, Cohen's d was calculated to determine the effect size. In addition, the statistical significance of the results was examined using a t-test and the Mann-Whitney U-test, which is more stable in the face of high dispersion, with subsequent determination of the two-sided p-value. Table 4 shows the results of statistical evaluations.

Table 4: Results of the statistical evaluation regarding learning success differentiated according to the use of the StudiSQ app

	N	Average Point Difference	Standard deviation	Cohen's d	t-Test	U-Test
With StudiSQ	32	-0.36	3.28	0.01	0.76	U = 485 z = 0.81 p > .05
Without StudiSQ	27	-0.63	3.22			

3.2 Usability and User Experience

Both the usability and user experience evaluations produced average to slightly positive results. Usability was rated with an average of 65.9 out of a possible 100 points (σ = 18.3). Regarding the operating systems used, it was found that the app StudiSQ was rated slightly better (67/100, σ = 19.0) when used on Android devices (70.7% of test environments), while the app scored slightly below average (63.4/100, σ = 16.0) when used on iOS devices (29.3% of test environments). In terms of user experience, the students rated an average of 0.61 (σ = 0.27). On the underlying scale of -3 to 3, this represents an average to slightly positive rating. Again, differences between Android and iOS devices were identified. Like usability, the app StudiSQ achieved a slightly above-average score of 0.68 (σ = 0.29) when used on Android devices, while the score when used on iOS devices was below the overall average with a score of 0.51 (σ = 0.38). Figure 3 shows the rating of the individual word pairs; the rating of all word pairs is in the positive value range, but the ratings of individual word pairs (e.g. confusing - clear) show only a slight difference to the neutral rating.

A statistical evaluation was also carried out for usability and user experience, which included Cohen's d, the t-test and the Mann-Whitney U-test. Tables 5 and 6 show the results of the statistical calculations for usability and user experience.

3.3 Evaluation of the tasks

On average, the students rated the intuitive solvability of the tasks with 3.45 (σ = 0.61) out of a possible 5 points but there were differences between the tasks to be solved. While the task of viewing their own badges, for example, received an average score of 4.22 (σ = 1.13), the tasks of playing a duel against a friend and viewing their own statistics received the lowest scores of 2.36 (σ = 1.11) and 2.86 (σ = 1.28) points. Table 7 shows a detailed list of the tasks and their respective scores including the associated standard deviations.

Figure 3: Evaluation of the StudiSQ app regarding the user experience based on the word pairs defined by Laugwitz et al. (Laugwitz et al., 2008)

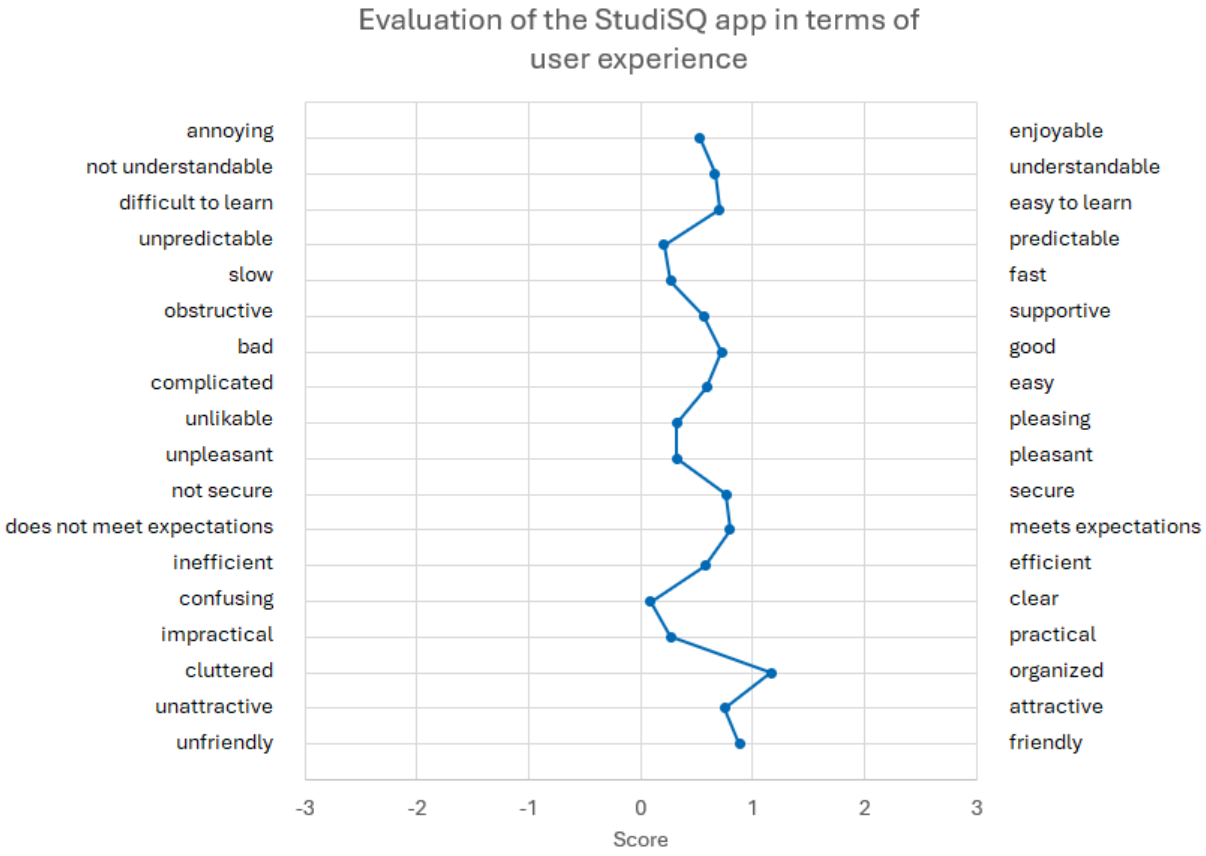


Table 5: Results of the statistical evaluation regarding usability differentiated according to the operating system used

	N	Average Usability	Standard deviation	Cohen's d	t-Test	U-Test
Android	41	67	19.04	0.03	0.24	U = 483 z = 2.3 p < .05
iOS	17	63.4	16.04			

Table 6: Results of the statistical evaluation regarding user experience differentiated according to the operating system used

	N	Average User Experience	Standard deviation	Cohen's d	t-Test	U-Test
Android	41	0.68	0.95	0.03	0.53	U = 441 z = 1.58 p > .05
iOS	17	0.51	0.89			

3.4 Evaluation of the overall impression

Considering the prototype status, the students rated the app an average of 3.29 ($\sigma = 0.83$) out of a possible 5 stars. Positive impressions cited included the simple option of checking knowledge, the range of quiz modes offered, the basic design and clarity of the app, the option of receiving explanations for incorrectly answered questions and the increase in motivation through playing a quiz together in duel mode. Suggestions for improvement made by the students included improving the search function (e.g. by adjusting the search so that you can search by category as well as by quiz), adding a tutorial, placing some functions (e.g.

statistics) more intuitively and generally fixing existing bugs (e.g. missing notifications for duels). Five students would like to see a larger pool of tasks and more variety in the selection of questions. Despite some criticism, 89.7% of respondents stated that they would use the app more often if it offered relevant content for their study program. Students who would not use the app cited the excessive size of individual quizzes, the current bugs, the time limit for answering the questions and the lack of fit between the didactic design and preparation of the learning content in the app and the personal learning habits of the user as reasons.

Table 7: Overview of the tasks to be completed as part of the evaluation, their associated average scores and their standard deviations

Task	Average rating	Standard deviation
Search for quiz and start in practice mode	3.83	0.81
Find quiz via catalogue and start in review mode	3.83	0.95
View statistics	2.86	1.28
Play a duel against a friend	2.36	1.11
Start quiz via QR code	3.22	1.25
Start quiz via room code	3.84	1.01
View badge	4.22	1.13

4 Discussion

The results presented in section 3 tend to show positive developments with regard to learning success, usability and user experience, the completion of the tasks set and the overall impression. The calculated effect sizes according to Cohen's d indicate a small effect in the area of learning success, usability and user experience. However, these effects proved to be statistically insignificant in the significance tests carried out (t-test and Mann-Whitney U test with two-sided p-value); the exception was usability, which showed statistical significance in the Mann-Whitney U test with two-sided p-value. Therefore, the results cannot be generalized and should be interpreted with caution. The results will be discussed in more detail below and approaches for increasing statistical significance will be outlined. Practical applications of the app are highlighted as well. First of all, there is a discussion of possible reasons for the average drop in student performance observed in the second knowledge test.

4.1 Discussion of the drop in performance

There may be various reasons for the drop in performance between the two knowledge tests. A (subjectively) different level of difficulty, the preparation for the second test and the influence of external disturbance variables are discussed below.

Difficulty level of the tasks One possible reason for the average deterioration could be a difference in difficulty between the two tests. Despite the identical level of competence, certain topics may have been subjectively perceived as more difficult in the second knowledge test, which could explain poorer performance. Regarding this aspect, the test results were examined more closely. While 40.7% of the participants deteriorated by more than one point, 23.7% of the participants were able to maintain the level of the first knowledge test (+/-1 point) and 35.6% of the participants improved by more than one point. The low average point deviation (20.9 compared to 20.4) suggests that the level of difficulty was comparable; however, this does not exclude individual difficulties with the second test. In addition to an individually perceived difference in the level of difficulty, inaccurate reading of the tasks may also be a cause of the drop in performance. Some of the tasks were retained and only varied in their possible answers. Due to inaccurate reading of the answer

options, participants may have chosen incorrect solutions that were correct in a modified form in the first knowledge test.

Preparation In addition, the individual preparation of the students may have been very different. Although the lecturer's view of the StudiSQ app shows that the preparation test was taken a total of 72 times, no conclusions can be drawn from this regarding the individual preparation of the students. The extent of preparation for the second knowledge test is not comprehensible, especially for students who did not use the app.

Influence of external factors In addition, individual and external disturbance variables (e.g. daily form, time of the test, environmental conditions) could have influenced the performance. These were not recorded and limit the validity.

4.2 Discussion regarding statistical significance

No statistically significant differences were found between the test groups or the operating systems used, neither in terms of learning success nor in terms of usability and user experience. Regarding learning success, however, a statistically significant difference between the test groups would be necessary to validly substantiate the effectiveness of the app. Possible reasons for the lack of statistical significance and ways of achieving this are discussed below.

Sample size One way to achieve statistical significance in subsequent studies is to increase the sample size. A total of 59 students took part in the knowledge tests, 32 students used StudiSQ to prepare, 27 students used conventional methods to prepare for the second knowledge test. The calculated effect size (Cohen's $d = 0.63$) indicates a medium to strong effect of the intervention. However, a power analysis shows that a sample size of at least 41 students per test group would have been required to statistically confirm such an effect at a significance level of 5% and a test power of 80%. The actual group sizes were below this threshold, which limits the significance of the significance tests.

Framework conditions of the knowledge tests In addition to increasing the sample size, the standardization of the framework conditions for conducting the test should also be considered for future studies. By establishing an examination situation, equivalent conditions can be created that exclude deviations in the students' performance due to unforeseeable influences or at least allow them to be documented.

Framework conditions for individual preparation The framework conditions for preparing for the second test should also be defined more clearly. As part of the evaluation, students were prepared individually with the aim of depicting learning scenarios that were as realistic as possible. The disadvantage of this approach, however, is that it is difficult to assess the actual amount of preparation. Although it can be deduced from the statistical survey of the StudiSQ app that the preparation test was completed 72 times, this cannot be mapped to the individual students, meaning that it is not possible to check how often each student completed the preparation quiz. Furthermore, the chosen preparation period was correspondingly short due to the location of the evaluation at the end of the semester. To be able to assess the effect of the app on learning success, a longer period would probably be necessary in which the course content taught could be repeated and consolidated during the semester using the app and thus the effect on students' learning success could be examined.

4.3 Possible practical applications

The evaluation presented in this paper used a course event but was not actively involved in it. In future evaluations, the app should be used didactically in a course and thus find practical

application over a longer period. A possible concept for the practical integration of the app into a course is presented below.

One possible way to integrate the app is to use it to prepare for and follow up on courses. At the beginning of the courses, students are asked which students used the app to prepare for the course and to what extent they prepared. Similarly, other preparation methods and the amount of time they took are also surveyed. In selected courses that have not been communicated in advance, this survey is supplemented by a test to check current knowledge. With this approach, learning progress can be evaluated, while at the same time the use of the app is tracked, and a standardized test environment is created. In addition to the evaluation context, the StudiSQ app and the quizzes it contains can be used by students for the individual preparation of courses and exams.

As part of a further evaluation, a closer look should also be taken at the subgroup of lower-performing students. The results of the knowledge test shown in Figure 2 and Table 3 which were presented in Section 3.1 show a smaller spread of points achieved when using the app, a stronger upward shift of the lower limit as well as a greater increase in points for lower-performing students who used the StudiSQ app for preparation. Based on these observations, the thesis can be formulated that lower-performing students can benefit from using the app. Due to the small number of participants, this hypothesis cannot be refuted or confirmed within the scope of this study and should be evaluated in follow-up studies.

5 Conclusion

In this paper, an evaluation of the app StudiSQ was presented with a focus on learning success, usability and user experience. A total of 58 students from the *Media Informatics and Interactive Entertainment* bachelor's degree program took part in the evaluation. Although the evaluation of usability and user experience was slightly positive and the students who used StudiSQ showed less poor results in the second knowledge test, no statistically significant results were obtained. During the discussion, reasons were discussed and adjustments for future studies were derived. Based on the findings, a further evaluation should be carried out with a focus on the learning success of the students, which creates uniform framework conditions for the knowledge test, makes stronger reference to the frequency and intensity of use of the app StudiSQ, increases the number of samples and ensures the use of the app over a longer period of time and is didactically anchored in a course. Such a practical application would also make it possible to evaluate the app from the lecturers' perspective. In addition, the app should be revised regarding the functionalities criticized during the evaluation and bugs documented by the students should be fixed. Despite the criticism, around 90% of students stated that they would actively use the app when content relevant to their studies is provided and the underlying concept was praised.

References

Alonso-Sánchez, J. A., Alonso, J. L. N., & Santana-Monagas, E. (2025). Gamification in Higher Education: A Case Study in Educational Sciences. *TechTrends*. https://doi.org/10.1007/s11528-025-01056-2

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, 100322. https://doi.org/10.1016/j.edurev.2020.100322

Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind., 189*.

Criollo-C, S., Guerrero-Arias, A., Jaramillo-Alcázar, Á., & Luján-Mora, S. (2021). Mobile Learning Technologies for Education: Benefits and Pending Issues. *Applied Sciences*, *11*(9), Article 9. https://doi.org/10.3390/app11094111

Groening, C., & Binnewies, C. (2019). "Achievement unlocked!"—The impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior*, *97*, 151–166. https://doi.org/10.1016/j.chb.2019.02.026

Ishaq, K., Rosdi, F., Azan, N., & Abid, A. (2020). Usability and Design Issues of Mobile Assisted Language Learning Application. *International Journal of Advanced Computer Science and Applications*, *11*(6). https://doi.org/10.14569/IJACSA.2020.0110611

Kherazi, A., & Bourray, M. (2024). Gamification and m-learning: An innovative approach to sustainable language learning. *E3S Web of Conferences*, *477*, 00066. https://doi.org/10.1051/e3sconf/202447700066

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (pp. 63–76). Springer. https://doi.org/10.1007/978-3-540-89350-9_6

Mazarakis, A., & Bräuer, P. (2023). Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human–Computer Interaction*, *39*(3), 612–627. https://doi.org/10.1080/10447318.2022.2041909

Mohtar, S., Jomhari, N., Omar, N. A., Mustafa, M. B. P., & Yusoff, Z. M. (2023). The usability evaluation on mobile learning apps with gamification for middle-aged women. *Education and Information Technologies*, *28*(1), 1189–1210. https://doi.org/10.1007/s10639-022-11232-z

Nurtanto, M., Kholifah, N., Ahdhianto, E., Samsudin, A., & Isnantyo, F. D. (2021). A Review of Gamification Impact on Student Behavioural and Learning Outcomes. *International Journal of Interactive Mobile Technologies (iJIM)*, *15*(21), 22. https://doi.org/10.3991/ijim.v15i21.24381

Pedraja-Rejas, L., Muñoz-Fritis, C., Rodríguez-Ponce, E., & Laroze, D. (2024). Mobile Learning and Its Effect on Learning Outcomes and Critical Thinking: A Systematic Review. *Applied Sciences*, *14*(19), Article 19. https://doi.org/10.3390/app14199105

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

Sailer, M., & Homner, L. (2020). The Gamification of Learning: A Meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Schlosser, M., Urban, M. W., Tran, T. D., Nobre, C., Kreißig, P., Labude, S., Breck, D., Roschke, C., Ritter, M., & Vodel, M. (2023). Evaluation of a Learning Application Designed to Improve Motivation and Engagement of Students in Higher Education. *International Journal of Intelligent Computing Research*, *14*, 1181–1186. https://doi.org/10.20533/ijicr.2042.4655.2023.0145