

Usability Improvement based on Hierarchical Task Analysis (Case Study on i-Caring)

Muhammad Zafif Muttaqy #¹, Angelina Prima Kurniati *², Yanuar Firdaus A.W #³

*School of Computing, School of Computing
Bandung, Indonesia*

¹ zaizhunter@gmail.com

² angelina@telkomuniversity.ac.id

³ yanuar@telkomuniversity.ac.id

Abstract

i-Caring (IT Telkom i-Gracias Collaboration and e-learning) is an LMS (Learning Management System)-based e-learning integrated in i-Gracias own by IT Telkom Bandung. This application is being used to facilitate the implementation of learning activities supporting the face-to-face discussion in class. I-Caring supports some functionalities, such as instructional material storage, assignment, quizzes, and communication through chats and forum. But the great functionalities are not supported by an interactive interface, which is not comply to QUIM (Quality in Use Integrated Measurement) standard. This paper explains our research results of evaluation and recommendation process. This research is using Hierarchical Task Analysis (HTA) which has some indicators to analyze the behavior of a sample lecturers and students which are randomly selected. Based on the questionnaires answers, we analyze the relations of each points on HTA principles using SPSS Statistics 20 tools. The recommendations are then formulated to improve i-Gracias user interface quality.

I. INTRODUCTION

IT Telkom Collaboration I-Gracias and E-learning (i-Caring) are LMS (Learning Management Sistem)-based E-Learning applications integrated in i-Gracias IT Telkom Bandung. I-Caring provides learning modules to support teaching and exercise materials, assignment, communication through chats and forums, and various types of quiz facility. However, these functions lack support of interactive user interface because it does not comply to QUIM (Quality in Use Integrated Measurement) Standard [13]. QUIM is the latest requirement standard and comprises 10 factors which consist of 26 sub-factors. As a result, better recommendation is generated. Nowadays, several researches to evaluate the usability have been conducted through various methods, such as Hierarchical Task Analysis (HTA) and Focus Group Discussion (FGD) (Kumar, 1987). HTA is a systematic method to describe how well-organized some work is to achieve the goal of the work [5]. HTA is selected because this method is developed to collaborate task analyst with people involved in operations. HTA is used to examine potential mistakes and their consequences, interface design verification, exercise procedure identification, operational procedure development and verification, workload assessment, and communication analysis [15]. HTA is very useful for interface designers because it provides task implementation model, enabling the designers to visualize the goals, task, sub-task, operation, and important plans for users. HTA helps to elaborate complex tasks to increase effectiveness in analyzing assignment yet has specific point of view on the task, and also functions as analysis framework which serves as a practical tool for the designers [4]. In accordance with the mentioned characteristics of QUIM and HTA, this method is suitable to evaluate usability of IT Telkom i-Caring. The authors selected QUIM as approach (model) tool to test the correlation of respondents answer score to identify equivalence coefficient between answer score and the method used. HTA is preferred due to its compliance with i-Caring condition in which there are structured steps to evaluate. The result will be used to upgrade website in the future.

II. LITERATURE REVIEW

A. Usability

Although ISO 9241-11 recommendation has been a standard for community specializing in usability, the most frequently adopted definition of usability is the one introduced by Nielsen. According to Nielsen [10], usability refers to: learnability, efficiency, memorability, errors, and satisfaction

B. Hierarchical Task Analysis (HTA)

Hierarchical Task Analysis (HTA) was introduced by Annett and Duncan to evaluate organizations training needs, in which the underlying technique, i.e. hierarchical decomposition, analyzes and represents the behavioral aspects of complex task such as planning, diagnosis, and decision making. HTA breaks task into sub-task and operations or actions. These tasks of components are then graphically represented using a structure chart. HTA requires tasks identification, categorization, subtask identification, and checking of the overall accuracy of the model [4].

Advantages of Using HTA: [5]

- 1) HTA is an economical method of gathering and organizing information because the hierarchical description needs to be developed only to the point where it is needed for the analysis purposes.
- 2) The hierarchical structure of HTA allows the analyst to focus on important aspects of the task which may impact installation safety.
- 3) When used as an input to design, HTA enables functional objectives to be specified at the higher levels of the analysis before final decisions about the hardware is made.
- 4) HTA is developed as a collaboration between the task analyst and people involved in operations. Therefore, the analyst develops the description of the task in accordance with the perceptions of the personnel who are responsible for effective system operation.
- 5) HTA is a starting point to use various error analysis methods to examine the error potentially occurs in the performance of the required operations.

Steps required in applying HTA method are shown in Figure 1.

<i>Step Number</i>	<i>Notes and Examples</i>
1. Decide the purpose(s) of the analysis	<ol style="list-style-type: none"> 1. Design of system/interface/operating procedures/manning. 2. Determine training content/method.
2. Get agreement between stakeholders on the definition of task goals and criterion measures	<ol style="list-style-type: none"> 1. Stakeholders may include designers, managers, supervisors, instructors, operators. 2. Concentrate on system values and outputs. 3. Agree performance indicators and criteria.
3. Identify sources of task information and select means of data acquisition	<ol style="list-style-type: none"> 1. What sources are available? e.g. direct observation, walk-through, protocols, expert interviews, operating procedures and manuals, performance records, accident data, simulations.
4. Acquire data and draft decomposition table/diagram	<ol style="list-style-type: none"> 1. Account for each operation in terms of input, action, feedback and goal attainment criteria and identify plans. 2. Sub-operations should be (a) mutually exclusive (b) exhaustive 3. Ask not only what <i>should</i> happen but what <i>might</i> happen. Estimate probability and cost of failures.
5. Re-check validity of decomposition with stakeholders	<ol style="list-style-type: none"> 1. Stakeholders invited to confirm analysis, especially identified goals and performance criteria. 2. Revert to step 4 until misinterpretations and omissions have been rectified.
6. Identify significant operations in light of purpose of analysis	<ol style="list-style-type: none"> 1. Identify operations failing $p \times c$ criterion 2. Identify operations having special characteristics, e.g. high work-load, requiring teamwork, specialist knowledge etc.
7. Generate and, if possible, test hypotheses concerning factors affecting learning and performance	<ol style="list-style-type: none"> 1. Consider sources of failure attributable to skills, rules and knowledge. 2. Refer to current theory/best practice to provide plausible solutions. 3. Confirm validity of proposed solutions whenever possible.

Fig. 1. HTA Steps [6]

C. Quality in Use Integrated Measurement (QUIM)

QUIM (Quality in Use Integrated Measurement) Standard is a consolidated model for usability. QUIM is hierarchical, consisting of, from the higher to lower level, usability factors, criteria, and metrics. In regard to this, QUIM conforms to IEEE 1061 (1998) which elaborates method to determine quality requirements as well as identify, conduct, analyze, and validate the process and the metrics of product quality [13]. The relationship between factors and criteria in QUIM is shown in Figure 2.

D. Likert Scale

Likert Scale is a scaling method of attitudinal statements using the distribution of responses as the basis to define the scale rating. This is the most frequently used scale in survey-based researches. This scale was first developed by Rensis Likert in

Criteria	Factors									
	Efficiency	Effectiveness	Satisfaction	Productivity	Learnability	Safety	Trustfulness	Accessibility	Universality	Usefulness
Time behavior	+			+						
Resource utilization	+			+						+
Attractiveness			+						+	
Likeability			+							
Flexibility		+	+					+	+	+
Minimal action	+		+		+			+		
Minimal memory load	+		+		+			+	+	+
Operability	+		+				+	+		+
User guidance			+		+			+	+	
Consistency		+			+	+		+	+	
Self-descriptiveness					+		+	+	+	
Feedback	+	+							+	+
Accuracy		+				+				+
Completeness		+				+				
Fault-tolerance						+	+			+
Resource safety						+				
Readability								+	+	
Controllability							+	+	+	+
Navigability	+	+					+	+	+	
Simplicity					+			+	+	
Privacy							+		+	+
Security						+	+			+
Insurance						+	+			
Familiarity					+		+			
Loading time	+			+					+	+

Fig. 2. Relationship between Factors dan Criteria in QUIM [13]

1932 to measure social attitude [3]. The scale rate of each statement is not determined by the degree of favorability but by the response distribution of agreement and disagreement from a group of respondents serving as pilot group. Scaling procedure using summed rating method is based on 2 assumptions [1]:

- 1) Each written attitudinal statement can be approved as favorable or unfavorable statement.
- 2) Response which reflects positive attitude must be rated higher than that of negative one

Figure 3 shows an example of Likert Scale.

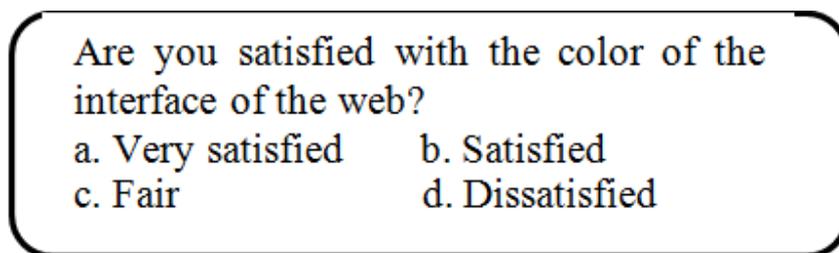


Fig. 3. Likert Scale example

III. SYSTEM DESIGN

A. System Overview

This research uses questionnaire to determine usability success level and web interface relevance. The success is indicated by dimensions retrieved from Hierarchical Task Analysis (HTA) method. The dimensions consist of reference points for determining the questions, as research limitation. Each dimension is broken down into several indicators in the form of questions. The strength of this method is the availability of dimensions. Thus, the success level of a web can be precisely defined, especially if it is equipped with questionnaire data collection method. Data retrieved from the questionnaire are processed using SPSS [11] to figure out whether or not the tested web is proper. General view of the research flow is shown in Figure 4.

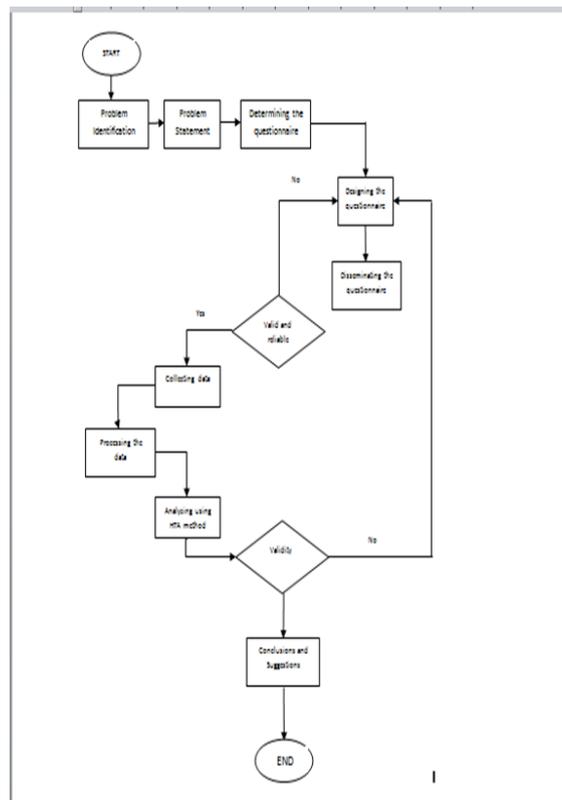


Fig. 4. Research Flow

Remarks: Process 1 is the first step to find out the problems in i-Caring. Process 2 is done by stating the problems in i-Caring. In regard to this, the research focuses on success level of i-Caring usability. Process 3 is conducted according to the desired or needed responses. This research requires responses in the form of precise value, in which the questions are correlated. Therefore, Likert Scale is preferred. Process 4 consists of designing and determining the content of questionnaire which must meet the following principles:

- 1) The questions of the questionnaire must be relevant to the respondents, i.e. familiar to them.
- 2) The clarity of the questions must conform to the respondents' capacity. This is important considering the intellectual level of the respondents might not be equal to that of the question author

Process 5 determines the distribution method of the questionnaire. The distribution type is selected personally considering the geographical distance is not that great. The language used is simple Indonesian language easier to understand by the respondents. Process 6 tests the validity and reliability of the questions. Invalid and unreliable questionnaire must be redesigned. Process 7 collects data of valid and reliable questionnaire. Then the data of the questionnaire is processed using SPSS. Process 8 analyzes data using Hierarchical Task Analysis (HTA) method. Process 9 re-tests the validity to obtain accurate result which concludes this paper.

B. Questionnaire Design

The questionnaire design is done in nine steps. The steps are shown in Figure 5.

Process 4 determines the questionnaire distribution using personally administered questionnaire technique, because the respondents' location is reachable. This technique enables the process of the questionnaire answering to be more personal

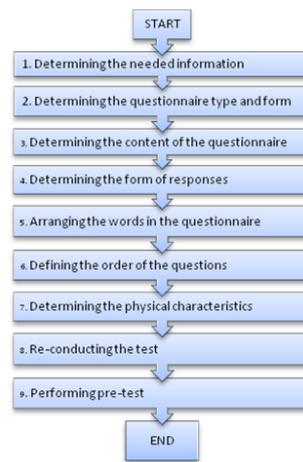


Fig. 5. Questionnaire Design Steps

and the result is more valid. Process 5 and process 6 must be done to ensure that the word choice in the questions and its sequence are ease to understand by the respondents. Process 8 and process 9 are conducted to ensure that the questionnaire is appropriate and in accordance with the needed information. In this research, 30 sheets of questionnaire are distributed to ensure its validity and reliability [1].

C. Questionnaire Dissemination

The questionnaire dissemination is done using interview and personally administered questionnaire method. The researchers read the content of the questionnaire for the respondents and record the responses. The strength of this method is the researches can encourage the respondents to give honest responses suitable with the information needed. The population of this research is the lecturers and students of IT Telkom Bandung registered in 2009 up to 2013. The sample is selected using non-probability judgement sampling method with which the dissemination is more accurate than that of probability sampling method. The number of sample is defined based on the population size, reliability level, and the variant/heterogeneity of the population.

D. Reliability and Validity Testing

Reliability test is performed using Cronbach Alpha [11] method, and the formula is shown in Formula 1

IV. TESTING AND ANALYSIS

Testing is done by defining the population located in an area where the questionnaire is easy to disseminate. The population covers several types of potential respondents background. Thus, the selected population consists of IT Telkom Students registered in 2009, 2010, 2011, 2012, and 2013, as well as IT Telkom lecturers. Based on Charles Gervitz method, population ≥ 5.000 can be represented by 100-500 samples. The selected samples are 100 students and 11 lecturers of IT Telkom

A. Pre-test Result

Pre-test is performed to find the validity and reliability value of each question and dimension distributed to the respondents. The number of respondents in pre-test is 30. The result of validity and reliability calculation of pre-test data is presented in Table 1.

All validity values meet the standard ≥ 0.3 and reliability values meet the standard ≥ 0.6 . Therefore, all questions are considered valid and reliable.

B. analysis Result using HTA

Evaluation of i-Caring Usability by students can collect 127 respondents consisting of students from various faculties at IT Telkom. The results of the evaluation are:

- 1) I-Caring application is difficult to use at first. It is recommended to provide user guide menu to help beginners to learn i-Caring application.
- 2) Popping-out Box Dialogue is unable to close with mouse or keyboard. This hinders users view. It is recommended to provide keyboard navigation system in i-Caring.

TABLE I
RESULT OF VALIDITY AND RELIABILITY OF PRE-TEST DATA

No	Dimension	Number of Question	Validity	reliability	conclusion
1	Efficiency	11	0.729 0.900	0.963	Valid dan reliable
2	Effectiveness	8	0.742 0.895	0.951	Valid dan reliable
3	Satisfaction	4	0.864 0.940	0.963	Valid dan reliable
4	Productivity	5	0.864 0.930	0.949	Valid dan reliable
5	Learnability	4	0.688 0.912	0.869	Valid dan reliable
6	Safety	2	0.949 0.949	0.974	Valid dan reliable
7	Trustfulness	2	0.915 0.915	0.953	Valid and reliable
8	Accessibility	2	0.860 0.860	0.924	Valid and reliable
9	Universality	8	0.793 0.943	0.970	Valid and reliable
10	Usefulness	3	0.882 0.954	0.957	Valid and reliable

- 3) System does not support automatic coma on decimal numbers. It is recommended to provide this feature to facilitate users in reading numbers.
- 4) Simple and frequently-performed tasks cannot be accomplished in 2 seconds. It is due to insufficient support of network system and server. It is recommended to improve network system and server.
- 5) When delay (>15 seconds) occurs in the system, users are not notified that the system is working. It is recommended to display feedback/warning that the program is processing >15 seconds.
- 6) The displayed system and menu organization are not well-arranged, clear, and easy to understand. For example, in calendar and my courses menu. It is recommended to rearrange both menus to make it easier for i-Caring users.
- 7) Instruction, notification, icon, and help information is not helpful enough to the users. It is recommended to upgrade notification and help menu to allow users to know incoming reports and to help users who find difficulties in using the system.
- 8) I-Caring protects users files and data. Users files and data protection in i-Caring application is not understandable by most users. Thus, it is necessary to provide information about the limitation of users files and data protection.
- 9) Users interest to access i-Caring application is low. I-Caring web UI is considered not attractive. As a result, students access this application only when they have assignments and online quizzes. It is recommended to enhance the system and web UI to increase users satisfaction level.
- 10) I-Caring web application is difficult to access. Respondents complain that i-Caring application can only be accessed through PC. It is recommended to provide mobile/Android version to make it easier to access.
- 11) Users find it difficult to resume their work after opening other web pages. Delay occurring when opening web pages hinders users performance. It is recommended to fix web design to quicken the loading process of web pages.
- 12) Icons are not familiar to the users. It is recommended to use standard icons and provide additional explanation on the icons.
- 13) The color is not attractive. Color combination used in the page is considered not interesting. It is recommended to harmonize the color combination of the web.
- 14) Text is difficult to read. It is recommended to adjust the size and type of the fonts.
- 15) Academic information is incomplete and inaccurate. Academic information displayed in the web is considered incomplete and inaccurate.
- 16) Academic information is out-of-date. It is recommended to update academic information regularly.

Furthermore, most of the evaluation results from lecturers as respondents are the same as the ones from the students, with the following additional evaluation:

- 1) Menus are not in a logical order and not suitable with the real work steps. This makes it difficult for users to comprehend and use the web. It is recommended to re-organize the work steps in my course menu.
- 2) The names of commands are not clear. The names of commands for several necessary actions are not clear and confusing. It is recommended to provide explanation appearing when cursor is directed to that action menu, to help users comprehend the functions.
- 3) Syllabus and materials to support the lecturing are difficult to upload. It is recommended to improve server and network connection to support the process of uploading syllabus and materials.

C. Mapping of usability problems

The mentioned evaluation results are then mapped into the menu in i-Caring web. The mapping of usability problems by students is shown in Figure 6, and the mapping of usability problems by lecturers is shown in Figure 7.

Based on the mappings of usability problems by both users above, it is found out that the menus need improving are: Login ID, Home, My Calendar, My Messages, My Courses, Subject Forum, My e-Safety, My Online Users, Notification, Opening Help, and User Guide Menu. The sample of recommended improvement for Home menu is shown in Figure 8.

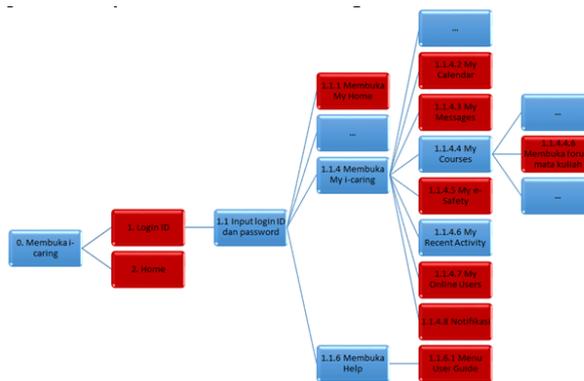


Fig. 6. Usability problems by students

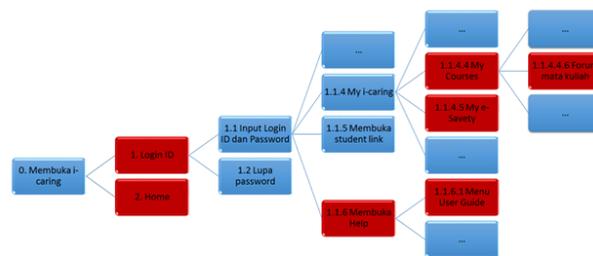


Fig. 7. Usability problems by lecturers



Fig. 8. Home menu improvement

V. CONCLUSION AND SUGGESTION

From this research, it can be concluded that:

- 1) I-Caring web application usability can be evaluated using QUIM (Quality in Use Integrated Measurement) and HTA (Hierarchical Task Analysis) evaluation method.
- 2) The result of double regression test using QUIM shows that:
 - a) Learnability variable in i-Caring usability gives the greatest effect to the satisfaction of IT Telkom students as users, i.e. as much as 63.2%.

- b) Effectiveness variable in i-Caring usability gives the greatest effect to the satisfaction of IT Telkom lecturers as users, i.e. as much as 91.4%.

Suggestions for further researches are as follows:

- 1) Conclusions can be made using various kinds of regression but the capacity and population data should be increased.
- 2) Evaluation method can be combined with other methods (FGD, GOMS, Heuristic, etc) then compared with the standards in HCI

REFERENCES

- [1] Azwar, S . 2012. Realibilitas dan Validitas, Edisi 4. Pustaka pelajar: Yogyakarta.
- [2] Bevan, Nigel. 2006. International Standards for HCI. London. UK.
- [3] Churc Jill, Gilbert A. 2005. Dasar-Dasar Riset Pemasaran. Edisi 4. Jilid I. Alih Bahasa Oleh Andrianti, Dkk. Penerbit Erlangga. Jakarta.
- [4] Cristal, Ellington. 2004. Task analysis and human computer interaction. New York. USA.
- [5] David. 2000. Task analysis techniques. Human Reliability Associates Ltd.
- [6] Diaper, et al. 2004. The handbook of task analysis for human-computer interaction. Lawrence Erlbaum associates. New Jersey.
- [7] Godbole, Nina S. 2005. Software quality assurance : principle and practice. Alpha Science International Ltd. India.
- [8] Kieras. Task analysis and the design of functionality. the computer science and engineering handbook. (second edition). University of Michigan.
- [9] Mantei, Mylopoulos. 2001. Information Systems Analysis and Design.
- [10] Matera, et al. Web Usability: Principles and Evaluation Methods. Milan. Italy.
- [11] Nugroho, Yohanes Anton. 2011. Olah Data dengan SPSS. Skripta: Yogyakarta.
- [12] Santoso, Singgih. 2010. Statistik Multivariat Konsep dan Aplikasi dengan SPSS. Elex Media Komputindo: Jakarta.
- [13] Seffah, Ahmed, et al. 2006. Usability measurement and metrics: A consolidated model. Springer Science + Business Media, Inc.
- [14] The Human Machine Interfaces. The International engineering consortium. www.iec.org.
- [15] Txomin. 2001. Compilation of task analysis methods. KTH industrial engineering and management. Swedia.