Perceived Usability of Educational Chemistry Game Gathered via CSUQ Usability Testing in Indonesian High School Students

Herman Tolle [1], Muhammad Hafis[2]
Ahmad Afif Supianto[3]
Faculty of Computer Science
Brawijaya University, Malang, East Java, Indonesia

Kohei Arai[4]
Graduate School of Science and Engineering
Saga University
Saga, Japan

Abstract—Educational game is now a commonplace among students and teachers alike. Recent researches show that studies regarding educational game general effectiveness in the learning environment are nothing new. However, usability studies in the educational game are rather rare compared to general non-game-related usability studies. This research synthesizes the result obtained from the Computer System Usability Questionnaire (CSUQ) and separated between multiple students pre-existing grouping such as genders, prior knowledge, as well as experimental treatment setup such as materials given before the game session. The metrics are tested in an Indonesian high school by using an educational game of chemistry regarding the topic of reaction rate with a total of 53 participants. General results show that there exist many differences of perceived usability aspects between male and female students, the existence of learning materials given before the game session, as well as the existence of students' prior knowledge. Overall, the main findings of this research show that usability in the educational game is affected by gender, materials existence, and previous knowledge existence.

Keywords—*Usability testing; CSUQ; educational game; male students; female students*

I. INTRODUCTION

In recent years, the digital educational game has emerged as one of the more sophisticated methods to augment the student learning process. With the increasing ease of access to technology [1], students are more exposed to computers and smartphones. They use it more than ever as the digital educational game approach can improve students' motivation; as such, the field of technology-enhanced learning is now getting more important than ever [2]. Some clear advantage of educational games is how the students perceive it to be useful for their learning experience. Recent studies show that educational game is perceived to be able to increase students' enjoyment during the learning process [3] as well as promoting skill and knowledge gain [4]. Digital educational games also offer an advantage in terms of enriched visuals as well as more appealing multimedia aspects [5]. In terms of subjects, multiple domains of educational topics have been adapted in the form of digital educational games such as art [6], language learning [7], and mathematics [8]. The current state-of-the-art of digital educational game shows that it is an emerging approach to supplement conventional generally-used instruction-based approach.

Generally, the essential aspect of a digital educational game is whether the game can enable students' knowledge acquisition. This is generally done by evaluating students' performance in a quasi-experimental setup and aimed to evaluate the effectiveness of the learning environment [9]. However, the way students as end-users interact with the game itself also plays a significant role in ensuring optimal knowledge acquisition process [10], as usability is positively strongly correlated with increased learning motivation [11]. Similar to other software, digital educational games also require proper quality assurance in terms of usability.

Usability is a broad term defined as "user-friendliness" of software and quality that attributes the ease-of-access of an interface in a software [12]. Usability in the digital educational game, however, focuses on ensuring students learn effectively and efficiently as well as maintaining students' interest in the game itself [13]. In terms of how general usability is being measured, there exist several approaches to acquire different information regarding usabilities, such as observational technique [10] and think-aloud technique [14]. Different metrics to quantify a different aspect of usability also exist and is used to quantify a different aspect of perceived usability such as System Usability Scale (SUS) [15] Usability Metric for User Experience (UMUX) [16], and Computer System Usability Questionnaire (CSUQ) [17]. Each metrics has a different purpose and assesses a different aspect of usability.

Recent existing research about usability categorizes the test subjects based on multiple classifications. Game-dependent skill-based classification usability has been done to detect whether there exists any difference in perceived usability between the classes [18]. A more general gender-based classification for usability testing has also been done before [13]. Regardless, the existing researches separate the usability criteria differently.

Specifically speaking, existing research focuses on using a particular metric for evaluation purposes [18]; however, in-depth research regarding each aspect and category of usability in a particular metric is also needed. A synthesis of information based on usability scores can gather critical aspects of the users when viewed from different demography

and classifications [19]. An in-depth usability study is able to analyze users' satisfaction and create a recommendation for system improvement in the future [20]. Moreover, an existing in-depth study is done for general software and systems, but not for digital educational games. The urgency of an in-depth usability analysis from the digital educational game perspective is needed since digital educational games, and general software is vastly different. The pedagogical aspect and the delicate nature of students compared to general users should be taken as a primary consideration compared to only general usability aspect.

Another perspective is how the educational game is being deployed to the students. Also, the different pre-existing conditions of the students themselves, such as its skill level and current grade [18] or its treatment during experimental setup [21], have to be considered. Different treatment may result in a different result, either from students' study performance results or its usability.

Recent studies and development have been done on a digital educational game for high school students focusing on the subject of chemistry, from its design phase [22] and its performance based on students' scores [23]. The result shows that the digital educational game is effective at improving students' knowledge acquisition process. However, the usability aspect of the developed digital educational game has not been analyzed in-depth. This paper aims to synthesize the gathered usability test result by using one of the existing metrics for usability (CSUQ) and analyze the result based on its end-users (students) details during the game's experimental setup treatments.

The paper is then organized as follows. After the introduction, the second section will cover some theoretical background and related works, specifically the ones related to digital educational games as well as CSUQ itself. The third section introduces the developed digital educational game and its basic mechanics. The fourth section will cover the experimental setup. The fifth section will cover the results and discussion. Finally, the sixth section concludes the paper and discusses some future works.

II. THEORETICAL BACKGROUND AND RELATED WORKS

In this section, several reviews, and theoretical background related to digital educational games, usability, the Computer System Usability Questionnaire (CSUQ) as well as existing usability studies for educational games are presented.

A. Digital Educational Game

The very definition of digital educational games is quite hard to pin down, since there are several terms related closely with digital educational games, such as *gamification*, *Game-Based Learning* (GBL) as well as the popular educational game itself. Before defining digital educational games, to clear up the taxonomy, closely related terms are explained first. Gamification definition can be simplified as "the use of video game elements in non-gaming systems to improve user experience" [24] in which a gamified system mostly classified as a non-game system. GBL ramps up the usage of game elements for educational purpose, instead of just using some elements, GBL incorporate the game as an instructional

strategy [25]. However, GBL does not necessarily mean that the adapted strategy is in a digital form. A narrower and more specific term related to GBL in regards to technology integration is *Digital Game-Based Learning* (DGBL) [26], which combines curricular contents and digital games to increase students' motivation [27]. With both related terms clearly defined to reduce confusion, the educational game can be finally defined.

Generally, the educational game can be defined as a game being designed and used for teaching and learning. Furthermore, it is also designed to help people to learn about a particular subject [28]. Compared to both gamification and GBL (or DGBL), educational game are much more focused on combining entertainment and learning in which the players do not feel like they are learning as in the conventional definition of learning [29]. However, the abovementioned definition of an educational game does not strictly define whether the educational game is in a digitized system. A more specific term for a digitized system for educational purposes is Digital Educational Games (DEG), in which the educational game is deployed in the form of software and has the purpose of teaching a particular subject [30]. Hence, a DEG is different from its related terms, whether from its definition, aspect of gaming being used, as well as how such aspect is being used.

B. Usability and Computer System Usability Questionnaire

Based on the ISO standard 9241-11 [31], usability is defined as to which extent a product can be used to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. Usability in computer systems is used to learn users' experience (UX) when using an application. Usability is also used to measure the general usefulness of a system. There exist multiple standardized questionnaires that can be used to measure perceived usability [32]; one of them is the Computer System Usability Questionnaire (CSUQ). CSUQ is an instrument to measure perceived usability, which was developed by IBM [17] consisting of 19 validated questions using a 7-point Likert scale with an alpha score of 0.89, which represents a high degree of reliability. The questions are shown in Table I shown below with questions 1 through 8 representing the system usefulness, questions 9 through 15 representing the information quality section, questions 16 through 18 representing interface quality, and the overall satisfaction represented in question 19.

C. Usability Studies in Educational Games

Usability studies in educational games are rather scarce. A recent study has been done to analyze student's difficulty in playing a particular game by comparing the results between genders [33] with a result of male students performing slightly better compared to female students. Even in a non-game perspective, there exists a difference in preference between male and female students [34]. Another usability evaluation also has been done on an educational game for visually impaired users [35] in which the results show a promising potency for educational purposes. Both types of research however do not use the aforementioned CSUQ metrics but instead uses a personalized questionnaire. Evidence shows that some research has adapted the CSUQ for usability studies in

an educational game [36] to acquire students' opinions towards a game. Another research also has been done with a modified CSUQ to increase relevance [37]. Generally, usability studies done in an educational context are done to test perceive students' perception towards an educational game, and the result of the studies is often used to improve the game. However, a further analysis of preexisting demographics and details of the students are often left out and not being used as a consideration to do such future improvements.

TABLE. I.   CSUQ Questions by Category

Category	Number	Question
System Usefulness (SYSUSE)	1	Overall, I am satisfied with how easy it is to use the system
	2	It was simple to use this system
	3	I could effectively complete my work quickly using this system
	4	I was able to complete my work quickly using this system
	5	I was able to efficiently complete my work using this system
	6	I felt comfortable using this system
	7	It was easy to learn to use this system
	8	I believe I could become productive quickly using this system
Information Quality (INFOQUAL)	9	The system gave error messages that clearly tell me how to fix problems
	10	Whenever I made a mistake using the system, I could recover easily and quickly
	11	The information (such as online help, on-screen messages, and other documentati on) provided with this system was clear
	12	It was easy to find the information I needed
	13	The information provided for the system was easy to understand
	14	The information was effective in helping me complete the task and scenarios
	15	The organization of information on the system screens was clear
Interface Quality (INTERQUAL)	16	The interface of this system was pleasant
	17	I liked using the interface of this system
	18	This system has all the functions and capabilities I expect it to have
Overall Usability (OVERALL)	19	Overall, I am satisfied with this system.

III. Rate of Reaction Chemistry Educational Game

In this section, the basic concept of rate of reaction in high school chemistry is explained, a digital educational game for high school students focusing on the subject of chemistry is explained.

A. Rate of Reaction in Indonesian High School Chemistry

The rate of reaction in Indonesian high school chemistry concerns mainly on three main sub-topics of focus, which are collision theory, determining the order of reaction, and factors affecting the rate of reaction. Factors affecting the rate of reaction, as well as its effects are shown in Table II [38].

Based on Table II, most of the factors are easily understood due to their linear relationship. However, the enlarged surface area may be easily misunderstood due to its counterintuitive action in which grinding or smashing, i.e., a tablet, will increase its surface area. Breaking down a tablet by using a mortar and pestle into flakes, which can further be broken down into powders is multiplied for each unit of that particular object, which means a powder will have a tremendous amount of unit count compared to a single tablet, hence making it has a larger surface area. At last, an addition of catalysts affects a particular reaction when added. As such, there exist an exhaustive list of catalysts that increase the specific rate of reaction.

B. Application Design

The application design and development for this research includes 12 levels divided into four categories, the first four levels have a single factor affecting the level, and there is no special condition added, the next two levels have two different factors affecting the level, the next six levels have three different factors affecting the levels.

C. Base Gameplay and Mechanics

The common main goal of the game being designed is to improve players' understanding on factors affecting the rate of reaction in high school chemistry subject, while the main goal of the player is to be able to control the factors that affect reaction rate [22]. From a game design perspective, the goal is to create an interactive simulation game in which the players are able to control different factors that affect the rate of reaction. From an educational game design perspective, the goal focuses on knowledge acquisition of the subject of matter while also able to create a fun and engaging experience while doing so.

Fig. 1 shows the main concept of the game in which the players are able to control the factors by using an interaction, it should be able to either accelerate or decelerate the rate of reaction; hence, an indication which shows the current rate of reaction is also required.

TABLE. II.   Factors Affecting the Rate of Reaction Learned in the High School Level

Factors	Rate of reaction increases if…	The rate of reaction decreases if…
Reactant Concentration	Added	Reduced
Surface Area	Enlarged	Shrunk
Temperature	Increased	Decreased
Catalyst	Specific Catalyst Added	No Catalyst/ Incorrect Catalyst Added

Fig. 1. Game Interface Layout.

IV. EXPERIMENT

A usability study is then executed to test the application's perceived usability. In this section, the participants of this experiment, as well as the experiment procedure, is then explained.

A. Participants

The participants of this study are taken from an Indonesian High School level. This study uses a total of 53 samples students. Based on its grade, the samples are divided into 2 (two) grades, 10th grade, and 11th grade, in which 10th grade has no prior knowledge regarding the topic, and 11th grade has prior knowledge regarding the topic. Based on its gender, the samples are divided into 2 (two) genders, male and female. Based on its prior material given, the samples are divided into 2 (two) groups, one with prior materials given to reinforce their study, one with no prior materials given and directly starts the game. Based on its experimental setup, the samples are divided into 2 (two) groups based on its grade, the 10th grade is done with post-test only study design, while the 11th grade is done with pretest-posttest study design. Additionally, the 10th grade has no prior knowledge regarding the rate of reaction topic while the 11th grade has prior knowledge.

B. Procedure

The experimental setup is done within 60-75 minutes of session divided into following:

- Ten minutes of introduction and account registration.
- Ten minutes of a pre-test quiz.
- Five minutes of re-reading subject materials (for experimental group only).
- Twenty minutes of playing the game.
- Ten minutes of a post-test quiz.
- Ten minutes of open interview and answering the usability questionnaire (CSUQ).
- Ten minutes of reserve time.

The experimental setup also sets several technical issues limitation as follows:

- Students bring and use their mobile phones.
- Students may only access the game by using Google Chrome or UC Browser.

Students may only access the game when the session is live.

V. RESULTS AND DISCUSSION

In this section, the reports regarding the application of CSUQ to gather questionnaire is gathered according to participants' subjective perception.

A. General Usability

The individual results of the usability test are shown in Fig. 2. The result indicates that Question 7 (Mean(μ) = 6.057, Standard Dev(σ) = 1.183) as well as Question 15 (Mean(μ) = 6.075, Standard Dev(σ) = 1.053) received a relatively high score compared to the other questions on the tests, suggesting that the game ease-of-use degree is relatively high as well as a relatively good organization of information, both question also has a moderate degree of standard deviation which may depict a consensus among the students.

Inversely, Question 9 (Mean(μ) = 5.226, Standard Dev(σ) = 1.396) yields a relatively low score compared to the other questions on the test, suggesting that the game wasn't able to clearly show error messages to the users, implying that the game lacks intuitiveness, although this question also shows a relatively high degree of standard deviation which may depict a much more spread-out view among the students in this test. Subsequently, Question 10 (Mean(μ) = 5.396, Standard Dev(σ) = 1.214) also indicates that the game has a degree of problem in term of error recovery, along with question 9, this shows quite a significant problem in term of how the game design displays the error and how to recover from such error. Additionally, 43 out of 53 (81.11%) participants are critical regarding the questionnaire, depicting a high degree of participation from the participants on telling their perceived usability in regards to the game.

Fig. 3 depicts the box plot of an averaged values of the CSUQ measures based on its category. Alongside that, Table III depicts the mean and standard deviation of the boxplot. In general, the result indicates a positively perceived usability from all of the participants based on the results across the categories.

System usability category (Mean(μ) = 5.915, Standard Dev(σ) = 0.858) indicates that generally, the students perceive the game to be easy to learn, simple, and useful. The information quality category (Mean(μ) = 5.722, Standard Dev(σ) = 1.002) scores relatively low compared to all of the other categories, as being stated before regarding Question 9 and 10, the lack of proper error display as well as error recovery may be one of the major issues the students are facing when using the game although the game information structure and organization is perceived to be quite good. The interface quality category (Mean(μ) = 5.899, Standard Dev(σ) = 0.999) indicates a consensus between the students that the interface is pleasant, likable, and achieved the students' expectations regarding the game. Lastly, the overall usability (Mean(μ) = 5.981, Standard Dev(σ) = 1.083) indicates that the students' general perceived usability towards the system is quite high as well.

The summary statistics rows of the individual results table:

QUESTION	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	STDEV_P
MEAN	5.868	5.925	5.962	5.792	5.981	5.943	6.057	5.792	5.226	5.396	5.792	5.849	5.755	5.962	6.075	5.943	5.811	5.943	5.981	
STDEV_Q	1.093	1.107	1.16	1.15	0.93	1.167	1.183	1.081	1.396	1.214	1.306	1.133	1.27	1.037	1.053	0.949	1.272	1.064	1.083	

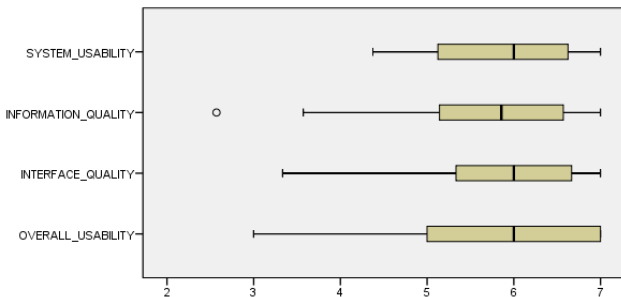Fig. 2.   Individual Results of CSUQ Test.

Fig. 3.   Average Values of CSUQ Categories.

TABLE. III.   MEAN AND STANDARD DEVIATION OF CSUQ CATEGORIES

Statistics	System Usability	Information Quality	Interface Quality	Overall Usability
MEAN	5.915	5.722	5.899	5.981
ST.DEV.	0.858	1.002	0.999	1.083

B. One-Way Analysis of Variance and Correlation

A one-way Analysis of Variance (ANOVA) has been performed to detect whether there exists any difference between the four categories of CSUQ. There was no statistically significant difference being observed as determined by the one-way ANOVA ($F_{(4,53)} = 0.6629$, $p= 0.5757$, $\alpha=0.05$) in which the *p*-value exceeds the α-value of 0.05. This result indicates that there is a similar perception for all the categories listed on the CSUQ test.

Table IV depicts the correlation between each CSUQ category in which the strongest correlation is observed between interface quality and overall usability (*corr*=0.845). This indicates a pleasant interface is vital for a higher overall usability score in case of an educational game system, in which the developed game has been able to reach based on previous results regarding the scores of each CSUQ category. Subsequently, a considerably strong correlation between interface quality and information quality also has been observed (*corr*=0.705). This indicates that the pleasant game interface could be reached with a proper information presentation. Although the aforementioned result is rather weak as shown in Question 9 and Question 10, a broader view of the result regarding information quality based on the average result of Question 9 to Question 15 is able to exceed and overshadow the weak result. This also indicate that no single weakness in information design in particular or game design in general that would be single-handedly responsible towards the usability score.

TABLE. IV.   CORRELATION BETWEEN CSUQ CATEGORIES

	System Usability	Information Quality	Interface Quality	Overall Usability
System Usability	1			
Information Quality	0.727	1		
Interface Quality	0.681	0.705	1	
Overall Usability	0.715	0.720	0.845	1

C. Educational and Practical Insight

In addition to the CSUQ usability results, this research also gathers several educational and practical insight by synthesizing the existing CSUQ categorical results with the different grouping of students during the tests. The result can be generally split into three major categories, which are gender differences, the existence of prior materials given before gameplay, and the study design as well as prior knowledge existence regarding the topic of the game.

A general insight of this section can is presented in Fig. 4 which depicts the result of the individual questions of CSUQ questionnaire of different participants when divided into different groups based on their gender differences, the existence of prior material given, as well as the existence of prior knowledge regarding the game. Subchapters in this sections' results will explains the result in a more in-depth fashion.

In general, male and female students differs the most in term of information quality in which male students has a significantly better perceived usability whereas female students view the game information quality to be somewhat inadequate. The effect of material given before the game session affects the system usability negatively in which the group with no material given before the session perceives the game to be more useful compared to the one with material given beforehand. The group of students with no prior knowledge of the topic conveyed in the game scored much higher in their perception towards the game system usability and game simplicity, the same group also perceive the system interface quality much better compared to the group with prior knowledge, the same group also perceive the overall usability to be much higher compared to the group with prior knowledge.

Fig. 5 depicts an interesting difference between male and female participants in this experiment. The overall usability shows that male and female students perceive the game differently. In general, male students rates the game much higher compared to their female counterparts.

As shown in the information quality category, there exist a stark contrast between male and female students scores in which the upper quartile on the male boxplot aligns with its upper whisker. This also happens on every single category in the male column. On the contrary, the female column shows more than an entire digit of difference between its median and the upper whisker. The same also happens on all other categories except the overall usability.

This result may indicate a difference between male and female students' perception towards educational game. Based on Fig. 6 and Fig. 7, male students are more likely to rate the game higher compared to their female counterpart, specifically, in general, female students rated Question 9 much lower compared to male students. This result could mean that female students require more intuitive design to cater to their expectation compared to male students. Similar result also depicted in Question 12, which male students generally rate it at least two digits higher compared to female students.

Another interesting result can also be seen in Fig. 6 on the male row which shows the median in several questions (Question 1, 2, 7, 12, 14, and 15) aligns with the upper whisker of the boxplots. This result may indicate how male students perceive educational game as a game and view it from a logical perspective by using the information given from the game much effectively compared to their female counterpart.

Based on these results, a general assumption can be made. The main difference found between male and female students are more focused on how the in-game information are being perceived as well as the perceived intuitiveness of the game. Male students may find the game to be easier to grasp compared to their female counterparts. To counteract this issue, an educational game design needs a clear depiction of information in order to improve the educational game perceived usability from female students' perspective. Hence, gender demographic in educational game may affect game design choices, especially in term of information structure in educational game.

DATASET	SYS_USE								INFO_QUAL							INTER_QUAL			OVERALL
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
BASE_AVERAGE	5.868	5.925	5.962	5.792	5.981	5.943	6.057	5.792	5.226	5.396	5.792	5.849	5.755	5.962	6.075	5.943	5.811	5.943	5.981
MALE_AVERAGE	6.036	6.036	5.964	5.857	6.036	5.929	6.107	5.893	5.643	5.607	6	6.214	6	6.25	6.286	5.893	5.821	6.036	6.143
FEMALE_AVERAGE	5.68	5.8	5.96	5.72	5.92	5.96	6	5.68	4.76	5.16	5.56	5.44	5.48	5.64	5.84	6	5.8	5.84	5.8
PRIOR_MATERIAL_AVERAGE	5.483	5.552	5.586	5.448	5.586	5.483	5.552	5.345	5.034	5.172	5.655	5.69	5.379	5.69	5.862	5.724	5.724	5.793	5.759
NO_PRIOR_MATERIAL_AVERAGE	6.333	6.375	6.417	6.208	6.458	6.5	6.667	6.333	5.458	5.667	5.958	6.042	6.208	6.292	6.333	6.208	5.917	6.125	6.25
PRIOR_KNOWLEDGE_AVERAGE	5.333	5.458	5.792	5.833	5.833	5.625	6.042	5.583	5.25	5.375	5.625	5.792	5.75	5.833	6	5.833	5.583	5.75	5.625
NO_PRIOR_KNOWLEDGE_AVERAGE	6.31	6.31	6.103	5.759	6.103	6.207	6.069	5.966	5.207	5.414	5.931	5.897	5.759	6.069	6.138	6.034	6	6.103	6.276

Fig. 4.   Individual CSUQ Questions Result of the Participants, Separated into different Groups of Gender, Material Existence, as well Student's Prior Knowledge.
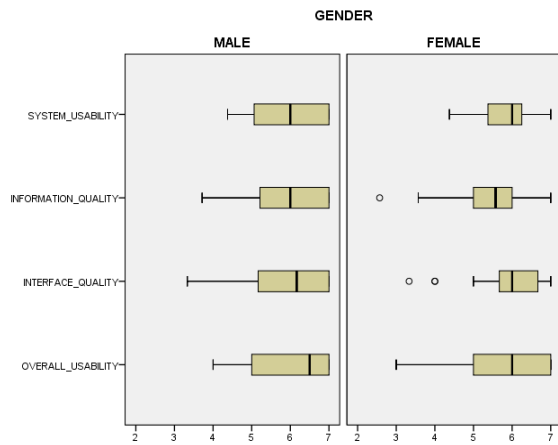
Fig. 5.   Average Values of CSUQ Categories Distribution between Genders.
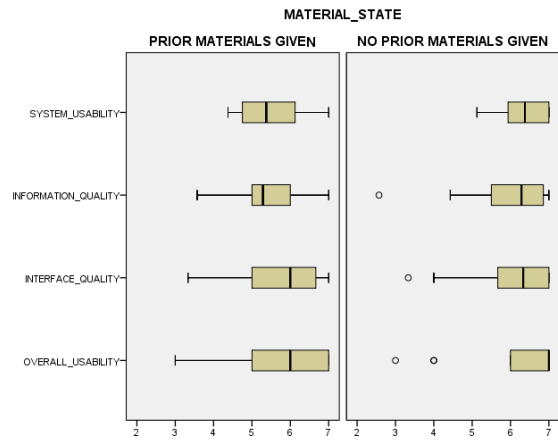

Fig. 6.   Average Values of CSUQ Categories Distribution based on the Existence of Materials given before the Game Session.

The existence of materials related to the game topic given before the game session shows an interesting result. Based on Fig. 8, overall, the usability was rated lower across the board on the group with materials given beforehand while the one with no materials given beforehand has much higher rating. The difference was particularly high especially in system

usability which the rating range between both groups are vastly different. This may indicate that the existence of materials given before the game session may affect the usability negatively as it may increase students' expectation towards the game. Generally speaking, sophisticated method was expected by students in this current era where smartphones is pervasive and entertainment gaming are much more graphically entertaining. The game was unable to reach such expectation and create a negative impression from the students. As mentioned before, this was particularly high in the system usability as students with materials given beforehand rated the game to be less useful as well as less sophisticated for their expectation.

In respect with Fig. 4, Fig. 9 and Fig. 10 shows that the first eight questions are rated much higher/lower depending on whether the material before the game session was given/not. Similarly, the range of scores was particularly different in Q7 in which the ease of use of the game was highly rated. The general result also shows a clear pattern that materials given before the game session affects the game usability negatively, however, a further investigation is needed whether this is just a case unique to this experiment or it is a general consensus that is adaptable to each game in existence.

Between the students' groups that has the knowledge regarding the game beforehand or not – in this case, the students have learnt about the topic of reaction rate beforehand – the usability scores show some difference in overall usability. Based on Fig. 11, the group with no prior knowledge rated the overall usability much higher compared to the ones with prior knowledge. As a student has no prior knowledge, the expectation towards the game may be lower, hence the usability rating is also higher.

Additionally, in respect to Fig. 4, Fig. 12 and Fig. 13 depicts a more specific result and difference especially in Q1 and Q2, but also in Q17 to Q19. The group with no prior knowledge perceives the game interface to be as high as their expectation as well as generally sees the game to be useful while the group with prior knowledge may either prefer the book instead or the game was simply not enticing enough for their expectation.
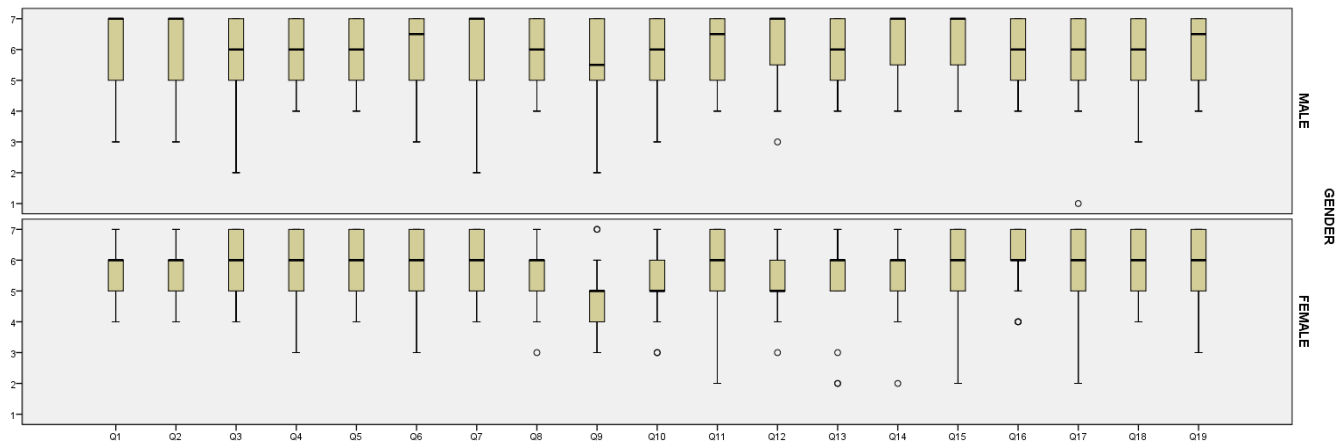

Fig. 7.   Boxplot of the Average Score of CSUQ Per Questions Separated between Genders.

DATASET	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
BASE_AVERAGE	5.868	5.925	5.962	5.792	5.981	5.943	6.057	5.792	5.226	5.396	5.792	5.849	5.755	5.962	6.075	5.943	5.811	5.943	5.981
MALE_AVERAGE	6.036	6.036	5.964	5.857	6.036	5.929	6.107	5.893	5.643	5.607	6	6.214	6	6.25	6.286	5.893	5.821	6.036	6.143
FEMALE_AVERAGE	5.68	5.8	5.96	5.72	5.92	5.96	6	5.68	4.76	5.16	5.56	5.44	5.48	5.64	5.84	6	5.8	5.84	5.8

Fig. 8.   Difference in Average Score of CSUQ between Genders.
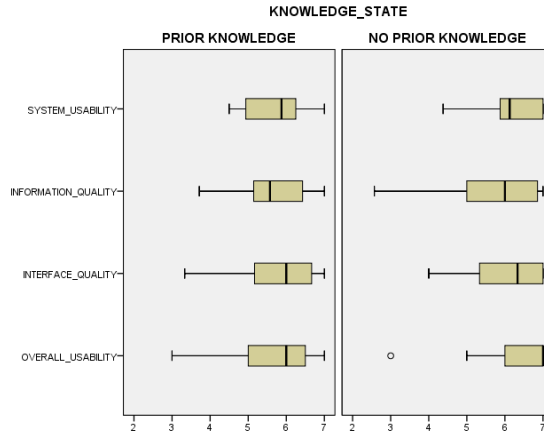

Fig. 9.   Average Values of CSUQ Categories Distribution based on the Existence of Students' Prior Knowledge before the Game Session.
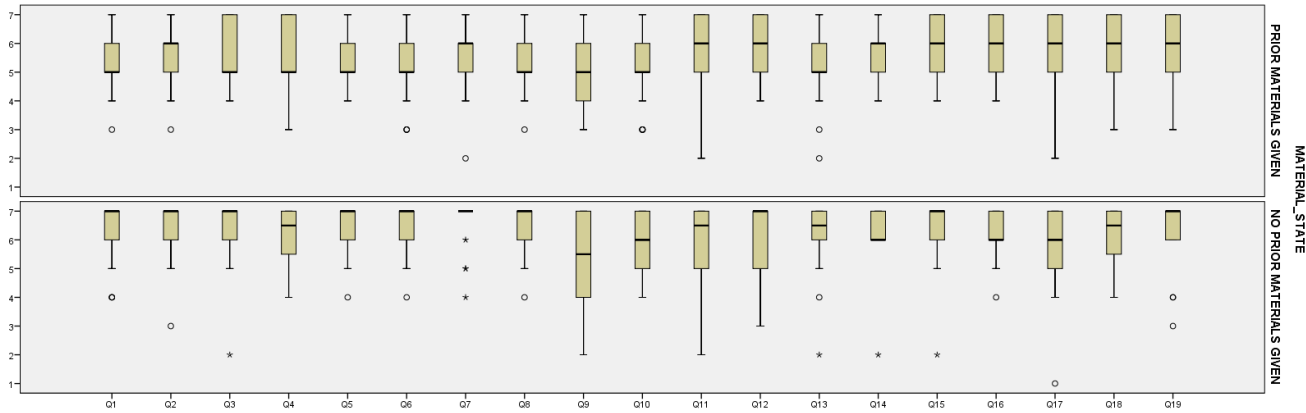

Fig. 10. Boxplot of the Average Score of CSUQ Per Questions Separated between Prior Material Existence.

DATASET	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
BASE_AVERAGE	5.868	5.925	5.962	5.792	5.981	5.943	6.057	5.792	5.226	5.396	5.792	5.849	5.755	5.962	6.075	5.943	5.811	5.943	5.981
PRIOR_MATERIAL_AVERAGE	5.483	5.552	5.586	5.448	5.586	5.483	5.552	5.345	5.034	5.172	5.655	5.69	5.379	5.69	5.862	5.724	5.724	5.793	5.759
NO_PRIOR_MATERIAL_AVERAGE	6.333	6.375	6.417	6.208	6.458	6.5	6.667	6.333	5.458	5.667	5.958	6.042	6.208	6.292	6.333	6.208	5.917	6.125	6.25

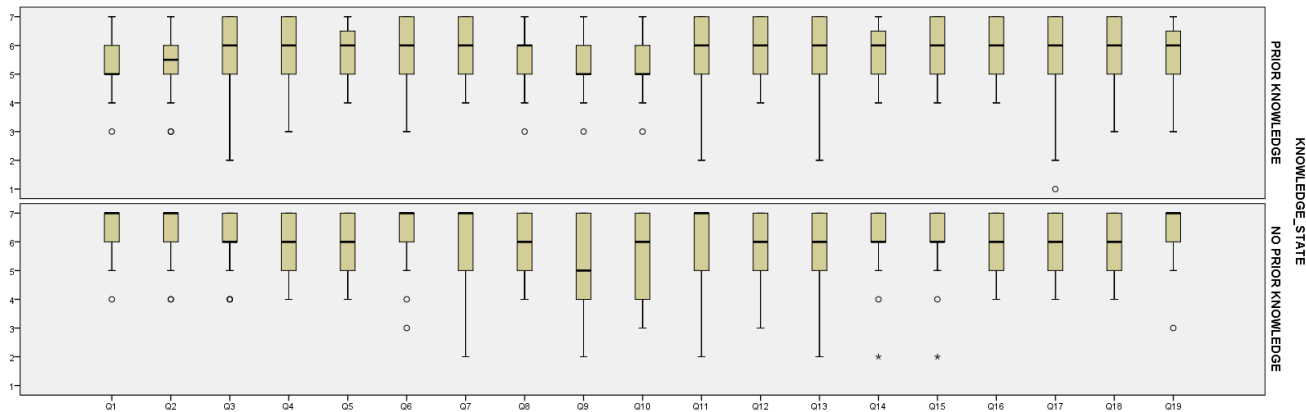Fig. 11. Difference in Average Score of CSUQ between Prior Material Existence.

Fig. 12. More Specific Boxplot of the Average Score of CSUQ Per Questions Separated between Prior Material Existence.

DATASET	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
BASE_AVERAGE	5.868	5.925	5.962	5.792	5.981	5.943	6.057	5.792	5.226	5.396	5.792	5.849	5.755	5.962	6.075	5.943	5.811	5.943	5.981
PRIOR_KNOWLEDGE_AVERAGE	5.333	5.458	5.792	5.833	5.833	5.625	6.042	5.583	5.25	5.375	5.625	5.792	5.75	5.833	6	5.833	5.583	5.75	5.625
NO_PRIOR_KNOWLEDGE_AVERAGE	6.31	6.31	6.103	5.759	6.103	6.207	6.069	5.966	5.207	5.414	5.931	5.897	5.759	6.069	6.138	6.034	6	6.103	6.276

Fig. 13. More Specific Difference (especially in Q1 and Q2, also in Q17 to Q19) in Average Score of CSUQ between Prior Material Existence.

VI. Conclusion and Future Works

This result presents the result of a usability testing viewed from the educational perspective of a high school chemistry educational game regarding the topic of reaction rate. The work done is contributing to the field of educational game development, specifically in terms of educational game evaluation, as well as actions that can be taken in an experimental treatment.

The initial result shows that general usability has been reached by the students as well as a high degree of correlation between each category in the usability test has been reached.

The result shows that, generally, usability scores in an educational game are affected by different gender groups, pre-game session materials, as well as prior knowledge regarding the game topic. The usability test scores show that each category in the usability tests yields a different result. In terms of system usability, the group with no materials given before the game session yields the best results. In terms of information quality, the group of male students yields the best results. In terms of interface quality, as well as overall usability, the group with no prior knowledge has a somewhat higher result.

All of the results on this, however, needs to be re-validated with a higher number of datasets to improve statistical significance, as well as done with different games to improve the validity of the result. Future works may also include more validation by relating the usability scores with students' performance as well as students' actions during the gameplay itself to see the educational effect of the game more precisely.

Acknowledgment

References

[1] J. H. Kuznekoff and S. Titsworth, "The Impact of Mobile Phone Usage on Student Learning," Commun. Educ., vol. 62, no. 3, pp. 233–252, 2013.

[2] R. Ferguson, "Learning analytics: Drivers, developments and challenges," Int. J. Technol. Enhanc. Learn., vol. 4, no. 5–6, pp. 304–317, 2012.

[3] J. L. Gómez-Urquiza, J. Gómez-Salgado, L. Albendín-García, M. Correa-Rodríguez, E. González-Jiménez, and G. A. Cañadas-De la Fuente, "The impact on nursing students' opinions and motivation of using a 'Nursing Escape Room' as a teaching game: A descriptive study," Nurse Educ. Today, vol. 72, pp. 73–76, 2019.

[4] K. Fellnhofer, "All-in-one: Impact study of an online math game for educational purposes," Int. J. Technol. Enhanc. Learn., vol. 8, no. 1, pp. 59–76, 2016.

[5] H. W. Lin and Y. L. Lin, "Digital educational game value hierarchy from a learners' perspective," Comput. Human Behav., vol. 30, no. 1, pp. 1–12, 2014.

[6] A. Fairuzabadi, A. A. Supianto, and H. Tolle, "Analysis of Players' Speed Thinking in Color Mix Game Application," Int. J. Interact. Mob. Technol., vol. 12, no. 8, pp. 113–122, 2018.

[7] A. A. Syahidi, A. A. Supianto, and H. Tolle, "Design and Implementation of Bekantan Educational Game (BEG) as a Banjar Language Learning Media," Int. J. Interact. Mob. Technol., vol. 13, no. 3, pp. 108–124, 2019.

[8] I. R. D. Renavitasari, A. A. Supianto, and H. Tolle, "Log Data Analysis of Player Behavior in Tangram Puzzle Learning Game," Int. J. Interact. Mob. Technol., vol. 12, no. 8, pp. 123–129, 2018.

[9] I. Wardani, H. Tolle, and I. Aknuranda, "Evaluation of an Educational Media on Cube Nets Based on Learning Effectiveness and Gamification Parameters," Int. J. Emerg. Technol. Learn., vol. 14, no. 14, pp. 4–18, 2019.

[10] N. M. Diah, M. Ismail, S. Ahmad, and M. K. M. Dahari, "Usability testing for educational computer game using observation method," in Proceedings - 2010 International Conference on Information Retrieval and Knowledge Management: Exploring the Invisible World, CAMP'10, 2010, pp. 157–161.

[11] O. Álvarez-Xochihua, P. J. Muñoz-Merino, M. Muñoz-Organero, C. D. Kloos, and J. A. González-Fraga, "Comparing usability, user experience and learning motivation characteristics of two educational computer games," ICEIS 2017 - Proc. 19th Int. Conf. Enterp. Inf. Syst., vol. 3, no. Iceis, pp. 143–150, 2017.

[12] J. Nielsen, Usability Engineering. Elsevier, 1994.

[13] C. Lu, M. Chang, Kinshuk, E. Huang, and C. W. Chen, "Usability of context-aware mobile educational game," Knowl. Manag. E-Learning, vol. 3, no. 3, pp. 448–477, 2011.

[14] C. G. Brown-Johnson, B. Berrean, and J. K. Cataldo, "Development and usability evaluation of the mHealth Tool for Lung Cancer (mHealth TLC): A virtual world health game for lung cancer patients," Patient Educ. Couns., vol. 98, no. 4, pp. 506–511, 2015.

[15] J. Brooke, "SUS-A quick and dirty usability scale," in Usability evaluation in industry, London, UK: Taylor & Francis, 1996, pp. 189–194.

[16] K. Finstad, "The usability metric for user experience," Interact. Comput., vol. 22, no. 5, pp. 323–327, 2010.

[17] J. R. Lewis, "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use," Int. J. Hum. Comput. Interact., vol. 7, no. 1, pp. 57–78, 1995.

[18] F. Adnan, B. Prasetyo, and N. Nuriman, "Usability testing analysis on the Bana game as education game design references on junior high school," J. Pendidik. IPA Indones., vol. 6, no. 1, pp. 88–94, 2017.

[19] J. L. P. Medina et al., "Usability study of a web-based platform for home motor rehabilitation," IEEE Access, vol. 7, pp. 7932–7947, 2019.

[20] R. Khajouei and F. Farahani, "The evaluation of users' satisfaction with the Social Security Electronic System in Iran," Heal. Technol., pp. 1–8, 2019.

[21] S.-J. Lu, Y.-C. Liu, P.-J. Chen, and M.-R. Hsieh, "Evaluation of AR embedded physical puzzle game on students' learning achievement and motivation on elementary natural science," in Interactive Learning Environments, 2018, vol. 4820, pp. 1–13.

[22] M. Hafis, H. Tolle, A. A. Supianto, I. C. Christian, L. S. Atmojo, and S. N. Rochmainy, "Game Design Elements and Educational Game Design for Rate of Reaction Topic in High School Chemistry Subject," in 5th International Conference on Science and Technology (ICST 2019), 2019, pp. 1–6.

[23] A. A. Supianto, M. Hafis, and H. Tolle, "Significance of Dynamic Difficulty Adjustment in Delivering Instructional Scaffolding on Educational Game for Rate of Reaction Topic in High School Chemistry Subject," in 3rd International Conference on Education and Multimedia Technology (ICEMT 2019), 2019, pp. 1–6.

[24] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon, "Gamificaiton: Using Game Design Elements in Non-Gaming Contexts," in CHI'11 extended abstracts on human factors in computing systems, 2011, pp. 2425–2428.

[25] P. Giani and C. G. Von Wangenheim, "How to Evaluate Educational Games : a Systematic Literature Review," J. Univers. Comput. Sci., vol. 22, no. 7, pp. 992–1021, 2016.

[26] M. Prensky, "Digital Game-based Learning Prensky," Comput. Entertain., vol. 1, no. 1, pp. 1–4, 2003.

[27] M. Papastergiou, "Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation," Comput. Educ., vol. 52, no. 1, pp. 1–12, 2009.

[28] R. Al-Azawi, F. Al-Faliti, and M. Al-Blushi, "Educational Gamification Vs. Game Based Learning: Comparative Study," Int. J. Innov. Manag. Technol., vol. 7, no. 4, pp. 131–136, 2016.

[29] G. Bente and J. Breuer, "Why so serious? On the Relation of Serious Games and Learning," Eludamos - J. Comput. Game Cult., vol. 4, no. 1, pp. 7–24, 2010.

[30] S. Aslan and O. Balci, "GAMED: Digital educational game development methodology," Simul. Trans. Soc. Model. Simul. Int., vol. 91, no. 4, pp. 307–319, 2015.

[31] T. Stewart, "Ergonomic Requirements for Office Work with Visual Display Terminals (VDTS): Part 11: Guidance on Usability, document ISO 9241," 1998.

[32] J. R. Lewis, "Measuring Perceived Usability: The CSUQ, SUS, and UMUX," Int. J. Hum. Comput. Interact., vol. 34, no. 12, pp. 1148–1156, 2018.

[33] S. AlDakhil, E. Al Taleb, M. Al Ghamlas, and S. Al-Megren, "Assessing the Usability of a Tangible Educational Game for Children," in 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 2019, pp. 1–5.

[34] A. Al-Sa' di, D. Parry, and P. D. Carter, "User interface preferences of young Jordanians using tablet devices," Int. J. Technol. Enhanc. Learn., vol. 10, no. 3, pp. 202–217, 2018.

[35] A. G. D. Correa, L. C. C. De Biase, E. P. Lotto, and R. D. Lopes, "Development and Usability Evaluation of an Configurable Educational Game for the Visually Impaired," 2018 IEEE Games, Entertain. Media Conf. GEM 2018, pp. 173–180, 2018.

[36] K. Maragos, "Web based Adaptive Educational Games - Exploitation in Computer Science Education," National and Kapodistrian University of Athens, 2012.

[37] M. W. M. Johnson, M. Eagle, and T. Barnes, "Invis: An interactive visualization tool for exploring interaction networks," in Educational Data Mining 2013, 2013.

[38] M. Trautz, "Das Gesetz der Reaktionsgeschwindigkeit und der Gleichgewichte in Gasen. Bestätigung der Additivität von Cv‐3/2R. Neue Bestimmung der Integrationskonstanten und der Moleküldurchmesser," Zeitschrift für Anorg. und Allg. Chemie, vol. 96, no. 1, pp. 1–28, 1916.