Journal of Computer Science 3 (5): 345-352, 2007
ISSN 1549-3636

An Integrated Quantitative Assessment Model For Usability Engineering

[1]Haidar S. Jabbar, [1]T. V. Gopal and [2]Sattar J. Aboud
[1]Department of Computer Science and Engineering, College of Engineering,
Anna University, Chennai 600-025, INDIA
[2]Department of Computer Information Science, Faculty of IT,
The Middle East University of Gradate, Studies, P.O. Box 42 -11610, Amman, JORDAN

Abstract: Many different quantitative assessment models have been proposed to measure and report the usability of a general software product for various business and design purposes. However, there are several problems coupled with existing models that consequently bias and affect the process and results of the usability assessment. Moreover, they do not aid their usage by analysts who are not experienced in the filed of usability. Therefore, an integrated, accurate, consolidated and simple usability assessment model is required, to provide an entire construct of usability for general software products. In this paper, we proposed an Integrated Quantitative Assessment Model for Usability Engineering (IQAMUE) for measuring and reporting usability for general software products. The contribution of the IQAMUE has been done at several points: (1) The investigation into existing models that represents usability factors, either by standard bodies or by well-known researches in the field of usability. As a result, we have proposed an improved comprehensive model, which integrates potential and general usability factors, and measure their related metrics in a standard way (2) We have proposed an adjustable sample size estimation model for usability assessment, which enhances the estimation process, by using historical data to gain an initial idea of the software product, and on present data to predict the complexity of the software product (3) For the applicability purpose of the proposed model, we have conducted an empirical case study for a local e-mail system (Eudora V7) to examine and practice the proposed model.

Key words: Usability Measurent Models, Motivation of the Model, IQAMUE, Eudora Case Study

INTRODUCTION

With the increase demand for software products and websites, the focus on the usability for these needs becomes essential[1,2]. Several studies in the field of usability engineering, have shown that in addition to functionality and other quality attributes, usability is a very important success factor with the huge delivery of diverse interfaces[3,4]. Moreover, usability is increasingly recognized as a vital quality factor for interactive software systems, since most researchers agree that unusable systems are probably the single largest reasons why encompassing interactive systems, computers plus people, fail in their actual use [5].

Generally, the quality of any interactive system can be measured in terms of its usability [6]. However, it is often noticed that people spend huge amount of money on fancy design for a software product rather than spending comparatively less to check this quality. This clarifies the growing numbers of research work in the literature that have devoted to the problem of how to measure and report usability. Several models have been proposed for measuring and reporting usability, however there are a number of critical problems associated with existing models that as a result bias and affect the process and results of the usability assessment. Major drawbacks in current models that they are reported individually, and they do not include most of common potential usability factors and metrics that has been defined by standard bodies[7,8,9] and researchers in the field of usability and they are further not incorporated into a consolidated model. Therefore, a fundamental research challenge for any quantitative assessment model is providing an integrated, accurate, comprehensive and simple usability assessment model[10]. In this paper, IQAMUE was a result of our intention to construct an integrated, comprehensive usability assessment model, for providing accurate statistical critical information for usability assessment.

This paper has been divided into 5 sections. In section 2, we have discussed the problems of existing models and the motivation of our research work. In section 3, we have described our proposed model IQAMUE in detail. In section 4, for the applicability purpose of the model, we have provided an empirical case study of a local e-mail system (Eudora V7) with

Corresponding Author: Haidar S. Jabbar, Department of Computer Science and Engineering, Anna University, Chennai, INDIA

real users and results, to examine the proposed model. Eventually, in section 5 we have concluded concerning our work and point toward the basis for future work.

MATERIALS AND METHODS

At the heart of usability engineering, usability assessment plays a fundamental role in discovering usability problems and benchmarking the level of usability for the software products[11]. As discussed early, an essential issue when developing a quantitative usability assessment model is providing an integrated, accurate, comprehensive and simple usability assessment model.

Throughout the literature, the methods and models of measuring and assessing usability can be divided into two broad categories: those that gather data from actual users[12] and those that can be applied without actual users[13]. In this proposed quantitative model we have not gain any focus on the second type of models, since the role of the users are limited, rather we have concentrated and give attention more on the data from actual users, since it employ representative users to perform the desired tasks. However, we can apply this model also to other types of usability assessment models, such as those in heuristic evaluations[14,15], in future versions of the model.

Nevertheless, existing models of analyzing and measuring usability have critical drawbacks, which later affect and bias the process and results of the usability assessment. Although there are many individual methods for assessing and measuring usability, they have associated problems associated with them.

Below are summary list of major research problems, which also are the motivation of our research work:

- Imprecise (small or high) number of users (sample size) selected for usability assessment.
- Current models are not homogenous.
- Current models are not open standard models.
- Current models only include single factors of usability.
- Current models are measured on different scales.
- They are difficult to use and to communicate.
- Results are reported individually.
- When single factor is used, the chance of high subjectively will exist.
- They do not provide enough meaning and interpretation for business and design matters.

Therefore, a comprehensive assessment model is required to provide an entire construct of usability for general software products. Within our proposed model, we have overcome these problems mentioned before and integrate all issues into one consolidated model. Now, before describing the proposed model IQAMUE and its improvements in detail, it is important to discuss existing problems of models to gain a comprehensive idea about them. We have given this discussion in 3 main sections below.

Usability Definitions and Factors: As mentioned earlier, usability has not been defined in a homogenous way across the standards bodies or usability researchers. Most of these various models do not include all major factors of usability. They are also not well integrated into a quantitative model. Therefore, the first step in IQAMUE was the investigation into existing models that represents usability factors, either by standard bodies such as ISO, ANSI and IEEE or by well-known researches in the field of usability. As a definition, the term usability has been defined in different ways in the literature, which makes it a baffling notion. To demonstrate this issue, we have listed some definitions of usability from different standards bodies next:

- "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", ISO/IEC 9126 [16].
- "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use", ISO 9241−11[7].
- "The ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component", IEEE Std.610.12,[9].
- The industry common standard ANSI[8] has included usability as a major factor, in terms of Efficiency, Effectiveness and Satisfaction in the measurement process.

Additionally, many usability researchers have proposed different factors to measure and report usability. Shackel [17] identified speed effectiveness, error effectiveness, both in Learnability flexibility and attitude as the major factor affecting usability. Nielsen's [18] five criteria of usability: Learnability, efficiency of use, Memorability, errors and satisfaction are often quoted. Hix and Hartson[19] related Learnability, long term performance, Retainability and long term user satisfaction to usability. Preece et al. [20] described usability in terms of throughput, Learnability and attitude. Wixon and Wilson [21] characterized usability by Learnability, efficiency, Memorability, error rates and satisfaction. Shneidermann[22] mentioned usability factor as time to learn, speed of performance, retention over time, rate of errors by users and subjective satisfaction. Constantine and Lockwood [23] defined usability in terms of efficiency in use, Learnability, Rememberability, reliability in use and user satisfaction.

Those various definitions from different resources, shows the confusion in the literature describing usability, which emphasize the need for a comprehensive usability assessment model. In the same time, the majority of models emphases the needs for the general factors of Efficiency, Effectiveness, Satisfaction, Memorability, Learnability, however they are not a well incorporated into existing models.

A Standardized Process: As discussed early, that the methods and models of measuring and assessing usability can be divided into two broad categories: those that gather data from actual users [12] and those that can be applied without actual users [13]. However, all these approaches assess usability independently. Individual measures do not provide a complete view of the usability of software product.

These metrics are typically reported individually on a task-by-task basis, with little context for interpreting the correlation among the metrics of a particular task or across a series of tasks. Therefore, we need to standardize the measurement process by using statistical methods to scale deviated measure and bring them into the same scale. We have used statistical methods to standardize the metrics, which have been already described and used in the literature many by researchers in [27,28].

Usability Assessment Sample Size: Another crucial problem in existing models and also a very important aspect for any usability assessment model is the estimation of the sample size desired for a software product. Once we start estimating the sample size needed for a usability assessment, a baffling question comes in mind: "how many users are enough for a given software product?". Throughout the literature, sample size estimation for usability assessment has been done either as simply guessing, or using a mathematical formula proposed by the researchers. A variety of international research work has been done on this topic, especially by the researchers: Virzi[24], Nielsen and Launder [25] and Lewis [26].

To the best of our knowledge, the majority of existing models estimates the sample size needed for a usability assessment, based on historical data (previous software products). A better estimation should be based on both historical data, which provide an initial idea from previous software products and based on present data, which provide a practical idea for a given software product, which aid us in predicting the complexity of the software product by conducting a simple pre-assessment.

Eventually, a consolidated model encompassing all these improved points was a challenge in existing models. Thus, in our proposed model, we have integrated and incorporated all these issues into a consolidated model IQAMUE, as described below in detail, in the next section.

An Integrated Quantitative Assessment Model for Usability Engineering: (Usability Definitions and Factors): As discussed above that there are many definitions, which confuse the definition of usability and the use of these factors in the models. However, for our proposed model we have selected the factors:

- Efficiency.
- Effectiveness.
- Satisfaction.
- Memorability.
- Learnability.

We have selected these factors for the reasons (1) there are often cited by most publication in the literature (2) there is a common agreement of these factors in the literature based on standard bodies and on usability researchers (3) since we are measuring general software products, we find that those proposed factors used in our IQAMUE, falls in the general nature of the software products. However, other usability factors may be plugged in future versions of the proposed model.

Now for each factor there are a number of related metric(s) that will be measured. Based on our review of existing usability measurement models, we identified a total of (6) specific usability metrics. However, we can add more metrics to a usability factor simply. We have selected: number of errors counted number of errors in a task) and task time (time required to complete the task) for efficiency, task accomplishment (whether the task has completed or not) for effectiveness, the objective satisfaction for satisfaction. These metrics are already known and used in the literature. The first 3 metrics are calculated from the first assessment (session). Now for Memorability we have calculated this metric by counting the number of recalled features from previous task. For Learnability we have calculated this metric by calculating the number of improvements that the user performed within a task in the second assessment. Those 2 metrics are calculated at the second assessment (session). The model uses a task-based assessment, which will be finish on 2 sessions. First assessment (session 1) will be used to calculate the metrics of efficiency, effectiveness and satisfaction. Second assessment (session 2) will be used to calculate the metrics of Memorability and Learnability.

Since this is a standard open model, we can add more metrics to a usability factor simply. As well as, we can plug more usability factors in future versions of the proposed model.

A Standardized Process: Now the second part of the IQAMUE is to standardize the calculated metrics into scalable values. This method is already described in general statistics and in [27, 28]. We have employed this method to complement the measurement process by standardizing the values into a standardized scale.

To standardize each of the usability metrics we need to create a z-score type value or z-equivalent. For the continuous and ordinal data (time, Learnability and satisfaction), we will subtract the mean value from a specification limit and divided by the standard deviation. For discrete data (task accomplishment, Memorability and errors) we will divide the unacceptable conditions (defects) by all opportunities for defects.

Now after calculating the z-scores for all the proposed metrics, we can effectively compare the measurements of the values, since all the values are now set on a standard percentage of (100%) scale. As discussed above, this type of model is not intended for collaborative assessment is intended to assess the usability of the software product at the final stages (pre release) with representative users; it is an independent assessment model, which could be applied only after completing the software product.

The range of products is typical in nature such as (Off-the-Shelf products), where there is large number of users using the software product with diversity of user profiles, an example may be: commercial, accounting, financial, management, banking, insurance, airlines, sales, and medical office products and may be window, web and terminal based interfaces. This model is not intended to assess special and individual software products that differ too much from the general nature as discussed above.

Usability Assessment Sample Size: Within the sample size of the usability assessment, we have proposed an improved and normalized estimation model for the better estimation of the sample size. The proposed model enhances the estimation process, by using historical data to gain an initial idea of the software product, and on present data to predict the complexity of the software product, which is described below:

- Estimating the historical problem discovery rate (λ_α), which is already recommended by the researchers.
- Estimating the adjustment (vertical / domain wise) problem discovery rate (λ_β), by conducting a pre-assessment usability study, to gain a practical idea about the complexity for a given software product.

Integrating both points, gave us an improved and normalized estimation model toward the sample size for the usability assessment studies, as described below in formula (1):

$$SS_{EST\text{-}Adjustable} = \alpha + \beta \quad (1)$$

where Alpha factor (α) is the initial historical estimation value based on $\lambda\alpha$ and Beta factor (β) is the adjustment (vertical / domain wise) value based on λ_β. α is estimated from the historical problem discovery rate either taken as a constant (0.31) or from a set of individual unique problem discovery rates of similar software products (indicated in point 1 above) and β is estimated from the specific discovery rate for a given software product (indicated in point 2 above).

α is already explained in numerous research works mainly by Nielsen, Virizi and Turner and has been discussed early. However, this value is estimated from historical data $\lambda\alpha$ (problem discovery rates). $\lambda\alpha$ is either taken (0.31) as suggested by Nielsen [29] or could be estimated from previous historical data to replace the above value of $\lambda\alpha$. If we go for the historical data, then $\lambda\alpha$ is estimated by creating a matrix of users' numbers and their related individual problem discovery rates.

β value is an adjustment factor, used to provide us crucial information related the sample size for a specific software product. λ_β is estimated by conducting a pre-assessment study of the highest task time of the software product. We have selected the (task time) to represent the software product complexity, because time always measures the whole complexity among other usability metrics [30]. Now for computing β factor, we first start to conduct a pre-assessment usability study for the highest task time for (2-3) users. We need to keep in mind that those 2-3 users should differ in their experience, at least 1 novice and 1 experienced user, to estimate the maximum variance of time among users.

At the present, this mean range time or a less time is the confidence interval value for that particular task. Within this value we start to match between the sample size and the confidence interval to reach an approximate number of users for a usability assessment, using the formula (2) of the continuous data method, reviewing above for this step we can ask the question in a different manner: rather than asking "how many users for a given product?" we ask "how many users for a given product with a desired confidence?".

$$SS_{EST\text{-}COD} = \frac{(t)^2 * (sd)^2}{(d)^2} \quad (2)$$

However, we need to keep in mind that here in this assessment, we do not have the complete sample of users, this means that we can not compute the standard deviation (sd) based on the sample, therefore we need to compute it from a small sample size, to solve this we will use the T-Distribution [31]. This distribution uses a value from T-table, which approximate the standard deviation value for this small sample size. Now altering places of the variables in formula (2), the confidence interval is computed as follows in formula (3):

$$d = \frac{f * sd}{\sqrt{SS_{EST-COD}}} \quad (3)$$

f is t-value, sd is the standard deviation of the small sample size, $SS_{EST\text{-}COD}$ is the number of users for the small sample and d is the confidence interval which is already known from the mean range of task time.

Usually, it does not take more than 5 simple computations to find an approximate sample size for a given software product. Unless we are obtaining a problem discovery rate $\lambda\alpha$ for α, then we also need to obtain problem discovery rate for λ_β from β and then average them to find an adjusted λ value, which is based on the historical data as well as for a given software product. Now for computing λ_β (at 90% likelihood), we use the formula (4):

$$\lambda_\beta = 1 - \sqrt[n]{0.10} \quad (4)$$

where n is the number of users and λ_β is the problem discovery rate for β factor.

Eventually, the final step is to estimate the sample size for the usability assessment based on both historical data and for a given product (at 90% likelihood), given below in the formula (5):

$$SS_{EST-Adjustable} = \frac{-1}{\log(1-\lambda)} \quad (5)$$

where λ is the average problem discovery rate and $SS_{EST-Adjustable}$ is the final estimated sample size number for the usability assessment study for a given software product.

RESULTS AND DISCUSSIONS

A case study: In this section, we have presented a conducted empirical case study for exercising and validating the proposed IQAMUE model. The goal of the usability assessment is to benchmark the usability for a Free (Open Source) local e-mail system (Eudora V.7) [32], with the factors of Efficiency, Effectiveness, Satisfaction, Memorability, Learnability and their related metrics for usability reporting and to explore the significance of the proposed model, the main window of the Eudora case study has been shown below in figure-1.

The Eudora software is a typical local e-mail system found in any computer like outlook express or office outlook with more functionality and features. This free open source software product has been setup on a free post office protocol (POP3) [33] and a free simple message transfer protocol (SMTP) [34], preparing it to start performing the desired tasks in all details.

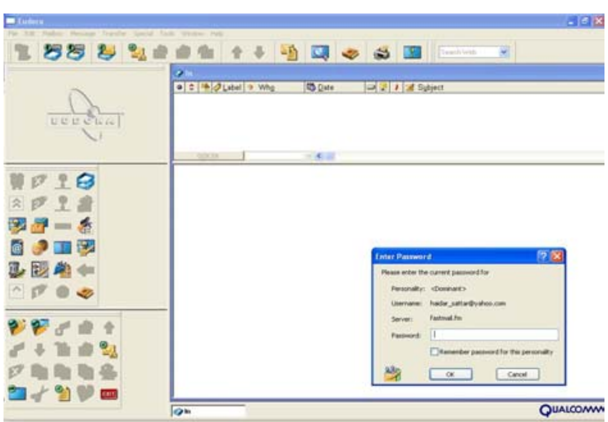

Fig.1: Eudora Main Window

The conducted case study was completed in approximately (2) weeks. A schedule was prepared and confirmed according to the convenience of the selected users. 10 users (5 novice and 5 experienced) were assessed in the first assessment to collect the metrics of efficiency, effectiveness and satisfaction and then again 10 users (5 novice and 5 experienced) were assessed again to collect the metrics of Memorability and Learnability and Satisfaction. The users were students at Anna University, from different academic levels and backgrounds.

Sample Size: As discussed early, the sample size estimation should be done in two parts: historical and real sample size data, if the historical data is available. In this case study, we do not have any historical data, thus we will use real data to predict the sample size needed for the Eudora case study by conducting a pre-assessment usability study based on 2 users. In this case study we have selected 2 users to conducted the longest task, one user is experienced (the usability analysts) and a novice user from the available 5 novice user group. The results of the pre-assessment assessment were: novice user t1 (95) and experienced user t2 (35). The range has been given below:

Time Range = 95 – 35 = 60 Sec.

Based on the above results the confidence interval (range of task time) for this task will be (60) Sec, however we have selected (30) to get better confidence in the results. We will make several attempts to reach an approximate match of the confidence interval; however we will assume here we want to shrink the mean time to (30) seconds (as already discussed). With (6) Attempt and the standard deviation of (42), t value is (2.65) and n is (7), using formula (3), the confidence interval is:

$$d = \frac{2.65 * 42}{\sqrt{7}} = 30.33$$

Unless (30.33) is approximately equal to 30 (30.33≈30), then the recommended number of users for this given product is 7 users. Sample Size = 7, however in this case study we have selected 10 users (5 users in the novice group and 5 users in the experienced group), based on the availability of the users and for illustration purpose.

Designing the Tasks: A well set of comprehensive tasks that represents the complete software product has been proposed. Five intensive tasks has been created, which represent most of the system usability of the

Eudora case study. Therefore, after a detailed analysis of the software product, the main tasks that represent this software are:

Task1: Receive e-mail(s) on a specific account and download them, Task2: Compose (new message), attach file and send, Task3: Add contact to address book, Task4: Add note and Task5: Schedule an event.

Task1: Receive e-mail(s) on a specific account and download them. For this task users are required to do the following:

- Check the mails.
- Enter user account details: name, Password and PoP3 receiving mail server if different for another E-mail account.
- Then go to the mail inbox and open the new messages.
- Then go to the attachments section, locate and download the attachment file in a specific folder.

For this task the possibilities of errors that may occur are (4) errors. We have given (60) seconds as an average time for the completion of this task, a satisfaction rate should be (3.9), Memorability should be (4) and the Learnability should be (1).

Task2: Compose (New Message), attach file and send. For this task users are required to do the following:

- Create New Message.
- Write the address of the receiver.
- Write the subject.
- Then go to attachment section, upload an attachment file from a specific folder.
- Write the message ('Hello Message').
- Send the message.

For this task the possibilities of errors that may occur are (6) errors. We have given (80) seconds as an average time for the completion of this task, a satisfaction rate should be (3.9), Memorability should be (5) and the Learnability should be (2).

Task3: Add contact to address book. For this task users are required to do the following:

- Locate the address book.
- Create new contact address.
- Enter information: name and nick name and other information.
- Enter E-mail ID.
- Add and save contact to the address book.

For this task the possibilities of errors that may occur are (5) errors. We have given (60) seconds as an average time for the completion of this task, a satisfaction rate should be (3.9), Memorability should be (5) and the Learnability should be (1).

Task4: Add a note. For this task users are required to do the following:

- Locate note.
- Select a folder to save the note.
- Write the note.
- Save the note.

For this task the possibilities of errors that may occur are (4) errors. We have given (60) seconds as an average time for the completion of this task, a satisfaction rate should be (3.9), Memorability should be (3) and the Learnability should be (1).

Task5: Schedule an event. For this task users are required to do the following:

- Locate the calendar of the e-mail system.
- Add the title of the event.
- Select event type.
- Enter date and time.
- Enter location.
- Write event note.
- Save event and activate.

For this task the possibilities of errors that may occur are (7) errors. We have given (90) seconds as an average time for the completion of this task, a satisfaction rate should be (3.9), Memorability should be (5) and the Learnability should be (2).

Quantitative raw results of the Eudora case study have been given below in table-1 for the metrics Efficiency, Effectiveness and Satisfaction, and the quantitative raw results for the metrics Memorability and Learnability have been given in table-2. The reported results have been presented based on the possibilities values of the metrics that was defined early by the usability analysts. Afterward the collection of the empirical results, we now used the standardization techniques from statistics methods as discussed early. We have created a z-score and a percentage score for the raw metrics values. At the moment, we are able to effectively compare the measurements, since they are all at the same scale; the normalized and percentage results of the Eudora case study has been given below on table-3 and table-4, respectively.

Table-1: Quantitative Raw Experimental Results of Usability Metrics in the First Assessment for the Eudora Case Study

		No.	Task Time (Seconds)					Task Accomplishment					User Satisfaction (1)					Number of Errors				
First Assessment	Novice Users	User 1	50	81	71	55	99	1	0	0	1	0	4	2	4	3	3.3	2	2	1	2	4
		User 2	55	77	57	49	88	1	1	1	1	1	3.5	4	4.2	4	3.7	1	3	2	1	3
		User 3	40	82	68	52	76	1	0	0	1	1	4.3	2.5	5	4.3	4.5	1	4	2	2	5
		User 4	60	66	50	60	58	0	1	1	0	1	3	5	2.2	1.9	5	4	6	5	4	7
		User 5	62	76	44	41	69	0	1	1	1	1	2.9	4.3	4	3.3	5	4	5	4	4	5
	Experienced Users	User 1	40	59	30	40	50	1	1	1	1	1	4	3.6	3.7	4	4	0	2	1	0	1
		User 2	44	64	33	33	49	1	1	1	1	1	4.2	3.9	4.4	4.1	3.9	0	2	0	0	1
		User 3	39	78	35	32	45	1	1	1	1	1	3.9	4	4.2	4.5	3.4	1	1	0	1	0
		User 4	38	46	41	29	64	1	1	1	1	1	5	4	5	4	4	0	0	0	1	0
		User 5	50	58	25	35	72	1	1	1	1	1	4.7	4.2	4.7	5	4.3	0	0	1	0	2

Table-2: Quantitative Raw Experimental Results of Usability Metrics in the Second Assessment for the Eudora Case Study

		No.	User Satisfaction (2)					Memorability					Learnability				
Second Assessment	Novice Users	User 1	4	2	4	3	3.3	4	2	2	0	2	0	1	0	1	1
		User 2	3.5	4	4.2	4	3.7	3	2	3	2	3	0	0	0	0	1
		User 3	4.3	2.5	5	4.3	4.5	3	3	3	1	2	1	0	1	1	1
		User 4	3	5	2.2	1.9	5	2	1	2	0	1	0	0	0	0	0
		User 5	2.9	4.3	4	3.3	5	1	1	1	1	0	0	1	0	0	0
	Experienced Users	User 1	4	3.6	3.7	4	4	4	4	5	3	5	1	2	1	1	2
		User 2	4.2	3.9	4.4	4.1	3.9	4	5	4	2	5	0	1	1	1	2
		User 3	3.9	4	4.2	4.5	3.4	3	4	4	3	4	1	2	0	1	2
		User 4	5	4	5	4	4	4	5	3	3	4	1	2	1	1	1
		User 5	4.7	4.2	4.7	5	4.3	3	3	5	3	5	1	2	1	1	2

Table 3: Quantitative Normal (Z-Score) Experimental Results of Usability Metrics in the Eudora Case Study

		Usability Metrics											
		Efficiency				Effectiveness		Satisfaction		Memorability		Learnability	
		Time		Errors		Task Accomplishment		Satisfaction		Memorability		Learnability	
		N	E	N	E	N	E	N	E	N	E	N	E
1	Task1	0.75	3.62	-0.25	1.65	0.25	6.41	-0.59	0.97	0.39	1.28	-1.10	0.24
2	Task2	0.57	1.64	-0.43	0.97	0.25	6.41	-0.27	0.18	-0.36	1.0	-1.4	0.45
3	Task3	0.17	4.59	-0.15	1.75	0.25	6.41	-0.02	1.01	-0.15	1.0	-1.79	0.35
4	Task4	1.21	6.41	-0.39	1.28	0.84	6.41	-0.64	0.97	-0.62	1.50	-0.73	0.73
5	Task5	0.75	2.97	-0.48	1.20	0.84	6.41	0.52	0.061	-0.47	1.41	-2.56	0.45

* N: Novice Users, E: Experienced Users

Table 4: Quantitative Percentage Experimental Results of Usability Metrics in the Eudora Case Study

		Usability Metrics											
		Efficiency				Effectiveness		Satisfaction		Memorability		Learnability	
		Time		Errors		Task Accomplishment		Satisfaction		Memorability		Learnability	
		N	E	N	E	N	E	N	E	N	E	N	E
1	Task1	0.77	0.99	0.4	0.95	0.60	1.0	0.28	0.84	0.65	0.90	0.14	0.60
2	Task2	0.72	0.95	0.33	0.83	0.60	1.0	0.39	0.57	0.36	0.84	0.08	0.67
3	Task3	0.57	0.99	0.44	0.96	0.60	1.0	0.49	0.84	0.44	0.84	0.04	0.64
4	Task4	0.89	1.0	0.35	0.90	0.80	1.0	0.26	0.83	0.27	0.93	0.23	0.76
5	Task5	0.77	0.99	0.31	0.89	0.80	1.0	0.70	0.52	0.32	0.92	0.01	0.67

* N: Novice Users, E: Experienced Users

CONCLUSION

Several usability measurement models have been proposed to measure and report the usability of a general software product for various business and design purposes. However, there are many problems coupled with existing models that consequently bias and affect the process and results of the usability assessment. Therefore, an integrated, accurate, consolidated and simple usability assessment model is required, to provide an entire construct of usability for general software products.

In this paper, we have proposed our IQAMUE model which improves the usability assessment measurement process at least in 3 points: a broader integration of general potential usability factors and metrics, employing statistical methods for standardizing the calculated metrics and an improved estimation of the usability assessment sample size. Both theoretical work and the Eudora case study discussed above, indicate that the proposed IQAMUE effectively resolve and integrates most of research problems found in the literature. However, future work is needed to be carried out to plug more usability factors and metrics and to investigate into the standard theoretical rationalization of the proposed model.

REFERENCES

1. Rubin J., 1994. Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests. John Wiley and Sons.
2. Haidar S. Jabbar and T. V. Gopal, 2006. User Centered Design for Adaptive E-Learning Systems. Asian Journal of Information Technology, 5, pp. 429-436.
3. Nielsen J., 1993. Usability Engineering. Boston Academic Press.
4. Patrick W. Jordan, 1998. An Introduction to Usability. Taylor and Francis.
5. Ahmed Seffah, Mohammad Donyaee, Rex B. Kline and Harkirat K. Padda, 2006. Usability Measurement and Metrics: A Consolidated Model. Software Quality Journal, 14, pp. 159-178.
6. Lauesen, S. and Harning S., 2001. Virtual windows: linking user tasks, data models and interface design. IEEE Software, 18, pp. 67-75.
7. ISO 9241. (1992/2001). Ergonomics Requirements for Office with Visual Display Terminals (VDTs), Geneva, International Organization for Standardization.
8. ANSI (2001). Common Industry Format for Usability Test Reports (ANSI NCITS 354-2001). New York: American National Standards Institute, Inc. See also, http://www.nist.gov/iusr.
9. IEEE-STD-610 ANSI/IEEE Std. 610., 1990. "IEEE Standard Glossary of Software Engineering Terminology.
10. Jean Scholtz, 2006. Metrics for evaluating human information interaction systems. Interacting with Computers, 18, pp. 507–527.
11. Haidar S. Jabbar and T. V. Gopal, 2006. An Integrated Metrics Based Approach for Usability Engineering. International Review of Computers and Softwares, 1, pp. 114-123.
12. Kantner L. and S. Rosenbaum, 1997. Usability studies of www sites: Heuristic evaluation vs. laboratory testing. Proc. of the 15th Annual International Conference on Computer Documentation, Salt Lake City, pp. 145-151.
13. Rosenbaum S., 2000. Not just a hammer: when and how to employ multiple methods in usability programs. UPA Proceedings, Ashville NC.
14. Patrick W. Jordan, 1998. An introduction to usability. Taylor and Francis.
15. Nielsen, J. and R.L. Mack, 1994. Usability inspection methods. New York Wiley.
16. ISO/IEC 9126, 2001. Software Product Evaluation-Quality, Characteristics and Guidelines for the User, Geneva, International Organization for Standardization.
17. Shackel, B., 1991. Usability-Context, Framework, Design and Evaluation. In Shackel, B., Richardson, S (Eds.). Human Factors for Informatics Usability. Cambridge University Press, Cambridge.
18. Nielsen, J., 1993.Usability Engineering. Boston: Academic Press.
19. Hix, D. and H.R. Hartson, 1993. Developing User Interfaces: Ensuring Usability Through Product and Process. John Wiley and Sons.
20. Preece, J., Y. Rogers, H. Sharp, D. Benyon, S. Holland and T. Carey, 1994. Human Computer Interaction. Addison Wesley.
21. Wixon, D. and C. Wilson, 1997. The Usability Engineering Framework for Product Design and Evaluation. In Handbook of Human-Computer interaction. Helander, M.G. et al., Elsevier North-Holland.
22. Shneiderman, B., 1998. Designing the User Interface: Strategies for Effective Human Computer Interaction. Addison-Wesley.
23. Constantine, L.L. and L.A.D. Lockwood, 1999.Software for Use: A Practical Guide to the Models and Methods of Usage-Centred Design. Addison-Wesley, New York.
24. Virzi R.A., 1992. Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough?. Human Factors, 34, pp. 457-468.
25. Nielsen J. and Landauer T.K., 1993. A mathematical Model of the Finding of Usability Problems, In Proceedings of ACM INTERCHI'93 Conference, Amsterdam, Netherlands: ACM Press, pp. 206-213.
26. Lewis J.R., 1994. Sample Sizes for Usability Studies: Additional considerations. Human Factors, 36, pp. 368-378.
27. Sauro, J. and Kindlund E, 2005. A Method to standardize Usability Metrics into a Single Score. In Proceedings of the Conference in Human Factors in Computing Systems, Portland, pp 401- 409.
28. Sauro, J. and Kindlund E, 2005. Using a Single Usability Metric (SUM) to Compare the Usability of Competing Products. In Proceeding of the Human Computer Interaction International Conference, Las Vegas, USA.
29. Nielsen J. Why You Only Need to Test with 5 Users, retrieved from: http://www.useit.com/alertbox/20000319.html.
30. Kasper Hornbaek, 2006. Current Practice in Measuring Usability: Challenges to Usability Studies and Research. International Journal of Human-Computer Studies, 64, pp. 79–102.
31. M. Abramowitz and I. A. Stegun, 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dove.
32. http://www.eudora.com.
33. http://www.freepops.org.
34. http://www.postcastserver.com