Video conferencing as a face-to-face online meeting app: user preference based on usability testing

Astrid Wahyu Adventri Wibowo[*], Berty Dwi Rahmawati, Hasan Mastrisiswadi

Department of Industrial Engineering, Universitas Pembangunan Nasional Veteran Yogyakarta, Jl. SWK 104 Condongcatur, Yogyakarta 55283, Indonesia

*Corresponding Author

Astrid Wahyu Adventri Wibowo
E-mail: astrid.wahyu@upnyk.ac.id

A B S T R A C T

The global pandemic of Corona Virus Disease (Covid-19) that hit Indonesia since March 2020 has changed the face-to-face system from offline to online. Video Conference (VC) becomes an alternative choice for delivering material, both learning and working. VC is a technology that allows users to hold face-to-face meetings at their respective places simultaneously. Various VC apps are becoming increasingly popular these days, such as Google Meet, Zoom, Youtube, Webex, Skype, GoTo Meeting, and Big Blue Button (BBB). This study is intended to provide a usability test of VC applications (such as Zoom, Google Meet, and BBB) and provide recommendations for VC as an online conference media based on user preferences. The usability measurement technique used is the System Usability Scale (SUS) and the USE questionnaire. The results of this study show that Zoom has the highest usability value compared to Google Meet and BBB, Google Meet is ranked second and BBB is ranked third.

1. INTRODUCTION

Coronavirus disease (Covid-19), which was announced to have entered Indonesia in March 2020, has changed the order of life that usually has been running. Covid-19, which spreads through human contact, becomes difficult to predict along with various interactions among humans. With this phenomenon, the World Health Organization (WHO) has officially designated Covid-19 as a global pandemic and established social distancing rules to stop the spread of the virus. This change has a broad impact in various sectors, such as tourism, economy, socio-culture, to education; the article is that almost all countries enforce regulations not to do activities outside the home.

Like it or not, the face-to-face system must change from offline to online, whether it's about work or study.

Video Conference (VC) is an alternative choice to answer the challenges of the online face-to-face system. VC is a technology that allows users to hold face-to-face meetings at their respective places simultaneously. Various video conferencing applications, such as the Google Meet application, Zoom, Webex, Skype, Go To Meeting, and the Big Blue Button (BBB), are becoming increasingly popular. And the technology offered is changing rapidly, following user needs. Video conferencing equipment manufacturers is now offering the flexibility of

communication services for conference rooms, desktops, and mobile terminals that can accommodate a wide range of groups and situations [1].

With so many video conferencing technologies being offered, this research focuses on which application provides the highest usability value. According to ISO 9241:11, usability explained as how a product can be usefull for specific users to accomplish the objective effectively, efficiently and obtain pleasure in the context of its users [2]. This problem will be answered by evaluating the usability, ease of access and credibility of various applications that focus on online face-to-face activities. In addition, this study aims to determine user preferences and characteristics in choosing VC applications as online face-to-face media.

The usability testing model used in this research is the System Usability Scale (SUS) and the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire, both of which have been widely used to assess the usability level of a system, product, and website. John Broke construct SUS for DEC's Integrated Office Systems Group (IOSG) in Reading, UK [3]. IOSG (the main product was ALL-IN-1) was a system which maintain integrated email, task and time management, word processing, and also improve environment for another office applications according to VAX/VMS systems.

The SUS has became to be a very malleable questionnaire [4]. A diversity of research have proven of the validity and sensitivity of the SUS. There were major differences in SUS ratings for each types of products [5]. For example, Excel had low SUS scores; otherwise Gmail had a high SUS scores. In 2015, the researchers discovered variation of SUS rating for iOS and Android (mobile operating systems) and types of devices (phones and tablets) [6]. SUS scores are sensitive to successful assignment completion, with those completing assignments successfully producing greater scores [7]–[9]. A study found a significant contact among a composite metric and SUS scores according to business indicators of success in the marketplace [10]. Other researchers have examined the validity, sensitivity, and reliability of the SUS. This research has persistently shown the SUS to have reliabilities at or just over 0.90 [11]–[13], which outstrip the typical criterion of 0.70 for measurements of sentiments [14].

The SUS model only covers aspects of satisfaction and learning ability. So in this study,

the USE questionnaire was also applied because it accommodates aspects of effectiveness and efficiency [15]. The USE questionnaire was developed to gain information applicable across domains [16]. Several studies related to the USE questionnaire have been carried out, including a user evaluation study of three non-invasive measuring heart rates [17]. There is also research that aims to explore the validity of tangible augmented reality (TAR) and augmented reality (AR) as a tool to evaluate the usability of a product [18].

The usability evaluation of video conferencing applications has been carried out, such as the Virtual Conference Center (VCC) application which has a high level of reliability and a high level of satisfaction [19]. An usability assessment to analyze usability errors and to calibrate user satisfaction with educational systems or software from the perspective of usability engineering discipline, according to a case of a Danish university's VCS use for the synchronous class, has been held [20]. Based on the literature study that has been carried out, there has been a usability evaluation of the video-conferencing application. However, not all applications have been evaluated. For this reason, this research will find out the usability evaluation of three VC applications, namely Zoom, Google Meet, and BBB.

2. RESEARCH METHODS

The type of research used is evaluation research. This research is used to develop a system, with the limitation that only Google Meet, Zoom, and BBB will be evaluated for usability. So, this study is limited to respondents who have used these three applications. Later, this research will put forward the results of the evaluation of the advantages and disadvantages in system usage and user preferences for various VCs based on their utility value.

A quantitative approach is also applied to this study. The data collection will be converted into numerical form to conclude the numerical data. There are two types of data sources: primary data and secondary data. The primary data were obtained from the SUS and the USE questionnaire, while the secondary data were pictures, documents, archives, and literature directly related to this research.

The first stage of the research was determining respondents. Respondents in this study were active Zoom, Google Meet, and BBB

applications (expert respondents), where the questionnaires will be distributed online and randomly. Most of the respondents were lecturers/teachers and students/students, most of whom use VC for schools, lectures, and teaching. The respondent's characteristics are further described in Table 1. The second research step was compiling usability test scripts based on SUS and USE questionnaires. After that, respondents carried usability testing, and data analysis from the questionnaire results was done.

Table 1. Characteristics of respondents

Characteristics of respondents	Zoom	Google meet	BBB
Number of respondents (person)	53	48	13
Male (person)	25	22	7
Female (person)	28	26	6
Age (mean ± SD years)	24,9 ± 4,7	24,8 ± 4,7	23,1 ± 4,5
Lecturer/teacher (person)	20	19	4
Private employee (person)	1	1	0
Office worker (person)	1	0	0
Student (person)	28	27	9
Others (person)	1	1	0
Entrepreneur (person)	2	0	0

2.1. *System usability scale* (SUS)

The System Usability Scale (SUS) is a reliable "quick and dirty" tool for measuring usability. This model has ten questions with five response options for respondents, from strongly agree to disagree strongly [20]. The SUS model was chosen because it has different characteristics from other questionnaires; it has been validated and tested for reliability even with a small sample value [3]. SUS consists of ten questions, so it is relatively rapid and simple for respondents to complete. SUS uses technology agnostic, which means it can be used widely and evaluates almost all types of interfaces [21]. The questionnaire results are single scores, ranging from 0 to 100 scores, and are relatively easy to understand by various disciplines, both individuals and groups [21]. To determine the SUS score, first, sum up the score contribution of each item. The contribution score of each item will range from 0 to 4. For items with odd numbers (1,3,5,7 and 9), the contribution score is a scale minus 1. While for even-numbered

items (2,4,6,8 and 10), the contribution score is five minus the scale. The overall SUS score is obtained by multiplying the total score by 2.5.

2.2. *USE questionnaire*

The USE Questionnaire is a form of question-naire to assist in the measure of products and services [22]. This questionnaire is also non-proprietary, which means that anyone can access it free of charge. The USE contains 30 items [15]. These items belong to 4 dimensions: usefulness, ease of use, ease of learning, and satisfaction. This questionnaire is filled using a 5-point Likert scale. Ranging from strongly disagree to agree strongly. Usability measurement is done by calculating the maximum score of the scale and the observation score for each dimension of the question with equations 1 and 2 [23].

$$Score_{max} = N \; x \; nbv \; x \; scale_{max} \qquad (1)$$

$$Score_{observe} = \sum_{k=0}^{N} \sum_{l=0}^{30} scale \qquad (2)$$

Where N: number of respondents, Nbv = number of valid item, Scale$_{max}$: max scale, Score$_{max}$: max score, Scale: the scale that filled in by respondent and Score$_{observe}$: observation score.

3. RESULTS AND DISCUSSION

3.1. Usability test results using SUS

After the respondents have filled out all the SUS questionnaires, the next step is to calculate the SUS score of each respondent and then look for the average value. The average value is then determined based on the results of the assessment. There are three ways to determine the level of assessment results (Fig 1) [3]. First, user accep-tance can be categorized into unacceptable, marginal, and acceptable based on user accep-tance. The second determination is based on the class scale, divided into six scales: A, B, C, D, E, and F. The third determination is based on the adjective rating, with six categories: worst imaginable, bad, ok, good, very good, and the best imaginable.

Evaluation of the Zoom application usability using the SUS model got an average SUS score of 73.63. Based on the provisions of Fig. 1, the results of the assessment of the Zoom application are as follows, the level of user acceptance is in the acceptable category, the determination of the grade scale is in the C category, and the adjective

rating is in a good category. As for the evaluation of the usability of the Google Meet application, the average SUS score was 60.78. The Google Meet application assessment results based on Fig. 1 are as follows: the level of user acceptance is in the low marginal category, the determination of the grade scale is in the D category, and the adjective rating is in the ok category. And for the evaluation of the usability of the BBB application, the average SUS score is 49.04. The results of the assessment of the BBB application based on Fig. 1 are as follows, the level of user acceptance is in the not acceptable category, the determination of the grade scale is in the F category, and the adjective rating is in the poor category.
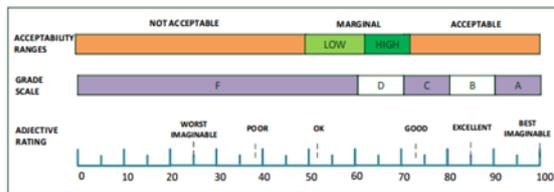
Fig. 1. Determination of the SUS scores result [24].

3.2. Usability test result using USE question-naire

In the Zoom application, the number of respondents involved is 52 people. The score for each item uses a 5-point Linkert scale. The number of valid items is 30 questions, so based on Equation (1) calculation, the maximum score is 7,800. At the same time, the observation values for each dimension are shown in Table 2.

Table 2. Measurement results with USE questionnaire for zoom app

No	Usability dimension	Number of valid items	Max score	Observation score	%	Average
1	Usefulness	8	2080	1654	79.5	3.98
2	Ease of Use	11	2860	2315	80.9	4.05
3	Ease of Learning	4	1040	858	82.5	4.13
4	Satisfaction	7	1820	1459	80.2	4.01
Total		**30**	**7800**	**6286**		
Average					**80.8**	**4.04**

The number of respondents involved in usability testing on the Google Meet application is 48 people. The assessment score for each item uses a 5-point Linkert scale. The number of valid items is 30 questions, so based on equation (1), the maximum score is 7,200. The observation values for each dimension are shown in Table 3. Lastly,

the number of respondents for the BBB app involved is 13 people. The assessment score for each item uses a 5-point Linkert scale. The number of valid items is 30 questions, then based on Equation (1) calculation, the maximum score is 1,950. The observation values for each dimension are shown in Table 4.

Table 3. Measurement results with USE questionnaire for google meet app

No	Usability dimension	Number of valid items	Max score	Observation score	%	Average
1	Usefulness	8	1920	1437	74.8	3.7
2	Ease of Use	11	2640	2001	75.8	3.8
3	Ease of Learning	4	960	744	77.5	3.9
4	Satisfaction	7	1680	1192	71.0	3.5
Total		**30**	**7200**	**5374**		
Average					**74.8**	**3.7**

Table 4. Measurement results with USE questionnaire for BBB app

No	Usability dimension	Number of valid items	Max score	Observation score	%	Average
1	Usefulness	8	520	307	59	3.1
2	Ease of Use	11	715	450	62.9	3.1
3	Ease of Learning	4	260	179	68.8	3.4
4	Satisfaction	7	455	263	57.8	2.9
	Total	**30**	**1950**	**1199**		
	Average				**62.2**	**3.2**

3.3. Discussion

After 25 years of the SUS model being published, it is still relevant to evaluate the usability of a product. From the beginning, some researchers have suggested slight changes to the wording of the elements. For example, "cumbersome" can be replaced with "awkward" [11], [25]. The original SUS elements refer to "system," but replacing the word "website" or "product," or using the actual website or product name seems not to affect the resulting scores [26]; these types of substitutions should be consistent across the items within a study). So in this study, the word "system" was replaced with the VC applications that used, like "Zoom", "Google Meet", and "BBB".

The Zoom application has the highest SUS Score compared to Google Meet and BBB, Google Meet is ranked second, and BBB is ranked third. The same thing also happened to the usability evaluation results with the USE questionnaire, Zoom was ranked first, Google Meet was ranked second, and BBB was ranked third (Table 5). The

Zoom application is more popular with respondents regarding usability.

Table 5. Comparison of VC application usability evaluation results

	Zoom	Google meet	BBB
SUS Score	73,63	60,78	49,04
USE Questionnaire	80.8	74.8	62.2

From the results of the USE Questionnaire, it can also be seen that the comparison value between the observation score and the max score for each usability dimension in the Zoom application is the highest. The Zoom application is more practical, easier to use, easier to learn, and more satisfying when compared to Google Meet and BBB. In addition, Zoom also has several advantages, namely the presence of an attractive virtual background and filter, the recording feature can work well, and it looks more user friendly.

Google Meet is included in the low marginal category, which means that respondents are quite satisfied with the presence of the application. According to some respondents, this application is accessible without a time limit, so it is an alternative to Zoom. Then, users do not need to install applications on the devices used to perform video conferencing in its use.
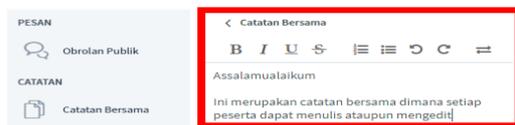
Fig. 2. Shared notes view [27]

BBB is included in the unacceptable category because respondents feel that the BBB application is more difficult to use, more difficult to learn, and unsatisfactory. BBB has almost the same features as Zoom, namely breakout rooms, polls, screen sharing, multi-user whiteboards, and so on. The low usability value of this application can also be caused because not many respondents are familiar with this application. Only 25% of the total respondents have ever operated BBB. Even though there are some features that Zoom and Google Meet don't have, the features are shared notes and create status. Shared notes can be used to create shared notes (Fig. 2). Everyone can write or edit the posts in these shared notes. And the results of these shared notes can also be downloaded in word or pdf form. BBB can also create a status (emoticon) next to the participant's

name, and it can be happy, sad, confused, raising hands, and so on (Fig. 3).
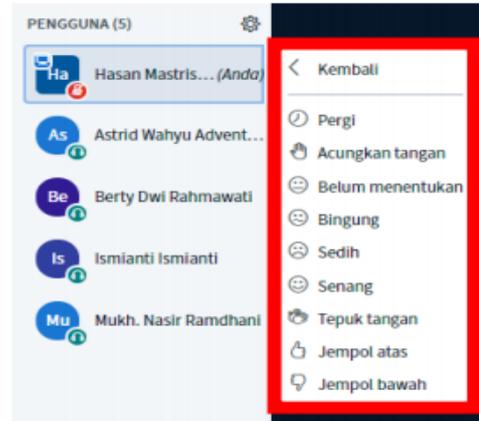
Fig. 3. Emoticon view [27]

Several other usability evaluation methods have been implemented, such as the think-aloud protocol and eye-tracking technology for usability research. Despite its value, analyzing think-aloud sessions can be onerous because it often entails assessing all users' verbalizations [28]. Then, eye tracking is a valuable method for analyzing user behavior and uncovering potential problems with a website or app design. This technique allows an in-depth understanding of how users interact with design elements, including colours, fonts, and layout. It can also help identify what users find most interesting about an app or what they miss completely. In addition to the large amount of valuable data that may be obtained, there are several disadvantages of this method. Users have to perform many actions prior to testing, such as setting up the camera and calibrating the eye-tracking software, where calibration can be performed multiple times to obtain precise eye-tracking data [29]. The researcher must calibrate the instrument for each respondent. So it can be said that these two methods, think aloud and eye tracking technology, are not worth usability testing because other methods are faster, easier, and more accurate.

4. CONCLUSION

Based on the calculation of the SUS Score and USE Questionnaire, Zoom has the highest usability value compared to Google Meet and BBB. The Zoom application is more practical, easier to use, easier to learn, and more satisfying. Google Meet is included in the low marginal category, which means that respondents are quite satisfied with the presence of the application. BBB

is included in the unacceptable category because respondents feel that the BBB application is more difficult to use, more difficult to learn, and unsatisfactory. The results of this study have positive implications for stakeholders, in this case, are video conference companies and users. Video conferencing companies know what users need; find out the usability level of each application so that it becomes the basis for future improvements. For example, BBB has a share note feature, where Zoom and Google Meet don't have it, but BBB has the lowest usability value. This result shows that something must be improved because usability is an important part of the overall user experience (UX).

SUS allows for relative judgments so that further research will add the adjective rating scale to these models. It may help practitioners interpret individual SUS scores and explain the results to non-human factors professionals. The adjective rating scale has a seven-point, adjective-anchored Likert scale that was used to determine if a word or phrase could be associated with a small range of SUS scores.

ACKNOWLEDGMENT

REFERENCES

[1] J. M. Denstadli, T. E. Julsrud, and R. J. Hjorthol, "Videoconferencing as a Mode of Communication," *J. Bus. Tech. Commun.*, vol. 26, no. 1, pp. 65–91, Jan. 2012, doi: 10.1177/1050651911421125.

[2] International Organization for Standardization, "9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)," *Int. Organ. Stand.*, vol. 45, no. 9, 1998. Available: https://www.iso.org/standard/16883.html.

[3] J. Brooke, "SUS: a retrospective," *J. usability Stud.*, vol. 8, no. 2, pp. 29–40, 2013. Available: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Brooke_February_2013.pdf.

[4] J. R. Lewis and J. Sauro, "Item benchmarks for the system usability scale.," *J. Usability Stud.*, vol. 13, no. 3,

pp. 158–167, 2018. Available: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Lewis_May2018.pdf.

[5] P. T. Kortum and A. Bangor, "Usability Ratings for Everyday Products Measured With the System Usability Scale," *Int. J. Human–Computer Interact.*, vol. 29, no. 2, pp. 67–76, Jan. 2013, doi: 10.1080/10447318.2012.681221.

[6] P. Kortum and M. Sorber, "Measuring the Usability of Mobile Applications for Phones and Tablets," *Int. J. Human–Computer Interact.*, vol. 31, no. 8, pp. 518–529, Aug. 2015, doi: 10.1080/10447318.2015.1064658.

[7] P. Kortum and S. C. Peres, "The Relationship Between System Effectiveness and Subjective Usability Scores Using the System Usability Scale," *Int. J. Human–Computer Interact.*, vol. 30, no. 7, pp. 575–584, Jul. 2014, doi: 10.1080/10447318.2014.904177.

[8] J. R. Lewis, J. Brown, and D. K. Mayes, "Psychometric Evaluation of the EMO and the SUS in the Context of a Large-Sample Unmoderated Usability Study," *Int. J. Human–Computer Interact.*, vol. 31, no. 8, pp. 545–553, Aug. 2015, doi: 10.1080/10447318.2015.1064665.

[9] S. C. Peres, T. Pham, and R. Phillips, "Validation of the System Usability Scale (SUS): SUS in the Wild," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 57, no. 1, pp. 192–196, Sep. 2013, doi: 10.1177/1541931213571043.

[10] A. Bangor, K. Joseph, M. Sweeney-Dillon, G. Stettler, and J. Pratt, "Using the SUS to Help Demonstrate Usability's Value to Business Goals," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 57, no. 1, pp. 202–205, Sep. 2013, doi: 10.1177/1541931213571045.

[11] A. Bangor, P. T. Kortum, and J. T. Miller, "An Empirical Evaluation of the System Usability Scale," *Int. J. Human–Computer Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008, doi: 10.1080/10447310802205776.

[12] J. R. Lewis, B. S. Utesch, and D. E. Maher, "Measuring Perceived Usability: The SUS, UMUX-LITE, and AltUsability," *Int. J. Human–Computer Interact.*, vol. 31, no. 8, pp. 496–505, Aug. 2015, doi:

10.1080/10447318.2015.1064654.

[13] J. C. Nunnally and J. C. Nunnaly, *Psychometric Theory*. McGraw-Hill, 1978. Available: https://books.google.co.id/books?id=arJq AAAAMAAJ

[14] A. Hodrien and T. P. Fernando, "A review of post-study and post-task subjective questionnaires to guide assessment of system usability," *J. Usability Stud.*, vol. 16, no. 3, pp. 203–232, 2021. Available: http://usir.salford.ac.uk/id/eprint/60928/.

[15] A. M. Lund, "Measuring usability with the use questionnaire12," *Usability interface*, vol. 8, no. 2, pp. 3–6, 2001. Available: https://www.researchgate.net/publication/230786746.

[16] E. Mathe and E. Spyrou, "User Evaluation of Low-Cost, Non-Invasive Methods for Heart-Rate Measurement," *J. Usability Stud.*, vol. 15, no. 3, pp. 145–165, 2020. Available: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Mathe_May2020.pdf.

[17] Y. M. Choi, "Applying Tangible Augmented Reality for Product Usability Assessment." *J. Usability Stud.*, vol. 14, no. 4, pp. 187–200, 2019. Available: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Choi_August2019.pdf.

[18] B. Jerman-Blazic and T. Arh, "Usability Evaluation and Study of a Video-Conferencing Service Provided via the Virtual Conference Centre," 2011. Available: http://www.thinkmind.org/download.php?articleid=icn_2011_20_20_10118.

[19] M. S. Khalid and M. I. Hossan, "Usability evaluation of a video conferencing system in a university's classroom," in *2016 19th International Conference on Computer and Information Technology (ICCIT)*, 2016, pp. 184–190, doi: 10.1109/ICCITECHN.2016.7860192.

[20] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation In Industry*, vol. 189, no. 194, CRC Press, 1996, pp. 207–212, doi: 10.1201/9781498710411-35.

[21] A. Sidik, "Penggunaan System Usability Scale (SUS) Sebagai Evaluasi Website Berita Mobile," *Technol. J. Ilm.*, vol. 9, no.

2, pp. 83–88, Apr. 2018, doi: 10.31602/tji.v9i2.1371.

[22] M. Gao, P. Kortum, and F. Oswald, "Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for Reliability and Validity," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 62, no. 1, pp. 1414–1418, Sep. 2018, doi: 10.1177/1541931218621322.

[23] A. Sasongko, W. E. Jayanti, and D. Risdiansyah, "USE Questionnaire Untuk Mengukur Daya Guna Sistem Informasi e-Tadkzirah," *J. Khatulistiwa Inform.*, vol. 8, no. 2, pp. 80–87, 2020. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/9135.

[24] K. Finstad, "The system usability scale and non-native English speakers," *J. usability Stud.*, vol. 1, no. 4, pp. 185–188, 2006, Available: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Finstad_Aug2006.pdf.

[25] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009. Available: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Bangor_May2009.pdf.

[26] J. R. Lewis and J. Sauro, "The Factor Structure of the System Usability Scale," in *International conference on human centered design*, Springer, 2009, pp. 94–103, doi: 10.1007/978-3-642-02806-9_12.

[27] H. Mastrisiswadi, I. Lestari, A. Istiani, R. Husaini, and R. P. Agusdin, *Panduan Spada Wimaya UPN "Veteran" Yogyakarta*. UPN''Veteran''Yogyakarta, 2020. Available: https://spada.upnyk.ac.id/.

[28] M. Fan, J. Lin, C. Chung, and K. N. Truong, "Concurrent Think-Aloud Verbalizations and Usability Problems," *ACM Trans. Comput. Interact.*, vol. 26, no. 5, pp. 1–35, Sep. 2019, doi: 10.1145/3325281.

[29] P. Chynał and J. M. Szymański, "Remote usability testing using eyetracking," in *IFIP Conference on Human-Computer Interaction*, 2011, pp. 356–361, doi: 10.1007/978-3-642-23774-4_29.