Usability Testing based on System Usability Scale and Net Promoter Score

1st Ginanjar Wiro Sasmito
Informatics Engineering
Politeknik Harapan Bersama
Tegal, Indonesia
https://orcid.org/0000-0002-3941-1769

2nd La Ode Mohamad Zulfiqar
Informatics Engineering
Politeknik Harapan Bersama
Tegal, Indonesia
https://orcid.org/0000-0003-4147-6403

3rd M. Nishom
Informatics Engineering
Politeknik Harapan Bersama
Tegal, Indonesia
https://orcid.org/0000-0003-0765-1044

***Abstract*— In a variety of software development methods and their models of the activity process, no one does not have testing activities in their application. System testing is the last phase of testing that done by the developer in the internal work-space. However, phase of testing is not the end of the whole testing phase, since there is another stage to pass before the product deployment process, it is the acceptance testing. A popular method to perform the user acceptance testing is the usability testing, which in this method, the result provides the indication of the essential quality for the client to decide whether to approve or to deny the product. In this research of paper, we are going to do the usability testing by utilizing the System Usability Scale technique and the Net Promoter Score technique to measure appropriateness of the population information systems in the city of Tegal while we are also finding optimum testers required using the sampling technique. The result we earn trough this research is according to the Usability Testing result the proposed system developed by researchers is better than the current running system and the total testers participating on system testing is able to optimize using the sampling method.**

Keywords—Software Testing, User Acceptance Testing, Usability Testing, System Usability Scale, Net Promoter Score

I. INTRODUCTION

In a variety of software development methods and their models of the activity process, no one does not have testing activities in their application. This proves that the software testing process is crucial in software advancement, both in the system scale and in the application scale [1].

Through the methodology of the software development lifecycle (SDLC) there are basic steps of testing, that are unit testing, integration testing, and system testing [2]. By the unit testing, the smallest part of code which is function, module or class from the systems are tested to ensure it will be able to run properly. After that, there is integration testing to make sure each of the smallest components are collaborated correctly. Then the last phase of testing is the systems testing to observe the overall systems working process [3].

Commonly white box testing is utilized for the investigation of the logical and structural internal code. Hence, the white box testing can be one of the technics to do the unit testing and the integration testing. While the black box testing technic is usually used to examine the systems testing since it is a technic of testing which analyze the fundamental aspects of the systems with a very less relevance from the structure of the logical inside the systems [4].

System testing is the last phase of testing that done by the developer in the internal workspace. However, phase of testing is not the end of the whole testing phase, since there is another stage to pass before the product deployment process, it is the acceptance testing [5]. Over the acceptance testing, the product of the software is verified by the end users to confirm the acceptability. There are several types of testing to work on this standing testing such as- Alpha or beta testing, user acceptance testing or business acceptance testing. A popular method to perform the user acceptance testing is the usability testing, which in this method, the result provides the indication of the essential quality for the client to decide whether to approve or to deny the product [1].

System usability scale (SUS) is one of the technics to evaluate the usability testing. With the most extensively adoption regulated questionnaire for the appraisal of perceived usability, this mechanism become so popular nowadays [6].

Another essential part of the software testing activity is the tester. The maximum and minimal numbers of the testers must also be considered when testing activity is worked. A sampling technic as one of the statistical methodologies for take a part of the whole population is able to be used on specify the optimum number of testers [7].

In this research of paper, we are going to do the usability testing by utilizing the SUS technic and the Net Promoter Score (NPS) technic to measure appropriateness of the population information systems in the city of Tegal. To examine its appropriateness, we are comparing the proposed system with current system used which is BPS system. Hence, we can achieve the percentage rate of the finest systems for user utilization. Furthermore, to find the novelty of this research, we are also utilizing the sampling method to determine optimum numbers of system's testers.

Then to present this paper, we divide the section into 5 sections. First section is introduction, then the next section we are going to discuss some related works in software testing using SUS, NPS, and sampling methods that we are adapting to do this research. After that in the third section, we are discussing how the system to be examined on this research works. The fourth section is result and discussion section which shows how we are finding the NPS, SUS, and optimum number of testers results. The last section is conclusion, consist the achievement that we gain through this research.

978-1-7281-4520-4/19/$31.00 ©2019 IEEE

II. Related Works

A. System Usability Scale Method

System Usability Scale (SUS) is the questionnaire to measure the utilizing perception. It is founded by John Brooke in the 1986 and formerly proposed for testing the office electronical systems [8].

In the method that applied on this research, we briefly depict whole process of the SUS activity as shown in figure 2. Which is on that picture shown there are 3 main stages, that are: questionnaire deployment, score calculation, and classified the rating based on grade.

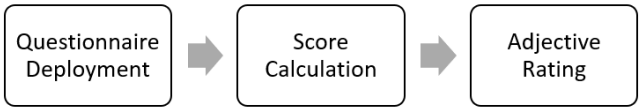

Fig. 1. SUS flow activity

- *Questionnaire Deployment*

In this phase, the questionnaire will be delivered to the participants. The questionnaire includes 10 questions adopted from the original SUS [9] (see Table I).

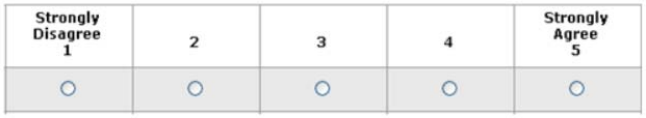

Fig. 2. SUS score scale

TABLE I. Form of Original SUS

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

Those 10 SUS is having the value scale of each point of answers from 1 to 5 (see figure 3) [10].

- *Score Calculation*

By the stage of calculation, the answers from participant will be scored through the phase below:

- Divide odd number questions and even number questions.
- For each odd number questions, reducing the value (x) by 1 (see Equation (1))

$$x - 1 \quad (1)$$

- For each even number questions, do the Eq. (2)

$$5 - x \quad (2)$$

- Combines the result value of odd questions and the even questions
- Multiplies the combination result by 2.5

The highest score of the SUS result is 100, thus the score more than it, is categorized as an error value, or not valid result.

- *Adjective Rating*

Through the adjective rating, the gain value from score calculation phase will be process into classification of the system's usability.

The table (Table II) below shows the interpretation of the SUS score.

TABLE II. SUS Adjective Rating

SUS Score	Grade	Adjective Rating
>80.3	A	Excellent
68-80.2	B	Good
67	C	Okay
51-66	D	Poor
<51	F	Awful

As it shown on the Table II, the SUS score is greater than 80.2 have the best adjective rating which is classified as an excellent product. For the score between 68 to 80.2 will classified as good product. The lowest standard of usable product is in SUS score 67. Under the rate of 67 which is 51 till 66 is categorized for disability product. The worst value of the score is below 51[11].

B. Net Promoter Score Method

Net promoted score (NPS) is the effective method to measuring and controlling the customers' satisfied levels. NPS is not only useful for meter the customers' loyalties but also for the tools monitoring their loyalties. NPS is one of the simple methods to categorized customers according to their answers through a single question. Commonly the question sound like this: *"how is the possibility of you will be recommending our product to your friends?"* [12].

Since our study case is taking place on the city of Tegal, so we are using the population on this research from its city. We are applying random sampling method as many as 12 samples of respondents.

Adopting this method, the goal is to calculating the obedience level of the customer on each product that we tested. Customer will be classified into 3 categories, that are: Promoter, Passive, and Detractor. Passive respondent will not be calculated on the formulation (eq. (3)).

$$Promoter\ \% - Detractor\ \% = NPS \quad (3)$$

The higher the promoter value, the lower value of detractor, then the loyalty value will be better. After NPS is gained, then it will be divided into 4 rate categories, which are: NPS star, NPS leader, NPS excellent and NPS good [13].

Through this research, 4 rate categories mentioned above are modified. We are divided 100% by 4, so NPS star is

acquired when NPS score is greater than 74%. For the NPS value under 75% until 50% will be categorized as NPS leader. NPS excellent is 50% reduced by 25%, and so on until the NPS good achieved its value.

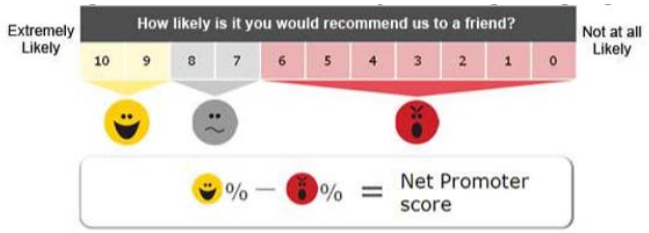

Fig. 3. Net Promoter Score

The NPS is providing the data quickly and it easy to understand the customers' feelings about the products that their use. Hence, the provider of the product can react faster to the negativity feedback from the customer [14].

C. *Sampling Method*

Selecting sample is an alternative for the researchers to collect data from the whole study cases. Hence, by that the researchers are able to minimize the number of cases on the optimum value.

The stages of sampling method are described on the following figure (figure 4).

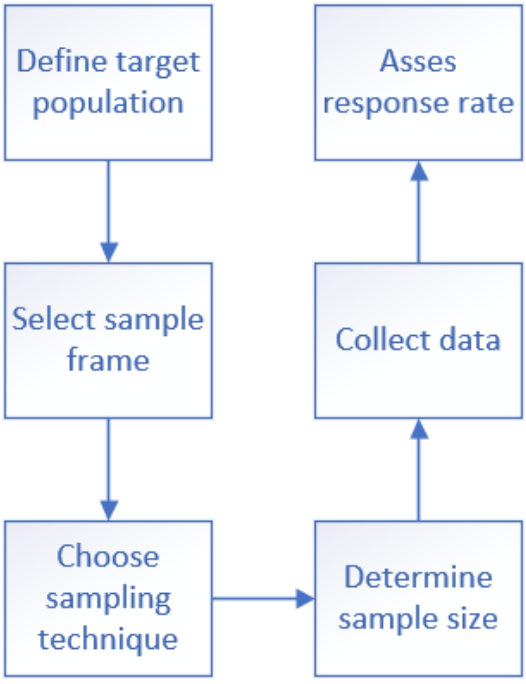

Fig. 4. Steps of sampling process

- Phase 1: Define Target Population

  On this first phase of sampling process, there is need to make sure the target of the population research field.

- Phase 2: Select Sample Frame

  The real cases list of sampling frame that represent the population is determined through this phase.

- Phase 3: Choose sampling technique

  In common, sampling techniques is derived into probability sampling and non-probability sampling. Probability sampling is done when each item on the population have the same opportunity of being inserted in sample. While the non-probability sample is used on the study case research design and qualitative research.

  Probability sampling have some techniques such as: simple random sampling, systematic sampling, stratified random sampling, cluster sampling, and multi-stage sampling.

  From all those probability sampling techniques, the stratified random sampling is the most popular techniques used. In this technique, the population is divided into strata (or subgroups_ and a from each subgroups the random sampling would be taken.

- Phase 4: Determine sample size

  For calculating the sample size, the general formula use is:

$$n = p\,(100)pz^2/E^2 \qquad (4)$$

  n is required sample size

  P is the percentage occurrence of a condition

  E is the percentage maximum error required

  Z is the value corresponding to level of confidence

- Phase 5: Collect data

  Collecting data is the key process of the activity

- Phase 6: Asses response rate

  Asses response rate is the last phase which responding the acceptable number of cases on research.

III. THE SYSTEMS IMPLEMENTATION

The population information of the Tegal City yet able to be served well, since the data obtaining technic still worked by the way of conventionally. Its process started when the census officers came to the citizens house in door to door manner to get the data samples. Then those gathered data will be collected before it given to the authorize party, Central of Statistical Affairs (BPS). The BPS represents the Indonesian government to manage all of the statistical information include census of the population data.

However, there is another party that worked to collect the population data census, which is The Office of Demographic Affairs. This party is collecting the data through the citizens that directly come to their office to be registered as local citizens. Actually, The office of demographic affairs have the responsibility to manage the government task in each of their autonomous region, and they have function to control the citizenship administration issues.

The difference technic worked between BPS and the office of demographic affairs in the process of data collection resulting some dissimilarities information. Comparing among BPS and the office of demographic affairs from which one presenting the more valid information is puzzling since those two bureaus have same signature for this concern.

An account of its problem, the population information system that that based on GIS was implemented. The advantage goals of developed this system was to ease the citizens from accessing the information about population's condition. Even more, this system is able to visualizing through maps, thus it will be more interesting for the end users. The guarantee of the data validation is the main issue consideration. The collecting data process start from the most bottom level of the government representation (shown in figure 1), which is neighborhood zone. In the neighborhood zone, it is inputting the data considering the variables from government. Inputted data will be validated by the next level authority, it is hamlet level. The hamlet will continue the valid data to the urban village, then continuing the data to the sub-district, and the last to the to the government. Verified data accepted by government will be automatically show on the official site of government itself.

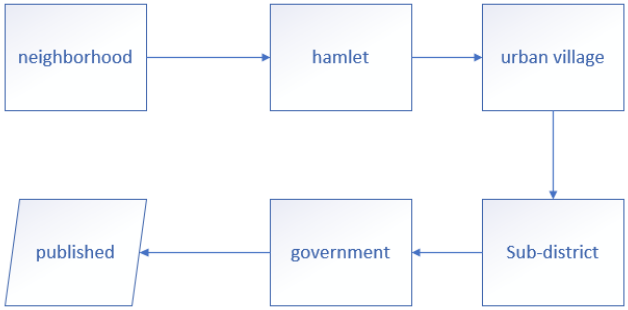

Fig. 5. Data collection flow

IV. Result and Discussion

Through this section we are going to show the calculation result and also the adjective rating of SUS that we gain after analyze the system built by BPS and the proposed system for population information system.

We hand the questionnaire to the 12 random respondents, and then they are instructed to make responds to the SUS questionnaire (fig. 3) and the NPS questionnaire (fig. 4) for both performance of the BPS system and the proposed system.

We worked the testing for each system, which on the following discussion, the proposed system would be discuss first (section B) then the BPS system as current running system is discussed next.

A. *Sampling of Testers*

- Phase 1: Define Target Population

  Our target population is based on the probability of age from user that will be utilized the systems. The population we are used is according to the BPS (statistic center of Indonesia).

- Phase 2: Select Sample Frame

  The sample frame we use is associated with the purpose of the research, which is to find the probability of best number for doing testing on examined systems.

- Phase 3: Sampling Techniques

  We are choosing the technique of stratified sample since the group of the population is already generated.

- Phase 4: Determine sample size

  The general formula on equation 4 is utilized to determine the sample size of this research.

  Percentage occurrence of the value is 50% as we choose random sampling then the probability that we could gain is half of maximum value of percentage. The level confidence condition that we are putting on this research is 2.57 which the value also equal to 0.01. To maximum error required is 5% tolerance. Hence, for the result we gain 12 respondents is presented to do the testing for both proposed system and current running systems (BPS's)

- Phase 5: Collect data.

  Data is collected according to the chosen respondents which are 12 total participants.

- Phase 6: Asses Response Rate

  Since we are going to test the systems, assessing response rate is done using NPS and SUS that will fully described on the next sub-session on this research.

B. *Proposed Systems*

On the proposed system we are earning the average SUS 87.50 (shown in Table IV) which is according to the SUS adjective rating, it categorized as excellent product. That result is appropriate with the NPS score obtained, which the NPS is 75 % (shown in Table III) from 12 respondents (see Table V).

TABLE III. NPS Proposed Systems

Net Promoter Score Calculation		
	Number	Percentage
Promoters	9	75%
Neutrals	3	25%
Detractors	0	0%
Total Responses	12	100%
Net Promoter Score		75

TABLE IV. AVERAGE SUS SCORE OF PROPOSED SYSTEM

Scales			
Odd items	Even items	SUS score (/100)	Grades
20	16	90	A
19	13	80	B
19	18	92.5	A
18	16	85	A
20	16	90	A
18	18	90	A
20	16	90	A
20	20	100	A
16	20	90	A
18	11	72.5	B
20	8	70	B
20	20	100	A
	Average SUS Score	**87.50**	A

C. *BPS Systems*

BPS systems is presenting data with the conventional way. It is using table to shows the number of the population on each region (see figure 4 and also the data displays is outdated (see Table 5) [15].

Fig. 6. Outdated prove of BPS

TABLE V. DATA PRESENTATION OF BPS

Kecamatan/Kelurahan		Laki-laki	Perempuan	Jumlah
Sub District/ Villages		*Male*	*Female*	*Total*
(1)		(2)	(3)	(4)
I	**TEGAL SELATAN**	**33,508**	**32,398**	**65,906**
1.	Kalinyamat Wetan	2,731	2,626	5,357
2.	Bandung	3,132	3,033	6,165
3.	Debong Kidul	2,858	2,683	5,541
4.	Tunon	3,337	3,193	6,530
5.	Keturen	2,548	2,380	4,928
6.	Debong Kulon	2,568	2,355	4,923
7.	Debong Tengah	6,800	6,574	13,374
8.	Randugunting	9,534	9,554	19,088

This technic of display is affecting the judgment of users that using the system. It is obviously seen in the result of SUS and the NPS.

It is shown that the product of BPS has SUS score 60 from total point 24 multiplied by 2.5. Thus, it makes the SUS adjective rating of BPS is poor product. The NPS score of the BPS also low. From the result that we acquire for NPS, BPS only got 10 percent.

V. CONCLUSION

We are comparing the systems that developed by between BPS and the researcher applying the method of Usability Testing. We found that the system developed by researches as proposed system for information system for population has a better result score according to SUS technique and NPS technique. The proposed system gains the average SUS score for 87.50% and the NPS shows it as an excellent product for the users. In another hand, the SUS score for the BPS system is 60% and it makes the adjective value of the product is poor according to the NPS calculation.

Another aim that we achieve is the appropriateness of using the sampling technique for determine the optimum number of the testers or participants to examine the systems. We found that with 99% percent of confidence level and 5 margin error, we can reduce the total participant that might be involved if we are not using the sampling methods for selecting the participants.

REFERENCES

[1] A. Uddin and A. Anand, "Importance of Software Testing in the Process of Software Development," *IJSRD-International J. Sci. Res. Dev.*, vol. 6, no. February, pp. 2321–0613, 2019.

[2] M. A. Jamil, M. Arif, N. S. A. Abubakar, and A. Ahmad, "Software testing techniques: A literature review," *Proc. - 6th Int. Conf. Inf. Commun. Technol. Muslim World, ICT4M 2016*, no. November 2017, pp. 177–182, 2017.

[3] G. D. Everett and R. McLeod, *Software Testing: Testing Across the Entire Software Development Life Cycle*. 2006.

[4] N. Dhingra and Mayank, "Contingent study of Black Box and White Box Testing Techniques," *Int. J. Curr. Eng. Technol.*, vol. 4, no. 5, pp. 3346–3352, 2014.

[5] R. S. Pressman, *Software Engineering A Practitioner's Approach 7th Edition*. 2010.

[6] J. R. Lewis, "The System Usability Scale: Past, Present, and Future," *Int. J. Hum. Comput. Interact.*, vol. 34, no. 7, pp. 577–590, 2018.

[7] H. Herdoost, "Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research," *SSRN Electron. J.*, no. January 2016, 2018.

[8] Z. Sharfina and H. B. Santoso, "An Indonesian adaptation of the System Usability Scale (SUS)," *2016 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2016*, pp. 145–148, 2017.

[9] J. Brooke, "SUS—a quick and dirty usability scale. 1996," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.

[10] H. Alathas, "Bagaimana Mengukur Kebergunaan Produk dengan

System Usability Scale (SUS) Score," *kelasux*, 2018. [Online]. Available: https://medium.com/kelasux/bagaimana-mengukur-kebergunaan-produk-dengan-system-usability-scale-sus-score-2d6843ca780a. [Accessed: 15-Sep-2019].

[11] J. Brooke, "SUS: A Retrospective," *J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, 2013.

[12] F. Reichheld, *The Ultimate Question 2.0 (Revised and Expanded Edition): How Net Promoter Companies Thrive in a Customer-Driven World*. Boston, Massaxhusetts: Bain & Company, 2011.

[13] M. N. Rajasekaran and M. N. Dinesh, "How Net Promoter Score Relates To Organizational Growth," *Int. J. Creat. Res. Thoughts*, vol. 6, no. 2, pp. 2320–2882, 2018.

[14] T. L. Keiningham, L. Aksoy, B. Cooil, T. W. Andreassen, and L. Williams, "A holistic examination of Net Promoter," *J. Database Mark. Cust. Strateg. Manag.*, vol. 15, no. 2, pp. 79–90, 2008.

[15] Badan Pusat Statistik (BPS), "BPS Sosial dan Kependudukan Kota Tegal," *Badan Pusat Statistik Kota Tegal*, 2019. [Online]. Available: https://tegalkota.bps.go.id/statictable/2017/01/05/199/jumlah-penduduk-kota-tegal-menurut-kecamatan-kelurahan-dan-jenis-kelamintahun-2015.html. [Accessed: 15-Sep-2019].