International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6194
***Article Received**: 22 December 2022 **Revised**: 19 January 2023 **Accepted**: 18 February 2023*

A Quantitative Evaluation for Usability under Software Quality Models

Hoijin Yoon
Dept. of Computer Engineering
Hyupsung University
Hwaseong-si, Republic of Korea
hj.yoon@uhs.ac.kr

Abstract—Usability evaluation in Human-Computer Interaction (HCI) is done by observing the user's behaviors and reactions while performing a given task. The observers examine users' behaviors while doing assigned tasks and describe their observations in terms of usability. The usability evaluation would depend on the observers' ability or experience. It proceeds qualitatively. We propose a quantitative evaluation, which adopts attributes and metrics from System and Software Quality Requirement and Evaluation(SQuaRE) published by International Standard Organization (ISO). Furthermore, we examine qualitative observations conducted in usability testing and apply our method to make it a quantitative evaluation.

Keywords- usability; SQuaRE; IEEE/ISO/IEC25000; usability attributes; usability metrics;

I. INTRODUCTION

Software usability is becoming increasingly important in an era where the market is becoming more competitive. In the past, developers have been very focused on software functionality. So testing, which we often refer to, mainly tests the software functionality. If we look at the V-model, introduced as one of the development processes that consider software testing, we can see that even there, there is much focus on the function of the software.

The V model [1] expresses the relationship between software development stages and test levels and matches the development stages that affect the design and execution of each test level. Among the testing stages, acceptance testing is conducted with users' participation, so we can think of it as usability testing. The tests presented in the V model are all stages of testing functional requirements rather than non-functional requirements, which are about software quality but functionality such as Performance, Reliability, Usability, and so on. Therefore, usability is outside the traditional testing range.

As seen in the V model, most software developers do not care much about software qualities other than functionality. They will likely complete the software's functions inside and then develop the user interface outside. It is called inside-out development. On the other hand, in the field of HCI, which regards usability as an essential quality, it is proposed to design and implement a concept model similar to the user's mental model as an outside-in development [2].

In HCI, usability is evaluated by observing the user's product use process. A typical example is Krug's usability evaluation method [3], which measures usability by watching how users perform given tasks. At this time, observers look at the user's screen in third place, listen to the user's thoughts, and write an evaluation result sheet through observation. Considering software V&V [4], it is necessary to examine whether there are evaluation criteria in observation. Suppose the usability evaluation is done without a measurement metric. In that case, it is not easy to trust the result because different evaluation results may come out depending on the observer's capacity or concentration.

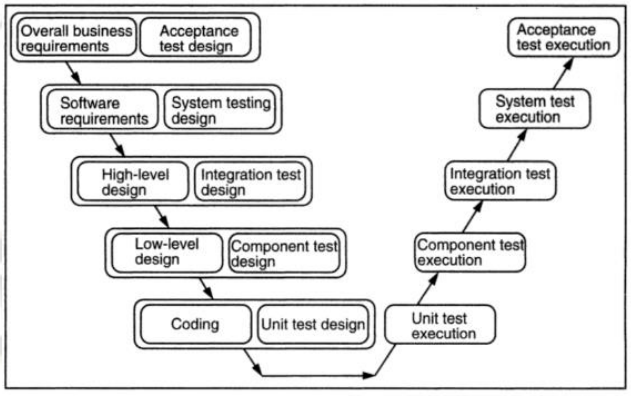

Figure 1. V Model [1]

This paper adds some measurement criteria to evaluate usability to Krug's usability evaluation process. Our method is developed according to the standards drafted by ISO. Furthermore, with several evaluation results obtained by the Krug method, we analyze the areas where they are lacking according to the evaluation criteria we established and show the contribution of the proposed method.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6194
***Article Received**: 22 December 2022 **Revised**: 19 January 2023 **Accepted**: 18 February 2023*

II. USABILITY EVALUATION

A. *Usability in HCI*

The HCI defines usability as a significant quality prior to functionality, which software developers mainly mind. To increase usability, experts in HCI design user interfaces, which would be called user experience (UX), in more depth. In a design supporting good UX, menus and screens are well laid out so humans can intuitively follow them.

The user can immediately know what action to take without any signifier. This characteristic is called affordance [5]. In other words, a design with affordance contributes to increasing usability. A product with good usability enables the user to achieve the purpose naturally and quickly without difficulty when using the product to reach his intended purpose. As a usability principle to express this, "Don't make me think" is suggested [3]. In other words, the product should express a conceptual model similar to the user's mental model [2].

The expert emphasizes that if a function is not findable or usable, that functionality could be eliminated without problems. So Nielsen Norman (NN) group suggests “puts at the center the user as a key element in achieving usability goals” as a policy for usability [6]. Also, the NN group defines usability as follows. Usability refers to a quality attribute that measures how easy it is for a user to use a user interface.

B. *Usability as Software Quality*

At a time when several software quality models were proposed, the need for a standard to eliminate confusion among them emerged. Therefore, ISO, an international standard organization, established ISO 9126 and published it as part of the ISO 9000 standard. ISO/IEC 9126 was announced in the name of "Software Engineering - Product Quality" in 1991. In ISO IEC 9126 Model, the software product quality attributes were classified into a hierarchy of characteristics and sub-characteristics; the highest level defines the quality characteristics, and the lowest level defines the software quality criteria. This model identified six characteristics: Functionality, Reliability, Usability, Efficiency, Maintainability, and Portability, all of which are further divided into 21 sub-characteristics.

In ISO IEC 9126, usability is "A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users" [7]. The sub-characteristics of Usability are Understandability, Learnability, Operability, Attractiveness, and Usability compliance. Each sub-characteristic will be further divided into attributes. An attribute is an entity that can be proved or estimated in the software product. However, the standard does not define attributes because they would vary between software products.

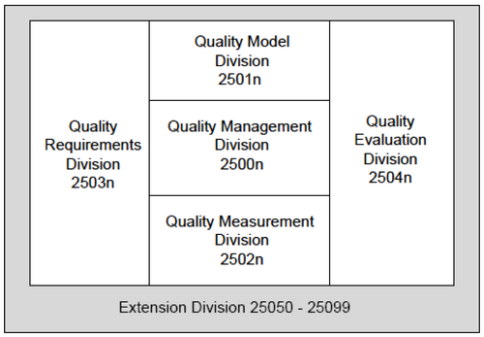

Figure 1. Organization of SQuaRE series of International Standards [8]

The ISO/IEC 9126 standard was first published in 1991 and revised in 2001. In 2011, 10 years later, the ISO/IEC 9126 standard was replaced by the IEEE/ISO/IEC 25000 titled in Systems and software engineering - *Systems and software Quality Requirements and Evaluation* (SQuaRE) - System and software quality [8]. SQuaRE is composed of 5 divisions, each shown in figure 1.

ISO/IEC 2500n, as the Quality Management Division, defines the terms of a standard model and provides guidelines for planning and management to support product requirements specification and assessment management. ISO/IEC2501n describes a detailed quality model for quality in use and data of systems and software products as a Quality Model Division. ISO/IEC2502n, as the Quality Measurement Division, provides mathematical definitions and applicable practical guidelines for measuring the quality of systems and software products. It defines *internal measures, external measures, and quality-in-use measures* for software quality. ISO/IEC2503n, as Quality Requirement Division, provides content that specifies quality requirements. ISO/IEC2504n, as a Quality Evaluation Division, describes requirements, recommendations, and guidelines for quality evaluation from the developer, acquirer, and evaluator perspectives.

C. *DIY Usability Testing*

Steve Krug presents a time-cost-effective approach to usability testing in his book [9] and introduces concrete testing guidelines. It is called *DIY usability testing* in his book. It selects a few participants from the product user group. It simultaneously decides on a representative task that can be a use case for the product and writes it down so that users can easily understand it. In addition to participating users, observers who observe their behavior participate in the test. Observers gather according to a space separate from the space where users use the product and observe the way users use the product through a camera. If the product is a computer program, the movement of the user's mouse must be observed, so the user's screen must be prepared for observers to see through a sharing program such as WebEx.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6194
***Article Received**: 22 December 2022 **Revised**: 19 January 2023 **Accepted**: 18 February 2023*

Figure 2. DIY Usability Testing [9]

The skill of the facilitator is essential in DIY usability testing. The facilitator plays the role of leading all steps of testing and also plays the role of eliciting feedback from users so that observers can acquire good points by asking meaningful questions to users while staying in the same space as them. At this time, a facilitator helps users participate in usability testing without difficulty. The left side of the figure below shows the space of users and facilitators, and the right side shows the space of observers.

In each test session, observers are supposed to write down the usability problems they recognized while watching the user's behavior and listening to his feedback or comments. They can take as many notes as they want, but they must identify the three most severe problems from the findings because those most serious problems are what to get fixed first.

Krug requested the observer to list and write the observation results and highlight the three severe problems to be dealt with as the top priority. The selected problems were written according to the judgment of the observer. In other words, it is a qualitative evaluation in which the evaluation result can come out very differently depending on the observer's capability or experience.

III. USABILITY METRIC

A. Common Industry Format for Usability

ISO/IEC 25066[10] is a Common Industry Format for Usability-Evaluation reports. According to ISO/ICE 25066, various possible types of usability evaluation are specified, and the following three kinds of evaluation are suggested for each of the various environments. The first approach to evaluation is an inspection. Moreover, the second is an observation. It is the observation used in the DIY usability testing mentioned above. ISO/IEC 25066 also introduces that there are two ways the observation, and those are qualitative observation and quantitative observation. Qualitative observation is to observe users' behavior to find actual usability problems, and quantitative observation measures users' performance and response with data regarding effectiveness and efficiency.

In addition to this observation method, the last third is obtaining the user's subjective opinion or information. This method can also be divided into qualitative and quantitative. At the same time, there are three evaluation environments: physical environment and facilities, technical environment, and evaluation administration tool. ISO/IEC 25066 also indicates how much evaluation is needed in each combination of the three evaluation types already explained above and the three different evaluation environments. The indicator is required, recommended, or optional. The table below shows this classification.

TABLE I. TYPE OF EVALUATION FOR EVALUATION ENVIRONMENT

	Type of Evaluation			
	Inspection	***Qualitative observation***	***Quantitative observation***	***Information from user***
Physical environment	*N/a*	*Required*	*Required*	*Optional*
Technical environment	*Required*	*Required*	*Required*	*Recommended*

As seen in Table I, both qualitative observation and qualitative observation are required in evaluation environments. Therefore, the DIY usability evaluation result described above is a qualitative observation. It is obtained by the observer monitoring the user's behavior. Also, quantitative observation is required simultaneously.

B. Usability Attributes

For quantitative observation, it is necessary first to examine the usability attributes. The attributes are identified in ISO/IEC 25023, which belongs to the quality measurement division of SQuaRE and deals with the quality measurement of systems and software products. ISO/IEC 25023 specifies the usability attributes as the following six [11].

- Appropriateness recognizability is the degree to which users recognize how much the product is appropriate enough to reach their goal. It will depend on whether there is an associated document or tutorials for the product.
- Learnability is how easily users learn a product's usage or manual to reach the goal.
- Operability is how easily users operate a product. For instance, it will express how well-deployed buttons, bars, and menus are in a control panel.
- User error protection is how much a product protects users against making errors while using it to achieve a goal.
- User interface aesthetics is how much a user is satisfied or pleased with the interface and the interaction with the user.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6194
***Article Received**: 22 December 2022 **Revised**: 19 January 2023 **Accepted**: 18 February 2023*

Attribute	Measurement [12]	Reference [7]
Appropriate recognizability	The ratio of X to Y X: number of input and output data with the user successfully understands Y: number of input and output data items available on the interface	Understandability
Learnability	Mean time taken to learn to use a function correctly	Learnability
Operability	The ratio of X to Y X: number of erroneous situations in which the user solves the problem correctly by controlling the product Y: number of possible erroneous situations	Operability
User error protection	1-(X/Y) X: number of the bad situation the user meets while selecting the menu or typing input data for a specific task. Y: max number of possible cases of user input	-
User interface aesthetics	The ratio of X to Y X: number of interface items a user is satisfied with Y: total number of interface items	Attractiveness
Accessibility	1-(X/Y) X: number of interface items the user cannot access Y: max number of interfaces	-

- Accessibility is the degree to which people can use a product or system regardless of their characteristics and capabilities to achieve a specified goal in a specified context of use.

Usability can also be measured as a product quality sub-characteristics identified in ISO 9126. The sub-characteristics are Understandability, Learnability, Operability, Attractiveness, and Usability compliance. These can match the six attributes of usability. Some of them hold the same name, and some have similar meanings in their explanations. The former is "Learnability," and the latter is "appropriate recognizability" of the attributes and "understandability" of the sub-characteristics. So the formula ISO 9126 defined for each sub-characteristics is one of the good references to figure out how to measure usability quantitatively.

C. Measurement

With the meaning and definitions of the five sub-characteristics of usability defined by ISO 9126 and the six attributes mentioned by SQuaRE, we propose the minimum measurement that evaluates them quantitatively. Since ISO 9126 has been withdrawn, the items for measurement are the currently available attributes of SQuaRE. At the same time, ISO 9126 is referred to as supportive data. Its contents are in the table below.

In ISO/IEC 25000, Quality in Use identifies Effectiveness, Efficiency, and Satisfaction as usability metrics. Effectiveness evaluates the accuracy and completeness of achieving a goal, and efficiency evaluates the use of resources related to achieving that goal. Satisfaction is a user's subjective opinion and can be evaluated based on how much the user enjoys using the product, including the level of complaints during use, the user's level of trust in the product, and the physical comfort of using the product [13].

Now, we propose how to measure the metrics. The key is the contribution of attributes for each metric, numbered from 0 to 100. The higher number is, the stronger the influence in the extent to which the attribute contributes to the metric. Any number between 1 and 100 can express contribution, but the sum of contributions of attributes for each metric does not exceed 100. Table III can be an example of setting these contribution values. This contribution value can appear very different depending on the product environment and domain. In Table III, the values presented are written considering the case of a general web application.

TABLE II. MEASUREMENT FOR USABILITY ATTRIBUTES

Metric / Usability Attribute	Effectiveness	Efficiency	Satisfaction
Appropriate recognizability	*30%*		
Learnability		*60%*	
Operability	*40%*		
User error protection	*30%*	*40%*	
User interface aesthetics			*90%*
accessibility			*10%*
	100%	*100%*	*100%*

Once the contribution is expressed as a percentage, as shown in Table III, each metric is calculated as the sum of the values obtained by multiplying the contribution by the corresponding attribute measurement value described in Table 2. As mentioned above, the contribution may differ in specific environments. Table III is an example. Below is how to calculate the metrics in the case of Table III, where M(x) is the measurement of the attribute, x. The measurement can be obtained by applying the measurement presented in Table II.

TABLE III. AN EXAMPLE OF AN ATTRIBUTE-METRIC ASSOCIATION

$$Effectiveness = 0.3 \times M(Appropriate\ recognizability) + 0.4 \times M(Operability) + 0.3 \times M(User\ error\ protection)$$
$$Efficiency = 0.6 \times M(Learnability) + 0.4 \times M(User\ error\ protection)$$
$$Satisfaction = 0.9 \times M(User\ interface\ aesthetics) + 0.1 \times M(Accessibility)$$

M(x) falls on a number between 0 and 1. Through the formula described above, Effectiveness, Efficiency, and Satisfaction will

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6194
***Article Received**: 22 December 2022 **Revised**: 19 January 2023 **Accepted**: 18 February 2023*

be valued from 0 to 1. Of course, a higher number of each metric supports better usability. In this paper, we call this method an *attribute-metric association*.

IV. QUANTITATIVE EVALUATION

A. *Usability Evaluation*

Without the metrics, we conducted usability evaluation under minimum settings in class. The document from the observation was very qualitative and straightforward since there is no template or guideline for evaluation. It would be totally up to the observers' capability. Unless the observers are experts, the evaluation result might not be enough. We share this experience to show that more than qualitative evaluation of non-experts is needed, and more quantitative evaluation is needed.

On October 2022, the students of the HCI class performed usability testing as a team project. The mission is to conduct DIY usability testing in Figure 3 for a web page service. Websites under testing and tasks were assigned as required by Krug's guidelines [9]. And then, a role of user, facilitator, and observer, were assigned to each teammate. Since it was a very in-class informal process, one observer was joined. So all team members gathered together when writing observer opinions. The scenario, given tasks, and websites under testing are as in Table IV.

TABLE IV. DIY USBILITY TESTING FOR CAR SHARING SITES

Settings	Description
Evaluation Scenario	Tom does not have a car. (User is named Tom.) nevertheless, he has just signed up for a class he can only drive a car. Besides, he needs a car occasionally.
Task #1	For going to class, Tom has to go on a 30 km round trip for 3 hours in the afternoon once a week, and he needs a car to handle various tasks all day on Saturday once a month. How much should he pay each month?
Task #2	Tom is concerned about whether the car is always available whenever he needs it. Check out how the site is ensuring the issue.
Website #1	SOCAR (https://socar.kr)
Website #2	GREEN CAR (http://greecar.co.kr)

We got three comments for each combination of two tasks and two sites, and the number of teams is 2. Finally, 24 comments are obtained from the project, but the comments are all narrative. The narrative comments-style evaluation will now be analyzed based on our method, attribute-metric association. This analysis leads to results in quantitative evaluation for usability.

B. *Attribute-Metic Association*

The total number of evaluation comments is 24, but twelve are for each website. Those twelve comments, 1) to 12), are matched to their associated one of the six usability attributes. Since the evaluation comments are not quantitative but in narrative style, it is hard to use the formula introduced in Table 2.

Therefore, we first determine whether the comment is semantically positive or negative. If it is negative, it is set to a value of 0.5 or less, centering on 0.5, and if it is positive, it is set to a value of 0.5 or more. A score difference is given with the number of comments associated with setting the value; 0.4 in case of one negative comment and 0.2 in case of two or more negative comments. The example explained in 3.1 is measured as shown in Table V. The Comments column next to the score is the list of comments related to that attribute, where each comment has its unique number.

TABLE V. MEASUREMENT OF SOCAR AND GREEN CAR

Attributes	*SOCAR*		*GREEN CAR*	
	Score	***Comments***	***Score***	***Comments***
Appropriate recognizability	0.2	1) 3) 11)	0.2	1) 2) 3) 4)
Learnability	0.2	2) 8) 12)	0.4	4)
Operability	0.2	5) 6) 7) 9) 10)	0.2	5) 6) 7) 8) 11) 12)
User error protection	0.4	4)	0.2	3) 4)
User Interface Aesthetics	0.2	1) 6)	0.2	5) 6) 9) 10)
Accessibility	N/A		N/A	

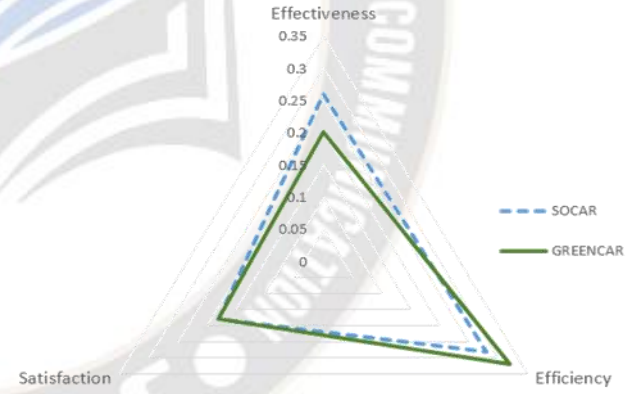

Figure 4. Metrics for SOCAR and GREEN CAR

Since the attribute's score has been calculated in Table 5, we now set up the relationship of the attribute-metric association as in Table III. According to Table III's setting, the formulas for Usability Metrics were already introduced; for Effectiveness, Efficiency, and Satisfaction. Through calculating numbers, the effectiveness for SOCAR is 0.26, efficiency is 0.28, and satisfaction is 0.18. For GREEN CAR, they are 0.2, 0.32, and 0.18, respectively. The result of the metrics is expressed in Figure 4.

V. CONCLUSIONS

Usability is not covered in the traditional software testing process. Typically, testing that software developers mind is unit

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6194
***Article Received**: 22 December 2022 **Revised**: 19 January 2023 **Accepted**: 18 February 2023*

testing, integration testing, system testing, and acceptance testing, as described in V-model. These are all not for usability but for functionality. Software developers mainly focus on implementing its functions and then designing the interface where users communicate with the functions. It is named the inside-out approach. Therefore, Usability testing or evaluation is not actively performed in the software development process.

Instead, HCI experts are more concerned with Usability testing than software developers. It is because they make an outside-in approach in contrast to developers' inside-out approach. They prioritize the interface of products. Some kinds of usability literature explain how to set up participants and environments for usability evaluation and have observers write their opinions while monitoring users' behavior. However, the guideline needs to include how to measure usability quantitatively in the process.

The outcome of the evaluation may depend on the capabilities or experience of the observer. The quantitative metrics for usability will standardize the evaluation result, which would be more meaningful and helpful in upgrading the product's usability. To solve this problem, we have proposed a method called attribute-metric association.

First, we have adopted the software quality model as a criterion. The quality models are included in the ISO standards series, which are ISO 9126 and ISO/IEC 25000 series. We have collected the contents related to usability from ISO9126 and ISO/IEC 25000, named SQuaRE, and developed the attribute-metric association method. It scores association degree in all combinations of six attributes of usability, defined in ISO 9126, and three metrics of usability, described in SQuaRE. Furthermore, the practical example was explained as applying our attribute-metric association method. The example upgrades the evaluation result obtained by Krug's usability testing to a more quantitative one. It showed that attribute-metric association takes out quantitative evaluation results from the existing qualitative evaluation results. In conclusion, the attribute-metric association method of this paper will contribute to improving the usability evaluation results into quantitative evaluation results.

The process of scoring usability attributes with narrative comments can be done by Natural Language Process(NLP) since the sentiment analysis of NLP classifies data as negative or positive. It is planned as future work.

ACKNOWLEDGMENT

This work was supported by the Hyupsung University Research Grant of 2020.

REFERENCES

[1] S. Desikan and G. Ramesh, Software Testing: Principles and Practice, Pearson Education, 2006
[2] A. Cooper, R. Reimann, D. Cronin, and C. Noessel, About Face: The Essentials of Interaction Design, 4th ed., Wiley, 2014
[3] S. Krug, Don't Make me think, Revisited: A Common Sense Approach to Web Usability, 3rd ed., New riders, 2013
[4] P. Ammann and J. Offutt, Introduction to Software Testing, 2nd ed., Cambridge Press, 2018
[5] D. Norman, The Design of Everyday Things, 2nd ed., Basic Books, 2013
[6] J. Nielson, "Usability as barrier to entry", https://www.nngroup.com/articles/usability-as-barrier-to-entry/ , 1999
[7] ISO/IEC 9126, Software Engineering - Product quality Parts 1-4, ISO, 1999
[8] ISO/IEC 25000, Software Engineering: Software Product Quality Requirements and Evaluation (SQuaRE), ISO, https://iso25000.com/index.php/en/iso-25000-standards, 2011
[9] S. Krug, Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems, New Riders, 2009
[10] ISO/IEC FDIS 25066, Systems and software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE) – Common Industry Format(CIR) for usability: Evaluation report, ISO, 2016
[11] ISO/IEC 25023, Systems and software engineering-systems and software quality requirements and evaluation(SQuaRE)-Measurement of system and software product, ISO , 2016
[12] K. Toshihiro, "Usability Evaluation based on international standards for software quality evaluation", NEC technical journal, Vol. 3, No. 2, 2008
[13] R. Schumacher, M. Lowry, and Z. Svetlana, "NISTIR 7742: Customized Common Industry Format Template for Electronic Health Record Usability Testing", NIST, 2014