ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 440
International Conference on Online and Blended Learning 2019 (ICOBL 2019)

System Usability Scale Measurement on Synhcronous Online Argumentation Learning System

Ayu Nuswantari
Graduate Network Learning Technology
National Central University
Taoyuan City, Taiwan
Instructional Technology
Yogyakarta State University
Yogyakarta, Indonesia
ayunuswantari.2017@student.uny.ac.id

Ying-Tien Wu
Graduate Network Learning Technology
National Central University
Taoyuan City, Taiwan
ytwu@cl.ncu.edu.tw

Herman Dwi Surjono
Instructional Technology
Yogyakarta State University
Yogyakarta, Indonesia
hermansurjono@uny.ac.id

Abstract*—Usability measurement on a system is an essential aspect to meet the user's needs. To resolve the user's problem and fill their needs, should be done in the effective way. There are many measuring instruments to examine the usability aspect in a system, but not many of them are effective and efficient. To measure usability perception in effective and efficient way John Brooke invent System Usability Scale (SUS) in 1986. The SUS was used to measure the usability aspect of electronic office systems at that time, but it's now applied on a more variety way, such on website or technology-based applications to measure how easy or difficult they are to use in order to improve the system. In this research, System Usability Scale was used to measure the usability aspect on a synchronous argumentation system. The participants were forty-two tenth grade students of Chi-Ying Senior High School in Taiwan. The result shows the score is (78.45) which means it is in "excellent" rank from six of the SUS grading key. Based on the result, the system's usability is high and qualified to use. This paper also proposed how to improve the system based on the SUS result.

***Keywords*: system usability scale, online system, usability, measurement, web-based system**

I. INTRODUCTION

The usability testing is an experimental way, the purpose of it is to check user attributes of the product or system and to identify the problems faced by the user during the system's application. The principal goal of the usability testing is to measure the product's usability in certain aspects such as efficiency, productivity and also satisfaction to develop a really high standard, useful, and desirable product. The advantage of usability testing compared to usability expertise is the usability testing discovers user's attributes of the product or system together with some indices calculated during the course of the user's interaction with the system. According to Niranjanamurthy's research [10], the importance of usability testing is high, and plays important role in the process of design. It can help to improve the system to be more acceptable by the users and gives significant impact in designing and delivering the product or system to the end-user. The usability testing can help the developers to discover potential bugs and other limitations in a system, which generally are not really visible to the developers themselves.

The measurement tools for the usability testing does the essential role to help the developer team to analyze the user's perception and experience on using the system which they have made in order to improve their product or system. In this paper, we use System Usability Scale (SUS) to measure the usability aspect on an online argumentation learning system. We didn't change the instruments and only adapt the instrument the way it is. This also supported by Lewis and Sauro research [9], the result proved that the System Usability Scale have an extremely high correlation (0.985), which means that practitioners and other researchers can continue to use the SUS as it is.

II. LITERATURE REVIEW

A. System Usability Test (SUS)

The System Usability Scale (SUS) is a ten-item attitude of Likert scale giving a global view of subjective assessments of usability which was developed by John Brooke in 1986 [1]. The scales are strongly agree, agree, neutral, disagree, and strongly disagree, which indicate the participant or the user's perception of the system. The SUS already have formula and grading key, which make it easier to be implemented in a research.

Other advantage from using SUS is, the SUS has been widely used in the evaluation of a range of various systems around the world and proven to be reliable and valid. As for the reference, other researcher like Bangor, Kortum and Miller (2008) have used this scale extensively over a ten-year period and also have produced normative valid data that allow SUS ratings to be positioned relative to others. Their research proposed an extension to the System Usability Scale to provide an adjective rating that correlates with a given score. The other researcher like Sauro and Lewis [9] also used the SUS and reviewed hundreds of usability studies, summarized them then proposed a curved grading scale for the average of SUS scores.

Other reason for using SUS is because it's reliable. The reliability refers to how consistently the user's

responds to the items or repeatability of the responses. The SUS has been shown to be more reliable and to detect differences at smaller sample sizes than other questionnaires and commercial tools. The SUS can be used on very small sample sizes such as only two samples and still generate reliable results. But the researcher should consider that small sample sizes generate imprecise estimates of the unknown user-population of SUS score. The researcher should compute a confidence interval around the samples and SUS score to understand the variability in the research estimate.

The SUS is proven to be valid. The validity refers to how well something can measure what it is intended to measure, for example in this case is perceived usability. The SUS has been shown to effectively distinguish between unusable and usable function in systems as well as or better than other questionnaires. The SUS also correlates highly with other questionnaire-based measurements of usability which is called concurrent validity. At only 10 instrument items that the SUS may be quite fast to administer and scorer, but the data of more than 5000 users and almost 500 different studies have suggested that SUS is far from dirty measurement. The SUS questionnaires are;

- I think that I would like to use this system frequently,
- I found the system unnecessarily complex,
- I thought the system was easy to use,
- I think that I would need the support of a technical person to be able to use this system,
- I found the various functions in this system were well integrated,
- I thought there was too much inconsistency in this system,
- I would imagine that most people would learn to use this system very quickly,
- I found the system very cumbersome to use,
- I felt very confident using the system,
- I needed to learn a lot of things before I could get going with this system.

The formula for scoring SUS;

Devine the data into two types, the odd items and even items. The odd items are measuring the positive statement of using the system which indicate that the system is easy to use, the numbers are; 1, 3, 5, 7, 9. The even items are measuring the negative or indicate that the user is not satisfy with the system and the system is hard to use, the numbers are; 2, 4, 6, 8, 10. The numbers are 5 for strongly agree and 1 for strongly disagree. Add up the converted responses for each user and multiply that total by 2.5. This converts the range of possible values from 0 to 100 instead of from 0 to 40. The average SUS score from various researches from about all 500 studies is about 68. The SUS score above 68 would be considered above the average and anything below 68 is below the average.

B. Online System

Online System or also known as NLS was actually revolutionary computer collaboration system which was developed in the 1960s by Douglas Engelbart and also implemented by researchers at the Augmentation Research Center (ARC) at the Stanford Research Institute (SRI). The online system was the first to introduce the practical use of hypertext links, raster-scan video monitors, screen windowing, and gain information organized by relevance, presentation programs, and many other modern computing concepts. NLS was also funded by the Defense Advanced Research Projects Agency, the U.S. Air Force, and NASA.

One of Online System most revolutionary features is the Journal, and it was developed in 1970 by an Australian computer engineer named David A. Evans as part of his doctoral thesis [7]. At that time, The Journal was a quite old style hypertext-based groupware program, which can be seen as a predecessor of all contemporary server software that could supports collaborative document creation such as wiki and any other document creation. The Journal was used by ARC members to discuss topic, debate opinion, and refine the concepts. Meyer [2] add statement that The Journal was used to store documents for the Network Information Center (NIC) and early network email archives at that time, and most Journal have been preserved in paper form or hard copy form and stored in Stanford University's archives. The Online or NLS was implemented using several domain-specific languages implemented with the Tree Meta compiler-compiler [4]. L10 was the eventual implementation language that was used that time [4]. For today's use, some of the full-interaction paradigm lives on in different systems and including the Hyperwords add-on for Mozilla Firefox till now. The Hyperwords concept was actually grew out of the Engelbart web-documentary Invisible Revolution [6]. The aim of the project was to allow the users to interact with all kind of words on the Web, not only the links.

In this research, the researcher use a system based on website or online. Which means that this system should be supported with computer and internet access. This system can be access anywhere and anytime as long the server is open and the internet connection is still working.

C. Argumentation Learning System

In an argumentation learning system, the characteristic is it has an argument map on its interface. In informal logic and philosophy explained that an argument map or often called argument diagram is a visual representation of the structure of an argument activity in a group. An argument map includes the essential components of the argument called the conclusion and the premises and also called contention and reasons [5]. In an argument map can also show co-premises, counterarguments, objections, rebuttals, and also lemmas. There are many different styles of argument map but those argument maps are often has functionally equivalent and represent an argument from individual or person's claims and the relationships between one to another.

Throughout the time on the 1990s to 2000s founded that many other software applications were developed for argument visualization activity. More than 60 such software systems existed by 2013 [13]. Whether collaboration is supported or not is one of the differences between these software systems [12]. Single-user argumentation systems which include Convience Me,

iLogos, LARGO, Araucaria, Carneades and Athena; small group argumentation systems include Digalo, Compendium, Belvedere, QuestMap, and AcademicTalk; community argumentation systems include Collaboratorium and Debategraph [12].

Argument maps have been applied in many areas for different purposes, most in educational, academic purpose and also business settings, including on design rationale [8]. In this research the argument map is used for educational purpose on argumentation learning activity. In this research, the online argumentation learning system was designed by the developer using principles and has two main parts on the user interface, which are the argument map and the argument chat activity. The system also has other functions that support the activity, such as argument history, and argumentation evaluation function that could record the student's argumentation learning activity.

III. METHOD

A. Research Subject

The participants were forty-two tenth grade Japanese language major students of Chi-Ying Senior High School in Taiwan. The reason for choosing this school is because based on the observation done by the researcher, the students are suitable to be the research subject based on the criteria such as familiar with computer and website, familiar with argumentation activity, and ever had experience in using other similar system. The experiment was conducted during IT lesson, and the data was taken on 22th April 2019.

B. Research Instrument

The instruments that used in this research are ten instruments of The System Usability Scale (SUS) for quantitative data. And also for other deep understanding about the perception of the participants, the researcher also used three questions to get qualitative data for strengthening the result which are also based on SUS, the following questions are: (1) What do you think about the system interface compared to the other system? (2) Will you use this kind of system in the future? (3) What do you think makes the system different from other system? And which one do you prefer?

C. Data Collection and Analysis

The data was collected after the participants tried the system. The data was collected with the ten instruments of The System Usability Scale (SUS) and the researcher took three students for sample to be interviewed and asked three questions with random sampling technique. The quantitative data of SUS instruments was analyzed with the SUS formula. And the grading technique used the SUS grading key. The qualitative data was analyzed by categorize the data into "positive perception" and "negative perception" towards the system. And for identify the function that can be improve from the system according to the SUS result, the researcher also did analyze the each of the items based on the average score. With these data the researcher can determine whether the result from the SUS and interview connect with each other and can get better understanding about the participant's perception.

IV. IMPLEMENTATION

A. Research Design Framework

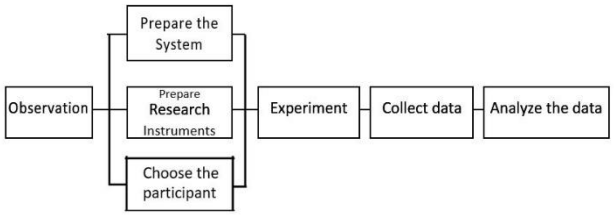

Fig. 1. Research Framework

The process of this research is by preparing the system that will be used on participants and prepare the instruments which is SUS instruments and select the participant based on the requirements. The experiment conducted and then the researcher collected the data from the participants using the instruments. After the data collected, then the researcher analyzed the data.

B. Activity

The forty-two tenth grade students of Japanese class in Chi-Ying Senior High School were separated into argumentation groups. Each group has two students in it and they did the argumentation activity using the system for thirty minutes. Before the activity start, the researcher showed short introduction video about the system to the students and also explain what kind of activity they would do.

In this activity, the students gain experience of using the system from the user interface and the user experience perspective. The researcher was there and observe their behavior and also help the students if they need.

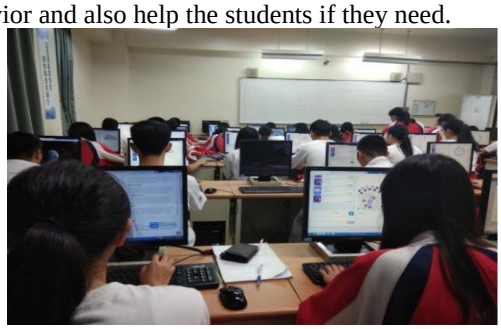

Fig. 2. Students using the system

The researcher also took sample from students output in the activity on the system screenshot.

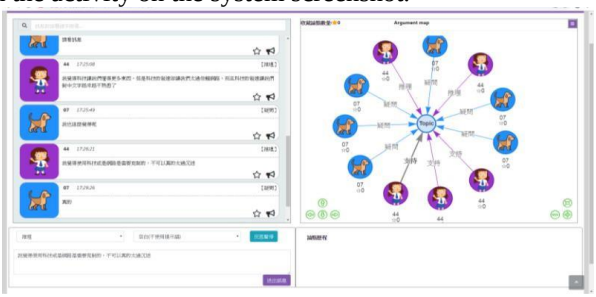

Fig. 3. One group student's activity output

The output of the activity with the system shows that students keep engaging with the topic and keep on discussing within thirty minutes. This activity shows that the students focus on the topic and discuss more

about the topic by build on the topic based on their knowledge and perception. The argumentation topic was "English Language is Important/Learning English is important".

V. RESULT

A. The SUS Result on The System

TABLE 1. SUS RESULT OF TENTH GRADE STUDENTS TESTED WITH THE SYSTEM

SUS Score (n = 42)	Average SUS Score (odd items + even items) x 25
The System	78.45

TABLE 2. SUS GRADING KEY

Scale	Grading SUS Key
86 – 100	Best Imaginable
74 – 85	Excellent
53 – 73	Good
39 – 52	Ok
26 – 38	Poor
1 - 25	Worst Imaginable

According to the System Usability Scale (SUS) result, proved that the average score for the system is (A.SUS= 78.45) from forty-two data that analyzed with SUS formula (odd items + even items) * 2.5. The result shows that the score is above 68 which means that the score is above average, and it shows that the score based on the grading SUS key is on the "Excellent" rate.

B. The Interview Result

TABLE 3. INTERVIEW OF TENTH GRADE STUDENTS TESTED WITH THE SYSTEM

Question	Student A	Student B	Student C
What do you think about the system interface compared to the other system?	I think the diagram of the system picture is **more clearly than other system.**	I think this system is **easier to read *than*** other system.	The system is quite similar with other system but I think **the pictures of this system are easier to show.**
Will you use this kind of system in the future?	In the future, **I will use it** if I need.	**I will consider using it** if I need to use it to discuss a topic with others in the future.	I will consider using it if **I will use it** to discuss it with others in the future.
What do you think makes the system different from other system? And which one do you prefer?	The other system is hard to match to people. It is difficult to identify whether it matches people. The picture is not as good as the system. This system is **easy to use** and can clearly indicate the matching person ID.	The system is quite similar with other system but I prefer this system because it is **easier to pair and use.**	The systems are very similar, but I prefer this one. This system personally, because it is **easier to use than the other system.**

The result of the interview shows positive perception toward the system. The indicators are; the participant think the system is easier to use than other system they ever experience with, and the participant would like to use this kind of system in the future for online argumentation learning activity. This finding support the SUS result (A.SUS= 78.45) which means the system is really give high usability aspect for the user in using the system.

C. The SUS Average Score of each Items

TABLE 4. AVERAGE SCORE OF EACH ITEM IN SUS

SUS Item	XI (positive)	X2 (negative)	X3 (positive)	X4 (negative)	X5 (positive)
Average Score	4.92	1.26	4.78	1.30	4.73
SUS Item	X6 (negative)	X7 (positive)	X8 (negative)	X9 (positive)	X10 (negative)
Average Score	2.14	4.80	3.02	4.76	4.90

According to the result, the item which shows positive perception towards the system (X1, X3, X5, X7, X9) are all high with the range of scale from 1 (strongly disagree) to 5 (strongly agree), they are all above 4.0, which indicate that the positive perception is high towards the system. The items that shows negative perception (X2, X4, X6) also show good result, with average score bellow 3.0 which indicate that the negative perception of the subject in the three items are low, and indicate good perception. However, the other items for negative perception (X8, X10) shows that the average score are above 3.0, which indicate that the positive perception toward the statement on these items are low, especially on item (X10).

Item X8; I found the system very cumbersome to use.

Item X10; I needed to learn a lot of things before I could get going with this system.

Both of the items are related to ease of use and the current skill and knowledge of the participants. The participant had perception that the system is cumbersome to use and they need to learn a lot of things before they could get going with the system. This finding also shown that the participants need more time to adapt with the system and some of the students are novice user in using this kind of system.

VI. CONCLUSION

The SUS measurement that has been used in this research was gave a fast, efficient and effective result. The result from using the SUS on the online argumentation system is that the system is good to use and has high usability average score at (78.45), and the grading for the system is "excellent" and the system is proved to be good to use in argumentation learning activity. The interview result shows that the participant's perception of the system is good. However, from the average score of each item shows that two items (X8, X10) are need to be improved to get better result and to improve the system based on the user's need.

VII. RECOMMENDATION FOR FUTURE PROJECT

A. For the User Interface and User Experience

According to the result of SUS average score for each items, the developer could try to improve the interface to meet user needs especially for the novice user. For example, the novice user need more guide, so it's better to use simpler interface but more essential function. Keep the text pop up when the mouse hit the button, and keep the tour to introduce the system before give the student a task using the system. Use one function prompt, for example if the student wants to propose new idea, the student only need to choose one action and when they want to post their idea then the prompt will automatically show when the student start typing (like in *facebook* or *email*), because the prompt that show as a bar is not often used when the student want to make a comment or propose idea. Make the step less to do an action, to make it more efficient, for example, the student doesn't have to do two steps to post a comment, etc.

B. For the Implementation

Based on the conducted research, it's highly recommended to planning ahead with the framework. It's better to give enough information about the system and also the SUS instruments before the experiment begin. The implementation should be appropriate to be implemented at school and the participant that would be the subject for the research, so do the user observation is highly recommended.

C. For The SUS Usage

It's highly recommended to keep on up to date of this System Usability Scale related research to get better understanding and also get newer concept that could be applied related with this usability scale. For the future research, the researcher suggested to test and do research about each items of the SUS could be study farther and deeper to improve the measurement scale itself. So the result would give more reference for the next future research.

REFERENCES

[1] Brooke, J. (1986). *"SUS: a "quick and dirty" usability scale"*. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (eds.). Usability Evaluation in Industry. London: Taylor and Francis.

[2] D. Meyer (July 31, 1973). *"Network Journal Submission and Delivery"*. RFC 543. Augmentation Research Center. Retrieved April 19, 2011.

[3] Douglas C. Englebart; Richard W. Watson; James C. Norton (June 4–8, 1973). *"The Augmented Knowledge Workshop"*. Proceedings of the National Computer Conference and Exposition: 9–12. CiteSeerX 10.1.1.729.1832. doi:10.1145/1499586.1499593. Retrieved April 20, 2011.

[4] Engelbart, D., *"Study for the development of Human Augmentation Techniques"*. Final Report, July 1968. Sections 4 and 5.

[5] Freeman, James B. (1991). *Dialectics and the macrostructure of arguments: a theory of argument structure*. Berlin; New York: Foris Publications. ISBN 978-3110133905. OCLC 24429943. Retrieved 2016-02-24.

[6] Frode Hegland and Fleur Klijnsma. *"Invisible Revolution"*. Web documentary. London. Retrieved April 13, 2011.

[7] It is important to not confuse Dr. Evans with the numerous other persons who share the same name. He was Managing Director and CEO of MRI magnet startup Magnetica and participated in the 1998 symposium honoring Engelbart's work. Archived July 18, 2008, at the Wayback Machine

[8] Kirschner, Paul Arthur; Buckingham Shum, Simon J; Carr, Chad S, eds. (2003). "*Visualizing argumentation: software tools for collaborative and educational sense-making. Computer supported cooperative work".* New York: Springer. doi:10.1007/978-1-4471-0037-9. ISBN 978-1852336646. OCLC 50676911. Retrieved 2016-02-24.

[9] Lewis, R. James and Sauro, Jeff. (2009). *"The Factor Structure of the System Usability Scale"*. IBM Software Group & Oracle, 1 Technology Way, Denver.

[10] Niranjanamurthy, M. Archikam, Nagaraj, Himaja, Gattu, & Puneeth K Shetty. (2014). *"Research Study on Importance of Usability Testing/ User Experience (UX) Testing"*. International Journal of Computer Science and Mobile Computing. p. 78-85.

[11] Sauro, J., & Lewis, J. R (2012). *"Quantifying the user experience: Practical statistics for user research".* Morgan Kaufmann, Waltham MA, USA.

[12] Scheuer, Oliver; Loll, Frank; Pinkwart, Niels; McLaren, Bruce M. (2010). *"Computer-supported argumentation: a review of the state of the art"*. International Journal of Computer-Supported Collaborative Learning. 5 (1): 43–102. CiteSeerX 10.1.1.322.2522. doi:10.1007/s11412-009-9080-x

[13] Walton, Douglas N. (2013). "*Methods of argumentation"*. Cambridge; New York: Cambridge University Press. doi:10.1017/CBO9781139600187. ISBN 978-1107677333. OCLC 830523850. Retrieved 2016-02-24.