IJCSI International Journal of Computer Science Issues, Volume 14, Issue 2, March 2017
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201702.1723

Usability Evaluation Method based on ISO/IEC SQuaRE : A Systematic Mapping Study

Tenia Wahyuningrum[1,2], Azhari[1]
[1] Computer Science & Electronics Department, Universitas Gadjah Mada
Yogyakarta, 55281, Indonesia

[2] Informatics Department, ST3 Telkom
Purwokerto, 53147, Indonesia

Abstract

Usability Evaluation Methods are essential in measuring the success of the software. Several methods have been able to measure some sub-characteristics that have been required in the ISO SQuaRE. However, there are some other sub-characteristics included in the measurement. Several methods such as heuristic evaluation, usability testing, and questionnaire are methods often used. There are 16 paper of 52 papers were selected for the assessment and analysis of the utilization of the ISO sub-characteristics. Most papers show that learnability is a sub-characteristic that widely used in the assessment, other papers proposed additional sub-components such as attractiveness, ease of use, and security.

Keywords— *Systematic Mapping Study; Usability Evaluation Method; ISO/IEC SQuaRE*

1. Introduction

ISO/IEC 25010 is a new model of Software Quality measurement derived from ISO/IEC 250n series (known as Software Quality Requirements and Evaluation-SQuaRE)[1]. This International standard is to substitute ISO/IEC 9126 and presents a detail of quality models for computer systems and software product, quality in use and data. Al Qutaish (2009) investigated the weakness of ISO 9126 modeling and identify four points of weakness by self-discovered and seven points by other researchers-discovered. He found a necessity to give the qualitative values to rank the result, for example, “excellent” label is similar with 100% percentage result, and “very good” label is used to represent 80% task completion[2].

Usability evaluation is one of the parts of Software Quality measurement, and it is similar to the term of Quality in Use in the ISO SQuaRE detailed (excludes freedom of risk and context coverage). Usability evaluation is necessesary before the software product marketed to users. It could be argued that usability is the determinant factor of the success of entire software system[3]. Usability has been defined in several studies in several ways, and it is all adequate system such as efficiency, effectiveness, and satisfaction [4][5][6][7][8] but all of the aspects still have some shortcomings dependent upon the software[9].

There are two categories od Usability Evaluation Methods (UEMs), analytical and empirical. Analytical UEMs also was known as the inspection method is an evaluation technique using the expert opinion, including Heuristic evaluation, Cognitive Walkthrough, or Guidelines[10]. Empirical UEMs meanwhile, is performed based on the user experience such as usability testing (also called as thinking aloud), user performance test, remote usability testing, or coaching method. Also, there is usability evaluation with statements involving the users such as user satisfaction questionnaire, field observation, focus group, or interviews[11].

This study focuses on the usability evaluation method mapping based on ISO/IEC SQuaRE to determine the extent of implementation of the standard measurement to show what parts need to be further investigated in the future. Here, the research used the Systematic Mapping Study Methodology based on Research Question. A method providing a structure of the type of research reports and result that have been published trough categorization them and gives a visual summary of the map as the results[12]. This paper is organized as follows. Section 2 discusses about related work; Section 3 presents the method of Systematic Mapping Study. Section 4 then describes the result of the Systematic Review and Section 5 presents the conclusions and suggests some areas for further investigation.

2. Related Work

Various methods for evaluating usability (Usability Evaluation Methods - UEMs) have been applied to assess the interaction between human and product. It used to identify the aspects of the interaction and expectation towards the usability enhancement. Some methods are rarely used because it is too complication and inefficient. The experts have performed inspection method also known as analytical method. The users have performed testing method can be called empirical method. The combination of several methods usability later is used by some researchers to evaluate the software. Lodhi (2010) used two methods, Heuristic Evaluation (conducting out by the evaluation specialist) and Usability Testing (carried out by the representative user), using 10 Nielsen"s Usability Heuristics as an assessment parameter[13]. Otaiza et al. (2010) proposed a method to choose the right usability evaluation on some transactional web applications, and Heuristic Evaluation is defined as the most efficient technique of comparison seven methods [14]. Freddy Paz (2015) argued that Heuristic Evaluation was conducted to complement the Usability Testing. The second technique is considered necessary to evaluate a software product and website. At this stage of Usability Testing, users will rate the website, and at the stage of Heuristic Evaluation, the experts will find the fault usability. In the end, it is not only to clarify which site is good or bad, but also to add those of experts in improving the performance of a website [15]. Therefore, important to know how to spread the use of empirical methods, analytical methods, and the combination of them.

3. Research Method

In the beginning, some medical researchers applied the guidelines in Systematic Mapping Study was, and now, it is widely used by software engineering researcher to reduce risk of bias and incompleteness in the review result [16]. The process of Systematic Mapping consists of five steps (Fig.1): (1) defining the research questions, (2) searching of papers, (3) screening of documents, (4) key wording of abstract, (5) data extraction and mapping process[12]. In the early steps, researchers define Research Questions (RQ) to find information from the existing research to provide an overview of a study area, and to identify the forums in which research in the area has been published to see a trend. In phase 2, the researchers identified, created and tested on the search strings on scientific databases to find the papers. From these result, it was found many papers related to the search string. Here, we can merge all the papers data, and choose the paper to the relevant criteria. Next step was to read all papers by title, abstract, and full, to find the papers. In the final part, data analysis and data plotting were given to report the results.

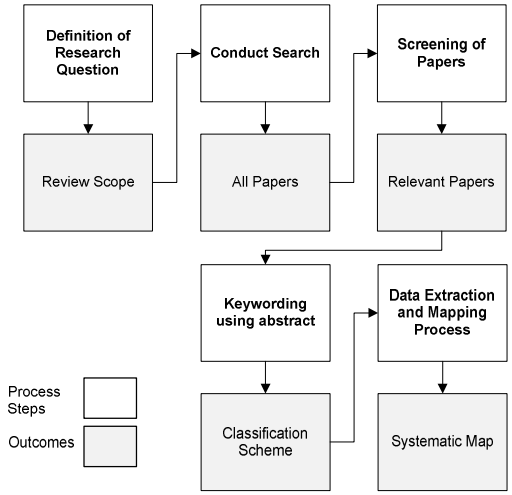

Fig.1: The Systematic Mapping Process[12]

4. Result and Discussion

4.1 Definition of Research Question

This step aimed to find the information of the existing research on the use of ISO SQuaRE model and Usability Evaluation Method to see the research trend. To formulate the Research Question, we followed the PICO (Population, Intervention, Comparison, and Outcome) structure [16]. Table 1 and Table 2 presents the study"s population and intervention and Research Questions.

Table 1 . PICO Structure

PICO	*Description*	*Formula*
Population	The target for the investigation	Software, application, website
Intervention	Specifying the research aspects or issues of interest to the researchers	Usability Evaluation Method, ISO 250n or ISO SQuaRE
Comparison	Aspect of the investigation with which the intervention compared to	N/A
Outcome	The setting of the intervention	Implementation UEMs using ISO SQuaRE trends

Table 2. Research Questions

Research Question		*Description*
RQ 1	How many papers that discuss the implementation of Usability Evaluation Method using ISO SQuaRE by year?	To identify the distribution of papers implementing the Usability Evaluation Method using ISO SQuaRE standard and its development in every year.
RQ 2	How many types of papers that discuss the implementation of Usability Evaluation Method using ISO SQuaRE?	To identify the distribution of paper type that implementing the Usability Evaluation Method using ISO SQuaRE.
RQ 3	What kind of sub characteristics of ISO SQuaRE model is widely used in the measurement of usability?	To determine the distribution and characteristics of the implementation of trend SQuaRE on paper.
RQ 4	What kind of method to evaluate usability using ISO SQuaRE?	To identify the most widely used method.

4.2 Searching Papers

The second step of Systematic Mapping Process was to search the papers from journals and proceedings related to the research question. The PICO structure is a good way to create the search string, and it should be driven by the research questions. It can be taken from each aspect of structure. From PICO structure in Table 1, some of the keywords used to filter paper be evaluated such as: "software"; "software application"; "website"; "usability method evaluation"; "ISO 25000"; "ISO 25010"; "ISO SQuaRE" by using Boolean operator "AND", "OR".

4.3 Screening Papers

The search string finds various kinds of papers in the topics. Here, we can merge all the documents data, and choose the paper to the relevant criteria. There are three commonly used research databases; IEEExplore (ieeexplore.ieee.org), Science Direct (www.sciencedirect.com), and Scopus (www.scopus.com), where we can search for the articles by typing string on a checkbox of title, abstract, and full content text. After retrieving the result, we applied selection criteria to filter the candidates. Figure 2 shows the selection process result, that consists of 7 steps.

- Automatic searching: The result based on search string derived from research question showed 52 papers including journal, proceedings, and lecturer notes.
- Removing repetitions: Several papers could be found in three databases, resulting in duplicate data. Hence, we have eliminated the same data to avoid data redundancy, and 20 papers independent papers were identified.
- Selection by title: The objective of this filter was to remove the documents with the scope that was clearly unrelated to UEMs using ISO SQuaRE. We found that all document titles were relevant with the Research Questions. Therefore, no documents were removed for these steps.
- Selection by abstract: At this stage, we discarded all those works that did not present the implementation of the Usability Evaluation Method. We removed one paper that was not relevant because it examined the usability weaknesses of the model, and did not discuss any implementation on the model of UEMs.
- Selection by the full paper : In the next process, we removed the papers which did not accomplish the following inclusion criteria properly. We removed one paper that was not relevant, because it used the Spanish language. We also removed four paper that not relevant to the inclusion criteria and exclusion criteria that described in Table 3.
- Accepted papers: Finally, we have sixteen of fifty two papers were selected for the evaluation and analysis.

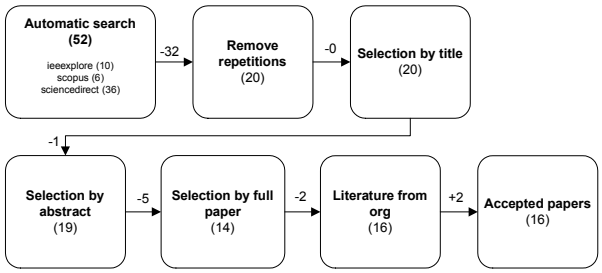

Fig.2: Selection Process Result[16]

Table 3. Inclusion and Exclusion Criteria

Criteria	*Description*
Inclusion criteria	Research focused on usability evaluation methods (title, abstract, keywords) In industrial and academic research on large and small scale The study discusses and compares the performance of the model in

	usability evaluation International publications (journals, proceedings and lecture notes) Publications between the years 2010-2016
Exclusion criteria	Not associated with Research Question The paper is not a journal, proceedings or lecturer notes

4.4 Keywording using abstract

Petersen argued that keywording was way to reduce the time needed in developing the classification scheme and ensuring that the scheme took the existing studies into account. In this stage, we read the abstracts and looked for the keywords and concepts reflecting the contribution of the paper. Reviewer could also read the introduction or conclusion when the abstracts were so poor in quality to allow meaningful keywords to be choosen[12]. We read the abstracts discussing the case study, implementing, evaluation, quantification, measuring the UEMs using ISO SQuaRE.

4.5 Data Extraction and Mapping Process

At this stage, we performed a mapping to answer each of Research Questions.

RQ 1, How many papers that discuss the implementation of Usability Evaluation Method using ISO SQuaRE by year?

Figure 3 presents the papers distributions based on year. From this chart, we were able to see that the implementation of UEMs using ISO SQuaRE increased in 2015. There were 25% papers published in this year and decreased in 2016. This is possible since the research has been done in the current year, so there are still many papers in the process of publishing or not indexed database.

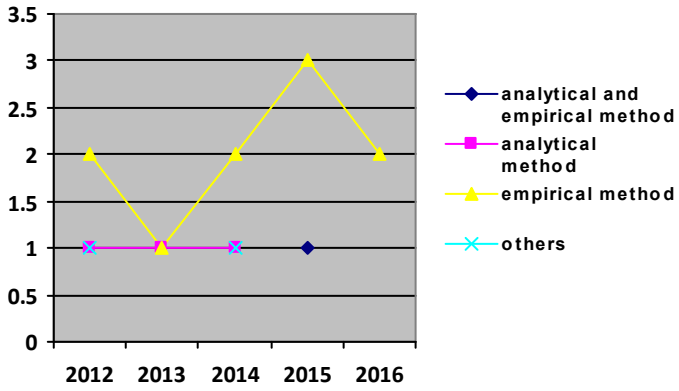

Fig.3. Papers distribution based on year

RQ 2, How many types of papers that discuss the implementation of Usability Evaluation Method using ISO standard SQuaRE?

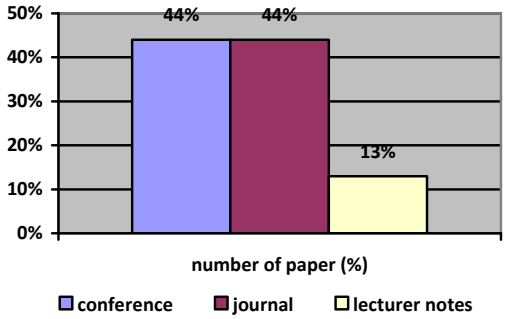

Fig.4. Type of papers

Figure 4 shows the same numbers of papers published in the conference in comparison to numbers of journal (44%). In Computer Science, journal and conference paper is the primary source for research. Traditional conference paper should be published as rapid publication papers instead of in conference proceedings. Papers presented at international conferences are considered as the primary literature if they are peer reviewed and published.

RQ 3 : What kind of sub-characteristics ISO SQuaRE model widely used in the measurement of usability?

The characteristics of usability comprise two aspects, there are 11 sub-characteristics, such as appropriateness recognisability, learnability, operability, user error protection, UI aesthetics, accessibility, effectiveness,

efficiency, satisfaction, freedom from risk, and context coverage. Not all of the characteristics used for the measurement of usability, some researchers reduce and add characteristics that were not appropriate. Some researchers added sub-characteristics of attractiveness to evaluate the video game software, word processor, and website [9], [17]–[20]. Ease of use sub-characteristic has added to evaluate software tools, like GQM (Goal Quality Metric) application and CSRML (Collaborative Systems Requirements Modelling Language) [18], [20], [21]. The type of software which is confidential and dangerous, like a poison control information system, it should be added sub-characteristic security [22]. Torrente using Sirius framework for evaluating usability based on heuristics to perform expert assessments that takes into account different types of websites [23]. Table 4 shows that learnability, appropriateness recognisability, operability, and efficiency are widely used in the evaluation of usability.

Table 4. Sub characteristic used in the usability evaluation

Sub Characteristics	***References***
Appropriateness recognisability	[16][18][20][21][23][24][25][26][27]
Learnability	[9][16][18][19][20][21][24][25] [26][27]
Operability	[16][17][18][20][21][24][25][26][28]
User error protection	[16][23][24][25][26][27]
UI aesthetics	[16][23][24][25] [26][27]
Accessibility	[16][17][18][25] [26][27]
Effectiveness	[19] [22][24][26][29][30]
Efficiency	[9][16] [19][22][24] [26] [29][30]
Satisfaction	[9][20][22][29][30] [31]
Other characteristics	[9][17][18][19][20][21][22][23] [24][26]

RQ 4 : What kind of method to evaluate usability using ISO standard SQuaRE?

Table 5 shows that the empirical method has been widely used for evaluating the usability. This method include observational study [17], usability testing [28][20][21] and Questionnaire [22][24][9][31][19][30]. The number of users that involved in the questionnare study is varied. Alves using sixty-nine respondents who have been using LACEN system. There are fifty-one technical staff, nine health care providers, seven managers, one information manager and one laboratory director. In the other study, Alves used forty-nine DATATOX users to participate in research, 65% respondents were attendants and 35% were physicians. From both of these studies, generate statistical data showing that LACEN system offers a good quality regarding effectiveness, efficiency, satisfaction, freedom form risk, functional suitability and security[24]. In the other hand, the performed analysis shows that the DATATOX system is have a good quality degree [22]. Dubey used a group of ten users to fill a fuzzy criteria in the questionnaire to evaluating MS Word 2003 usability. Using Fuzzy Multi Criteria Approach, this model considered as the base model essentially covers and integrates maximum number of factors and attributes [9]. Sivaji used 241 primary users, four secondary users and twenty indirect Malaysian URANUS users. The results show that URANUS was successful in improving the usability[30]. Although not described in detail the number of respondents involved in the research questionnaire, however, we can use the slovin formula.

The advantages of this method are feedback given directly by the user, easy collected, good in information detail and simple. The analytical method commonly used is heuristic evaluation [23]. It is uses the tested rules and can find many issues. Biscoglio together with software quality expert evaluators to conduct experiments in determining preservation needs and requirements to set up quality evaluation process in audio video preservation context [25]. The expert reviewers and game designer also conducts an inspection method for building three stages in evaluating the usability of video games, namely (1) the establishment of evaluation requirements, (2) initial usability evaluation, (3) usability evaluation in use [18].

Some researchers are combining two Usability Evaluation Methods. Nwasra [26] proposed a framework that shows the procedural flow between stakeholders (decision makers, evaluators, developers and end users). The questionnaire, interviews, automated tools, and task experiments method are used to measure functional quality, content quality, and appearance quality. Some paper does not mention in detail the methods of evaluation used. But they propose a quality model utilized for the measurement. Dominguez develop a set of Quality Characteristics and Sub-Characteristics for Model-Driven Web Engineering approaches based on ISO/IEC standards [17]. Oriol evaluated the current state of the art of the proposed quality models for web services. The study shows that reliability, security and performance efficiency are the most characteristics explicitly defined in at least half of the surveyed proposals [16].

Table 5. Usability Evaluation Methods

UEMs	*Evaluation Detail*	*References*
Analytical and empirical method	The questionnaire, interviews, automated tools, task experiments	[26]
Analytical method	Expert evaluation	[25]
	Inspection method	[18]
	Heuristic evaluation	[23]
Empirical method	Questionaire	[22][24][9] [31][19][30]
	Usability testing	[28][20][21]
Others	-	[16][17]

5. Conclusions

This paper systematically mapped the implementation of UEMs using ISO SQuaRE. Based on the results, it has been found that the use of the ISO SQuaRE model to measure the usability of software fluctuates every year. There was a decrease in the number of publications in 2016 allegedly because the research was conducted still in the year, hence, some publications were not indexed search engines and still in the publication process. The number of conference paper was the same number of journals, due to the rapid publication needs in the field of computer science. The researchers prefer to publish in advance at international seminars, symposia, and conferences. Learnability is the most sub characteristic from ISO SQuaRE model used for measurement usability, and some researchers have added a new characteristic to support Usability Evaluation Methods. There are more numbers of empirical methods compared to the analytical method. Several methods such as heuristic evaluation, usability testing, and questionnaire are methods often used. In the next study we will attempt to propose a merger of two methods to determine whether there is a significant difference to the usability of software.

References

[1] F. Febrero, C. Calero, and M. Á. Moraga, "Software reliability modeling based on ISO/IEC SQuaRE", Inf. Softw. Technol., vol. 70, pp. 18–29, 2016.

[2] R. E. Al-Qutaish, "An Investigation of the Weaknesses of the ISO 9126 International Standard", Computer and Electrical Engineering, 2009. ICCEE "09. Second International Conference on, vol. 1. pp. 275–279, 2009.

[3] H. Roder, "Specifying usability features with patterns and templates", 2012 1st Int. Work. Usability Access. Focus. Requir. Eng. UsARE 2012 - Proc., pp. 6–11, 2012.

[4] J. Nielsen, Usability Engineering. London: Academic Press, 1993.

[5] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, Human-Computer Interaction, vol. Third, no. January. 2004.

[6] H. Sharp, Y. Rogers, and J. Preece, Interaction design: beyond human-computer interaction, vol. 11. 2002.

[7] International Organization for Standardization/International Electrotechnical Commission, "ISO/IEC 9126-1:2001, Software Engineering, product quality, Part 1 : Quality Model", Geneva, Switzerland, 2001.

[8] International Organization for Standardization, "ISO 9241-11:1998, Ergonomic requirements for office work with visual display terminals (VDTs), Part 11 : Guidance on Usability", Geneva, Switzerland, 1998.

[9] S. K. Dubey, A. Gulati, and P. A. Rana, "Usability Evaluation of Software Systems using Fuzzy Multi-Criteria Approach", Int. J. Comput. Sci. Issues, vol. 9, no. 3, pp. 404–409, 2012.

[10] W. D. Gray and M. C. Salzman, "Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods", Human–Computer Interact., vol. 13, no. 3, pp. 203–261, 1998.

[11] N. E. Jacobsen, "Usability Evaluation Methods The Reliability and Usage of Cognitive Walkthrough and Usability Test", 1999.

[12] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering", 12Th Int. Conf. Eval. Assess. Softw. Eng., vol. 17, p. 10, 2008.

[13] A. Lodhi, "Usability heuristics as an assessment parameter: For performing usability testing", ICSTE 2010 - 2010 2nd Int. Conf. Softw. Technol. Eng. Proc., vol. 2, pp. 256–259, 2010.

[14] R. Otaiza, C. Rusu, and S. Roncagliolo, "Evaluating the usability of transactional web sites", 3rd Int. Conf. Adv. Comput. Interact. ACHI 2010, pp. 32–37, 2010.

[15] F. Paz, F. A. Paz, D. Villanueva, and J. A. Pow-Sang, "Heuristic Evaluation as a Complement to Usability Testing: A Case Study in Web Domain", 2015 12th Int. Conf. Inf. Technol. - New Gener., pp. 546–551, 2015.

[16] M. Oriol, J. Marco, and X. Franch, "Quality models for web services: A systematic mapping", Inf. Softw. Technol., vol. 56, no. 10, pp. 1167–1182, Oct. 2014.

[17] F. J. Domínguez-Mayo, M. J. Escalona, M. Mejías, M. Ross, and G. Staples, "Quality evaluation for Model-Driven Web Engineering methodologies", Inf. Softw. Technol., vol. 54, no. 11, pp. 1265–1282, Nov. 2012.

[18] A. Fernandez, E. Insfran, S. Abrahão, J. Á. Carsí, and E. Montero, "Integrating usability evaluation into model-driven video game development", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7623 LNCS. ISSI Research Group, Department of Information Systems and Computation, Universitat Politècnica de València, Camí de Vera s/n, 46022, Valencia, Spain, pp. 307–314, 2012.

[19] S. Kolahdouz-Rahimi, K. Lano, S. Pillay, J. Troya, and P. Van Gorp, "Evaluation of model transformation approaches for model refactoring", Sci. Comput. Program., vol. 85, no. PART A, pp. 5–40, 2014.

[20] M. F. Terenciani, G. B. Landre, D. M. B. Paiva, and M. I. Cagnin, "A plug-in for Eclipse towards supporting business process lines documentation", 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA). pp. 1–8, 2015.

[21] M. A. Teruel, E. Navarro, V. López-Jaquero, F. Montero, and P. González, "A CSCW Requirements Engineering CASE Tool: Development and usability evaluation", Inf. Softw. Technol., vol. 56, no. 8, pp. 922–949, Aug. 2014.

[22] J. M. Alves, D. B. L. Albino, M. C. Resener, M. Zannin, A. Savaris, C. G. Von Wangenheim, and A. Von Wangenheim, "Quality evaluation of poison control information systems: A case study of the DATATOX system", in Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2016, vol. 2016–Augus, pp. 30–35.

[23] M. C. S. Torrente, A. B. M. Prieto, D. A. Gutiérrez, and M. E. A. De Sagastegui, "Sirius: A heuristic-based framework for measuring web usability adapted to the type of website", J. Syst. Softw., vol. 86, no. 3, pp. 649–663, 2013.

[24] J. M. Alves, A. Savaris, C. G. v. Wangenheim, and A. v. Wangenheim, "Software Quality Evaluation of the Laboratory Information System Used in the Santa Catarina State Integrated Telemedicine and Telehealth System", 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS). pp. 76–81, 2016.

[25] I. Biscoglio and E. Marchetti, "A case of adoption of 25000 standards family establishing evaluation requirements in the audio-visual preservation context", Software Engineering and Applications (ICSOFT-EA), 2014 9th International Conference on. pp. 222–233, 2014.

[26] N. Nwasra, N. Basir, and M. F. Marhusin, "A framework for evaluating QinU based on ISO/IEC 25010 and 25012 standards", 2015 9th Malaysian Software Engineering Conference (MySEC). pp. 70–75, 2015.

[27] J. R. Oviedo, M. Rodríguez, and M. Piattini, "Certification of IPavement applications for smart cities a case study", Evaluation of Novel Approaches to Software Engineering (ENASE), 2015 International Conference on. pp. 244–249, 2015.

[28] J. R. Oviedo, M. Rodriguez, and M. Piattini, "Certification of IPavement Applications for Smart Cities A Case Study", pp. 1–6, 2014.

[29] N. B. N. Rozali and M. Y. B. Said, "Usability testing on government agencies web portal: A study on Ministry of Education Malaysia (MOE) web portal", 2015 9th Malaysian Softw. Eng. Conf. MySEC 2015, pp. 37–42, 2016.

[30] A. Sivaji and S. Soo, "Understanding, Enhancing and Automating HCI Work Practices: Malaysian Case Studies", Procedia - Soc. Behav. Sci., vol. 97, pp. 656–665, Nov. 2013.

[31] H.-J. Jung and S.-J. Hong, "The quality control of software reliability based on functionality, reliability and usability", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7709 LNCS. pp. 112–118, 2012.

Tenia Wahyuningrum, S.Kom., M.T. Currently is pursuing her doctoral program in Computer Science and Electronics Departments, Faculty of Mathematics & Natural Science, Universitas Gadjah Mada Yogyakarta. She took her undergraduate (S.Kom) at STMIK Widya Utama Purwokerto in 2005, and Master (M.T) in Electrical Engineering Department at Institut Teknologi Bandung in 2010. Her research area of interest are Human Computer Interaction, Software Quality, Soft Computing, and Software Engineering.

Dr. Azhari, M.T. He had his undergraduated (S.Si) in Statistics from Universitas Gadjah Mada Yogyakarta. He was granted a degree of Master level by Institut Teknologi Bandung. He presented his dissertation and was declared to pass the exams Ph.D in Computer Science from Universitas Gadjah Mada Yogyakarta. His research area of interest includes Software Engineering, Project Management, Intelligent Agent.