EVALUATION OF USER EXPERIENCE TLX TRAINING GATE FOR COMPETITIVE PROGRAMMING LEARNING USING USER EXPERIENCE QUESTIONNAIRE AND SYSTEM USABILITY SCALE

Guntur Eka Saputra[a], Rakhmi Khalida[b], Ratu Nurmalika[c]

[a,b] Faculty of Industrial Technology / Information Technology, Universitas Gunadarma
[c] Faculty of Civil Engineering and Design , Universitas Gunadarma
Coresponding_Email: gunturesaputra@gmail.com

ABSTRACT

Usability is a measurement of software applications to determine how easy the software is. Usability testing is done by involving end users or it can also involve experts. User Experience Questionnaire (UEQ) and System Usability Scale (SUS) are two approaches that can be used to measure Usability. Based on the provisions for determining the results of the SUS assessment, the percentile rank score, which is seventy-five, the results of the respondents' assessment on the TLX Training Gate application, the user acceptance level is in the acceptable category, the scale grade level is in category C, and the rating adjective is in a good category and with UEQ approach There are 3 (three) scales in the Above Average range and 2 (two) scales still below average, and 1 on the good scale. The result of the stimulation score (1,35) and attractiveness (1,27) is because this system is attractive and comfortable for students so the value of this aspect has a positive impression. Based on research, for 6 (six) aspects this shows that system development is needed so that students as users can maximize their learning potential in competitive programming learning.

Keywords: Competitive Programming, SUS, TLX Training Gate, UEQ, Usability.

1. INTRODUCTION

Programming language or often known as computer language is a command/instruction technique to command a computer. This programming language is a set of syntax and semantic rules used to define computer programs. Programming languages are known to be quite difficult to understand by ordinary people and seem boring [1]. Understanding programming languages requires diligent and continuous learning. In college, programming language is used as one of the courses, but due to time constraints, students need a medium to learn programming languages outside of lecture hours.

One of the programming languages learning and training logic to solve problems is competitive programming. Competitive Programming (CP) is a form of IT competition in which participants are given questions that are solved by designing a program that processes a number of inputs into the required output within a limited period of time [2]. Students are required to implement the program correctly and efficiently so that the answers are accepted. Competitive Programming focuses on studying various types of algorithms, data structures, and implementation techniques that are often used to solve CP questions through regular training and students can actively participate in CP competitions that are held regularly. Along with the development of technology, today's learning media is not only in formal rooms, but also online in the form of a website that can be followed by simply registering online and attending training as needed. One of the websites that will be discussed in this paper is TLX Training Gate as a medium for learning and practicing programming languages.

TLX Training Gate is a website that is used to learn and practice programming. The site can be accessed on the https://tlx.toki.id/training page. This site is not only used for learning but also for preparation for the Competitive Programming (CP) competition. This site is developed and regularly updated by the Indonesian Computer Olympiad Alumni Association. On this site, material, practice questions, practical exercises, and submissions are presented which are practical tasks to code the program with a file name, then submit the user's work so that the system can assess whether the program is correct or not. Figure 1 presents the front page view of TLX Training Gate [3].
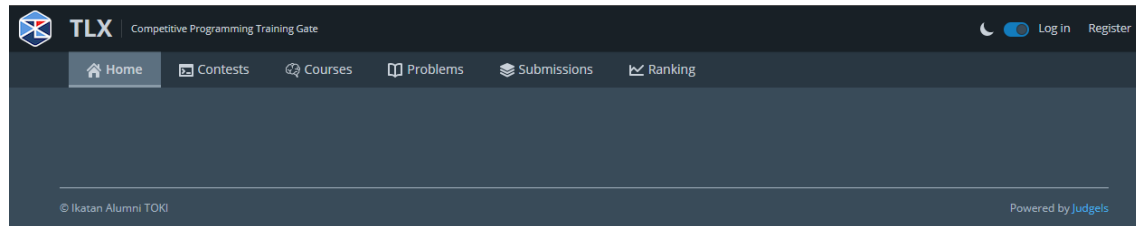
Figure 1. Front Page TLX Training Gate

Usability or what is often known as usability is a measurement of software applications to determine how easy the software is, which is seen from five aspects, namely learnability efficiency, memorability, errors and satisfaction. Usability analysis can be done using the User Experience Questionnaire (UEQ) and System Usability Scale (SUS) approaches. Usability testing is done by involving end users or it can also involve experts. Each usability testing approach has its own characteristics so that each approach can complement each other for the software evaluation process. These characteristics include the number of respondents involved, measurement instruments, measurement steps, scoring system, assessment results [4] [5].

User Experience Questionnaire (UEQ) is a questionnaire used to measure user experience with a product [6]. The purpose of using this questionnaire is to quickly assess the perceived user experience of a product. There are 6 measurement scales in the UEQ, namely Attractiveness, Perspicuity, Perspicuity, Dependability, Stimulation, and Novelty. Research on usability analysis using the User Experience Questionnaire (UEQ) on the web or application was previously conducted by Sonia et al. His research compared the user experience obtained when using three e-commerce websites that have the same characteristics and business processes [7]. Other related research is on usability analysis of the Ruang Guru application with the User Experience Questionnaire (UEQ) method [8]. Another research related to UEQ is a study of academic information systems on how students can use the system well without feeling confused or ambiguous [9].

System Usability Scale (SUS) is a test by involving end users in the process, so that the evaluation results will be more in line with the real situation [4] [5]. The previous research Using the System Usability Scale (SUS) for the web or application on the Palembang guide interface as a supporting media for the XVIII Asian Games. In this test, ten respondents were used as the test sample. The results showed that the assessment score was at 83, which means that the Palembang Guide software was declared acceptable and included in grades A and B with an excellent rating [10]. SUS testing has also been carried out on the Bandung Regency Livestock Service personnel application using five respondents as the test sample. The personnel application gets index B and can be applied [11]. Thus, it is hoped that the Palembang Guide software test using the SUS instrument will get accurate test results. This study focuses on how the responses of students who use TLX Training Gate to be able to learn well and easily use this platform in terms of user experience and user scalability.

2. BASIC THEORY
2.1. User Experience
User experience refers to how the behavior of users using the software [12]. Positive user behavior is when the software provides an experience in which the user's goals (finishing work) are met. One of the international standards, namely ISO FDIS 9241-210 defines user experience as an individual's perception and response resulting from the use and anticipation of the usefulness of a product, system, and service. The user experience conceptualization has been carried out in several ways and is divided into three parts [13], namely:
1. Explore indicators of user satisfaction in terms of usability.
2. Emphasis on user performance.
3. User experience is the parent of all user perceptions and responses that are measured subjectively and objectively.

2.2. User Experience Questionnaire (UEQ)
The User Experience Questionnaire is one of the questionnaires whose results can be used in usability testing to quickly measure the level of user experience of a product. UEQ can be downloaded at www.ueq-online.org. There are six scales with a total of 26 elements which are categorized based on the measurement scales contained in the UEQ [6]. From the six scales of the UEQ, it can be grouped into three main groups, namely attractiveness, pragmatic quality, and hedonic quality as in figure 2. The main aspect of the UEQ is attractiveness which is the overall impression or attractiveness of the product from the user's point of view. The pragmatic quality group is a technical aspect related to the user's point of view regarding the use of the available system products or services, while the hedonic quality group is a non-technical aspect regarding the

emotions or feelings experienced by users when using the product. User Experience Aspects in the questionnaire in figure 2.
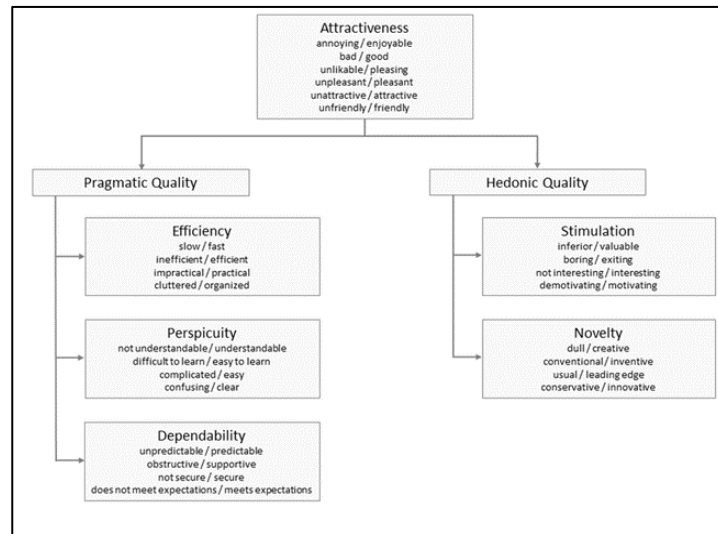

Figure 2. User Experience Questionnaire Aspects [6]

1. Attractiveness: The impression of the product. Do users like or dislike the product? How big is the appeal of a product. For example: good or bad, attractive or not attractive.
2. Efficiency (Efficiency): The possibility of a product can be used quickly and efficiently. How much the user can complete the task without great effort or efficiency. For example: sooner or later, practical or impractical.
3. Perspicuity: Is it possible to use this product? Is it easy for users to get used to using this product? How much perspicuityof a product. For example: easy to understand or difficult to understand.
4. Dependability: Does the user feel in control of the interaction? Are interactions with the product safe and predictable? How much precision is felt by the user through the control he has. For example: predictable or unpredictable, support or hinder.
5. Stimulation: Is it interesting and enjoyable for the user to use the product? Do users feel motivated to continue using the product? How much motivation to use the product. For example: useful or less useful, interesting or unattractive.
6. Novelty: Is the product design innovative and creative? Is the product able to get the user's attention? How creative or great is the novelty of the product.

UEQ data can be processed using the UEQ analysis tool. To find out the meaning of the values generated by each UEQ scale, it is done by comparing the calculated values for each scale with the values in Table 1 below.

Table 1. Benchmark Interval for UEQ Scale

Category	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Excellent	≥1,75	≥1,9	≥1,78	≥1,65	≥1,55	≥1,4
Good	1,52≥n<1,75	1,56≥n<1,9	1,47≥n<1,78	1,48≥n<1,65	1,31≥n<1,55	1,05≥n<1,4
Above Average	1,17≥n<1,52	1,08≥n<1,56	0,98≥n<1,47	1,14≥n<1,48	0,99≥n<1,31	0,71≥n<1,05
Below Average	0,7≥n<1,17	0,64≥n<1,08	0,54≥n<0,98	0,78≥n<1,14	0,5≥n<0,99	0,3≥n<0,71
Bad	<0,7	<0,64	<0,54	<0,78	<0,5	<0,3

2.3. System Usability Scale (SUS)

System Usability Scale (SUS) is a test that the evaluation results will be more in line with the real situation. Testing with the System Usability Scale (SUS) approach has 10 statements as a testing tool, and also does not require a large number of samples so as to minimize testing costs [3] [4]. 10 statements from the system usability scale as shown in table 2 below.

Table 2. System Usability Scale Statements

No	Statement	Scale
1	*I think that I would like to use this system frequently*	0-4
2	*I found the system unnecessarily complex*	0-4
3	*I thought the system was easy to use*	0-4
4	*I think that I would need the support of a technical person to be able to use this system*	0-4
5	*I found the various functions in this system were well integrated*	0-4
6	*I thought there was too much inconsistency in this system*	0-4
7	*I would imagine that most people would learn to use this system very quickly*	0-4
8	*I would imagine that most people would learn to use this system very quickly*	0-4
9	*I would imagine that most people would learn to use this system very quickly*	0-4
10	*I needed to learn a lot of things before I could get going with this system*	0-4

SUS develops an assessment strategy. Each item has a score contribution ranging from 0 to 4. For odd statement items, namely 1, 3, 5, 7, and 9 (positive statement items) and the score contribution is the position of the scale minus 1. While for even statement items, which is 2, 4, 6, 8, and 10 (negative statement items), the contribution is 5 minus the value of the scale range. Then, do the calculation by multiplying the total score by 2.5 to get the overall SUS score.

3. RESEARCH METHOD

The steps taken when carrying out the research are described in the figure 3 below.
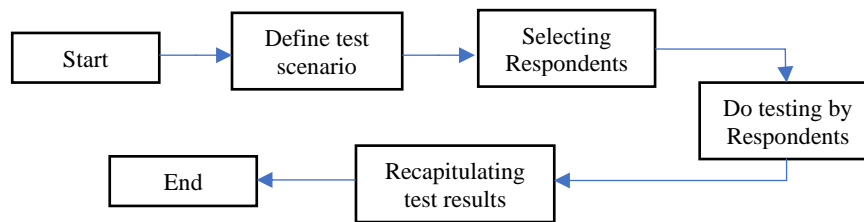

Figure 3. Research Method Flow

3.1. Define Test Scenario

In this research, the initial stage is a study of the literature related to user experience, user experience questionnaire, and usability scale system as well as previous research that can be used as a reference for this research.

3.2. Selecting Respondents

The respondents of this research are students who take part in the competitive programming community at universities who have active status in the even semester of the 2021/2022 academic year. The number of samples is determined by using the Slovin formula in equation 1 below:

$$n = \frac{N}{1 + (N \; x \; e^2)}$$

Where:
n = number of samples
N = number of populations
e = margin of error

In this study, the total population was 65 students with a margin of error of 10% or 0,1. So, by using equation 1 the minimum number of samples is calculated as follows:

$$n = \frac{65}{1 + (65 \; x \; (0,1)^2)}$$
$$n = \; 39,39$$

Based on the results of calculations using equation 1, the minimum population is 39,39 which is rounded up to 40.

3.3. Do Testing by Respondents

After calculating the minimum number of samples, the next step is to collect data using UEQ and SUS. The evaluation process using UEQ starts from the respondent being asked to use all the existing features. The goal is for users to get good feedback and accurate UEQ results. Users can fill in the UEQ according to the impression they feel when using the website, both in terms of function, color, type of writing, layout and others.

After using the system, the user fills out a given questionnaire. Users fill out the UEQ questionnaire, through a google form that has been distributed to users. Filling in the UEQ will be directed according to the UEQ provisions, where point 7 is not always the highest score on this questionnaire. In this study, user experience measurements were carried out using the UEQ questionnaire which consisted of 26 statements as shown in Figure 4.

	1	2	3	4	5	6	7		
menyusahkan	O	O	O	O	O	O	O	menyenangkan	1
tak dapat dipahami	O	O	O	O	O	O	O	dapat dipahami	2
kreatif	O	O	O	O	O	O	O	monoton	3
mudah dipelajari	O	O	O	O	O	O	O	sulit dipelajari	4
bermanfaat	O	O	O	O	O	O	O	kurang bermanfaat	5
membosankan	O	O	O	O	O	O	O	mengasyikkan	6
tidak menarik	O	O	O	O	O	O	O	menarik	7
tak dapat diprediksi	O	O	O	O	O	O	O	dapat diprediksi	8
cepat	O	O	O	O	O	O	O	lambat	9
berdaya cipta	O	O	O	O	O	O	O	konvensional	10
menghalangi	O	O	O	O	O	O	O	mendukung	11
baik	O	O	O	O	O	O	O	buruk	12
rumit	O	O	O	O	O	O	O	sederhana	13
tidak disukai	O	O	O	O	O	O	O	menggembirakan	14
lazim	O	O	O	O	O	O	O	terdepan	15
tidak nyaman	O	O	O	O	O	O	O	nyaman	16
aman	O	O	O	O	O	O	O	tidak aman	17
memotivasi	O	O	O	O	O	O	O	tidak memotivasi	18
memenuhi ekspektasi	O	O	O	O	O	O	O	tidak memenuhi ekspektasi	19
tidak efisien	O	O	O	O	O	O	O	efisien	20
jelas	O	O	O	O	O	O	O	membingungkan	21
tidak praktis	O	O	O	O	O	O	O	praktis	22
terorganisasi	O	O	O	O	O	O	O	berantakan	23
atraktif	O	O	O	O	O	O	O	tidak atraktif	24
ramah pengguna	O	O	O	O	O	O	O	tidak ramah pengguna	25
konservatif	O	O	O	O	O	O	O	inovatif	26

Figure 4. Instruments for Testing the User Experience Questionnaire Approach

The data that has been collected from the respondents is then entered into the UEQ data analysis tool and used as an input value to calculate the user experience value of TLX Gate Training. Testing the reliability of the questionnaire data was carried out using the Cornbach alpha coefficient. The cornbach alpha coefficient describes the consistency for all UX factor scales. The test results show that the Cornbach alpha value of all UX factors is above an acceptable scale (>0.7), namely attractiveness (0.9), perspicuity (0.79), efficiency (0.88), accuracy (0.83), stimulation (0.81), and novelty (0.83). This shows that the data analyzed using UEQ is reliable. For the calculation of the SUS scale using 10 questions in table 2, and following the rules of the percentile rank score in figure 5 [4] [5]. There are 6 (six) grade scale, namely A, B, C, D, and F (E include to F) grade scale).
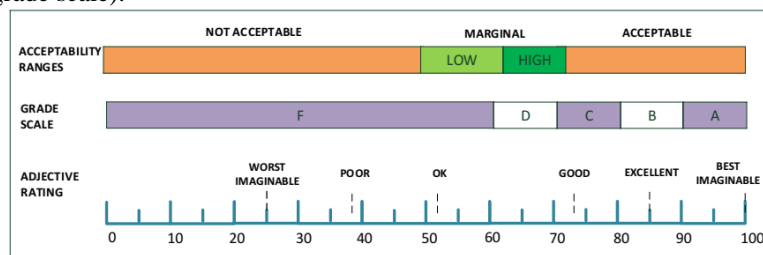
Figure 5. SUS Percentile Rank Score

34

3.4. Recapitulating Test Results

Processing of the data obtained from the respondents was then processed using UEQ tools and SUS grade scale. The results of data processing are then analyzed to obtain user experience information from TLX Training Gate.

4. RESULT AND DISCUSSION

At this stage, it discusses the results of respondents' testing of the TLX Training Gate application. Testing through the User Experience Questionnaire approach, respondent data is entered into the Data Analysis Tool which has been downloaded from the official UEQ website, this is a calculator that calculates UX with UEQ statistical calculations. Incoming data is analyzed first for an inconsistency scale analysis which describes the respondents filling out the questionnaire seriously or not seriously (less understanding of the test items).

A total of 40 respondents have filled in and 36 were declared valid or consistent in filling out. Respondents are web users to learn competitive programming. The next step is to check the Cronbach Alpha coefficient. The Cronbach Alpha coefficient is a measure for scale consistency. In general, the scale assumption must show an Alpha value > 0.7, which is considered to have high consistency. The next step is to calculate the mean and variance. This is done to find out all respondents' responses that have been grouped on each scale, and to show variations in the distribution of the data. The next step is to calculate the pragmatic and hedonic qualities. This is done to find out which qualities have high and low values to be used as a reference for making decisions.

From the results of calculations with UEQ, it is found that the attractiveness, efficiency, stimulation, and novelty scales are above the average results from the comparison with the benchmark, and the results show that the perspicuity scale is 50% of results better and in below average and 25% dependability indicates that the results are still below the average. Detailed results as in table 3.

Table 3. Measured Scale Mans UEQ

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	1,27	Above average	25% of results better, 50% of results worse
Perspicuity	0,85	Below Average	50% of results better, 25% of results worse
Efficiency	1,12	Above Average	25% of results better, 50% of results worse
Dependability	1,13	Below Average	50% of results better, 25% of results worse
Stimulation	1,35	Good	10% of results better, 75% of results worse
Novelty	0,81	Above Average	25% of results better, 50% of results worse

If grouped into three major parts of the 6 components of the assessment as shown in table 4 below.

Table 4. UEQ Results

Category	UEQ Scale Mean Value	UX Aspect	UEQ Scale Value
Attractiveness	1,27	Attractiveness	1,27
		Perspicuity	0,85
Pragmatic Quality	1,03	Efficiency	1,12
		Dependability	1,13
Hedonic Quality	1,08	Stimulation	1,35
		Novelty	0,81

Based on the results of processing in table 4, the highest value is the attractiveness aspect, which is 1.27 because this system is attractive and comfortable for students so the value of this aspect has a positive impression and the highest of the 3 aspects. In pragmatic quality, the highest is dependability, which is 1.13. This shows that the interaction system is appropriate for students who want to learn about competitive programming, as well as the features they have to support the learning process. The efficiency aspect received a positive impression, this shows that students in completing training and submission tasks can be done quickly on the TLX Training Gate system.

The test results obtained the average value of the hedonic quality aspect is 1.08. Stimulation and novelty are included in aspects of hedonic quality. The result of the calculation of the value of the stimulation aspect is 1.35. This shows that the TLX Training Gate system is useful for students and motivates them to use it. The UEQ value for novelty is 0.81 which indicates the design of this system is novel.

Based on the results of this study, it was found that the evaluation results for all aspects were still said to have positive values, where all UX aspects scored above 0.8. This is in accordance with research that has been carried out [14] which states that the average value of impressions between -0.8 and 0.8 is a normal evaluation value, a value > 0.8 is a positive evaluation, while values < -0.8 is a negative evaluation. From all aspects assessed, the average score for the novelty aspect is lower than the other aspects. This is because according to users, namely students, it is still monotonous and common. This is in accordance with the results of the monotonic factor test and usually has a normal impression value, so it requires efforts to increase this value by making more creative and innovative designs. The average value of the measurement results for each variable from the novelty aspect is as shown in table 5.

Table 5. Novelty Aspect Mean Score

Mean	Variance	Std. Dev	No.	Left	Right	Scale
⇨0.6	3.2	1.8	36	Creative	Dull	Novelty
⇨0.3	2.6	1.6	36	Inventive	Conventional	Novelty
⬆0.9	1.5	1.2	36	Usual	Leading Edge	Novelty
⬆1.4	0.8	0.9	36	Conservative	Innovative	Novelty

Each average value of the UEQ calculation results is compared with the benchmark values contained in table 1 to find out the meaning of each value. The results of the UEQ scale measurement are compared with the benchmark value as shown in Figure 6.
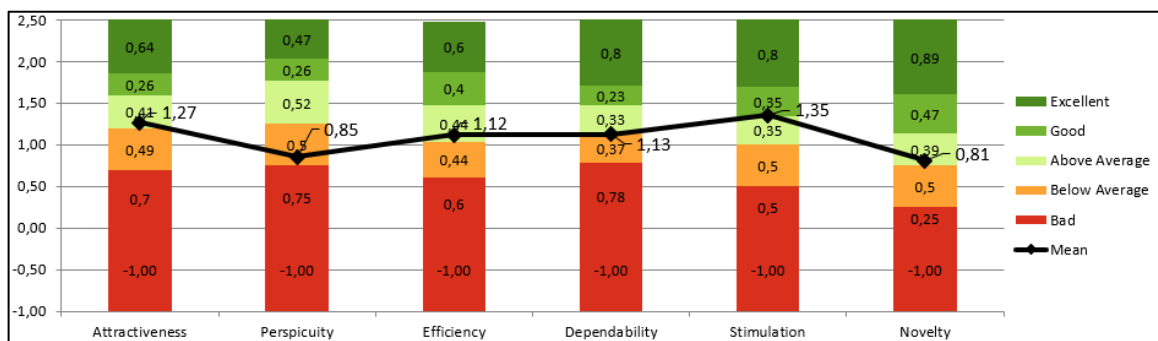

Figure 6. Comparison Results with Benchmark Scale

Based on the comparison results of the benchmark scale, the attractiveness, efficiency, and novelty aspects are in the above average category, and the stimulation aspect is in the good category. Meanwhile, aspects of perspicuity and dependability are in the below average category. These results indicate that efforts need to be made to improve the aspects of perspicuity and efficiency, so that students who learn are more motivated and can be used with fun.

Furthermore, the calculation of respondents' assessment results through the System Usability Scale approach, where the final scale value is obtained based on the rules of the SUS Score Calculation method. There are two ways to grade the results of the SUS approach, which can be used, namely in terms of the level of user acceptance, grade scale and rating adjectives and the SUS percentile rank score. Brooke's assessment of SUS can be seen in figure 5.

Based on the provisions for determining the results in the SUS assessment, the percentile rank score means that the results of the respondents' assessment in the TLX Training Gate application of 75 (seventy-five) are in grade C and if you look at the process of determining the results of the research that has been carried out using acceptability, grade scale, adjective rating, looking at figure 5, the user acceptance rate is entered. in the acceptable category, the scale grade level is in the C category, and the rating adjective is in the good category.

5. CONCLUSION AND SUGGESTION
5.1. Conclusion

Based on two approaches to determining research results, have been carried out either by UEQ or SUS, it can be seen the difference in the final results of the assessment. This condition is caused by different points of view in the process determination of assessment results. Approach with UEQ or SUS can be used as an

evaluation tool for assessing software interfaces that are measured and structured accurately. There are various ways to determine the results of the assessment evaluation.

The measurement results for each aspect obtained the attractiveness value of 1,27, perspicuity of 0,85, efficiency of 1,12, accuracy of 1,13, stimulation of 1,35 and novelty of 0,81. All aspects get a positive impression and the SUS value is 75, it is still considered acceptability.

5.2. Suggestion

In the novelty aspect, there are factors that need attention, namely monotonous and ordinary which have normal impression values. There is a need for improvement in aspects of attractiveness, pragmatic quality (perspicuity, efficiency, accuracy), and hedonic quality (stimulation and novelty) so that the TLX Training Gate system used by students is better.

REFERENCE

[1] C. L. Hanselman and G. Chrysanthos. "A Mathematical Optimization Framework for The Design of Nano Patterned Surfaces.", *American Institute of Chemical Engineers AICHE Journal*, volume 62, issue 9, 2016, pages 3250-3263.

[2] HIMAKOM UGM. (2020). Competitive Programming. [On-line]. Available: https://omahti.web.id/divisi/competitive-programming [July 30, 2022].

[3] Ikatan Alumni TOKI. (2015). TLX Competitive Programming Training Gate. [On-line]. Available: https://tlx.toki.id/training [July 30, 2022].

[4] J. Brooke. "SUS: A Retrospective." *Journal of Usability Studies*, volume 8, issue 2, 2013, pages 29-40.

[5] U. Ependi, T. B. Kurniawan, F. Panjaitan. "System Usability Scale Vs Heuristic Evaluation: A Review." *Jurnal Simetris*, volume 10, issue 1, 2019, pages 65-74.

[6] M. Schrepp, M. P. Cota, J. Thomaschewski, dan R. Goncalves. *User Experience Questionnaire handbook.* 2019.

[7] S. Elisurya, H. M. Az-Zahra, N. H. Wardani. "Evaluasi Pengalaman Pengguna Menggunakan Usability Testing dan User Experience Questionnaire (UEQ) (Studi pada E-Commerce Fashion)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK),* volume 3, issue 5, 2019, pages 4327-4332.

[8] S. P. Kritanto, L. Hakim, F. Hariyati. "Usability Evaluation In Ruang Guru Applications Using User Experience Questionnaire (UEQ).", *Jurnal Mantik*, volume 4, issue 1, 2020, pages 181-186.

[9] S. R. Henim, R. P. Sari. "Evaluasi User Experience Sistem Informasi Akademik Mahasiswa pada Perguruan Tinggi Menggunakan User Experience Questionnaire." *Jurnal Komputer Terapan*, volume 6, issue 1, 2020, pages 69-78.

[10] U. Ependi, F. Panjaitan, Hutrianto. "System Usability Scale Antarmuka Palembang Guide Sebagai Media Pendukung Asian Games XVIII" *Journal of Information System Engineering and Business Intelligence*, volume 3, issue 2, 2017, pages 101-107.

[11] B. J. Pudjoatmodjo and R. Wijaya. "Tes Kegunaan (Usability Testing) pada Aplikasi Kepegawaian Dengan Menggunakan System Usability Scale (Studi Kasus: Dinas Pertanian Kabupaten Bandung)" in *Semnasteknomedia STMIK Amikom Yogyakarta*, 2016.

[12] J. J. Garet. "The Elements of User Experience: User Centered Design for The Web Beyond", 2nd. USA: New Riders. 2011.

[13] N. Bevan. "What is the difference between the purpose of usability and user experience evaluation methods" in *Proceedings of the Workshop UXEM,* volume 9, 2014.

[14] M. Rauschenberger, M. Schrepp, M.P. Cota et al. "Efficient measurement of the user experience of interactive products- How to use the User Experience Questionnaire (UEQ)" *International Journal of Interactive Multimedia and Artificial Intelligence*, volume 2, issue 1, 2013, pages 39-45.