Journal of Engineering and Applied Sciences 12 (Special Issue 5): 6898-6903, 2017
ISSN: 1816-949X

Usability Metrics for a Web-Based Test Blueprint System

Amando P. Singun Jr.
Department of Information Technology, Higher College of Technology, Muscat, Sultanate of Oman

Abstract: Vast range of educational systems and technologies seamlessly support or facilitate the educational setting which necessitate an increased attention on their usability. Usability is a relative key concept in the field of Human Computer Interaction (HCI) where main focus is on how the human uses the computer (or any interactive system) as a tool to perform, simplify or support a task. Improving the usability of such systems and technologies can increase productivity while the use of available organizational resources could be maximized, at the pleasure of the stakeholders of the institution. The ISO 92141-11 usability framework is used in this study, focusing on the use of usability metrics to measure the usability goals of the web-based test blueprint system in terms of its effectiveness and efficiency as well as user's satisfaction. Effectiveness is measured through completion rates and number of errors whereas efficiency is measured based on the time-based efficiency or overall relative efficiency. Post-task user's satisfaction and test-level user's satisfaction are measured through standardized survey instruments such as Single Ease Question (SEQ) and System Usability Scale (SUS) questionnaire, respectively. The usability goals are dependent on its context of use ranging from the user characteristics, embedded technologies, required tasks and environment. The results of the usability metrics are influenced by the context-of-use or circumstances surrounding the system. The web-based test blueprint system has a SEQ score of 5.55 and SUS score of 90.00 which imply that the web-based system is highly-effective and highly-efficient which remarkably satisfies the users. The SUS mean value is interpreted as a promoter score because users tend to promote the system to others because of its high acceptability level and remarkable ease-of-use.

Key words: Usability metrics, usability evaluation, system usability questionnaire, web-based test blueprint system, acceptability, characteristics

INTRODUCTION

A vast range of educational systems and technologies seamlessly support or facilitate the educational setting which necessitate an increased attention on their usability. Improving the usability of such systems and technologies can increase productivity while the use of available organizational resources could be maximized at the pleasure of the stakeholders of the institution. Usability is a relative key concept in the field of Human Computer Interaction (HCI) where main focus is on how the human uses the computer (or any interactive system) as a tool to perform, simplify or support a task.

On this regard, usability should be seriously addressed so that the humanity could experience the true and universally-accepted meaning of usability which is "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" (Bevan *et al*., 2015).

Usability is a relative concept which is dependent on several factors because different users have different usability requirements in using a specific interactive system. Usability is not a single or one-dimensional property but rather aims for multiple usability goals. The following usability goals describe the usability qualities of a user interface which can be measured through the use of usability metrics:

- Effectiveness which refers to the accuracy and completeness of user goal achievement
- Efficiency which refers to the resources spent by user in order to ensure accurate and complete achievement of the goals. The key measured resource normally is time spent by user in order to achieve the goals
- Satisfaction which refers to the subjective thought of the user regarding the user's attitude, level of comfort, acceptability of use and relevance of application

In this study, usability metrics are used as the statistics of the usability goals of the educational web-based test blueprint system. This study may help those system's usability evaluators who simply rely on descriptive properties of systems without digging into the quantitative bases of measuring usability.

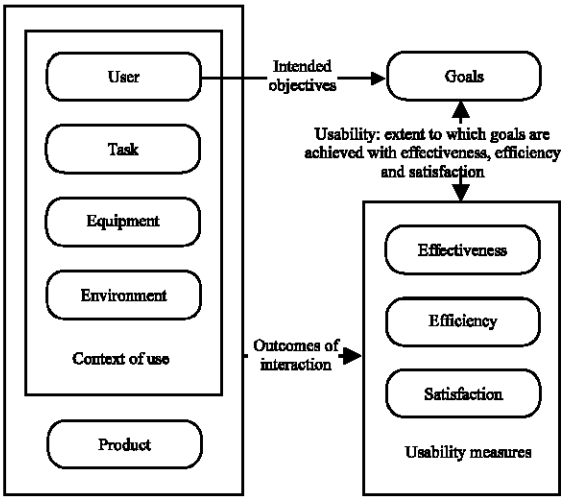

Fig. 1: Usability framework

Conceptual framework: ISO 9241 covers ergonomics of human-computer interaction (Bevan *et al*., 2015). Figure 1 illustrates the usability framework which explains that every system is designed for a specific goal. Such a goal is decomposed into specific and measurable intended objectives which are achieved through a usable interaction between the user and the system (or product). It is noteworthy that each system (or product) is dependent on its specific and distinct context-of-use (i.e., users, tasks, equipment and environment). To measure the usability of a system the usability goals are measured using appropriate usability metrics. The results of the usability metrics are influenced by the context-of-use or circumstances surrounding the system (or product).

In light of the conceptual framework, the researcher developed a web-based test blueprint system that facilitates an automated preparation of a test blueprint that helps teachers to write test questions that are valid, fair, reliable and comprehensive. The intended users are the teachers who will do the following tasks; to prepare a test blueprint successfully, to undergo the automated approval procedures, to prepare a mapping matrix between the test paper and the test blueprint and to view or print required reports. Educational institutions integrate technology into their educational processes in order to increase effectiveness, efficiency and satisfaction. It is easier to conduct usability evaluation on an actual operational system and/or real and deployed systems. For instance, the web-based test blueprint system under study allows the participants to experience the actual look-and-feel of the system which is very interactive, completely functional, user-driven and easily navigated. A screenshot of the web-based system developed by the researcher is shown in Fig. 2.

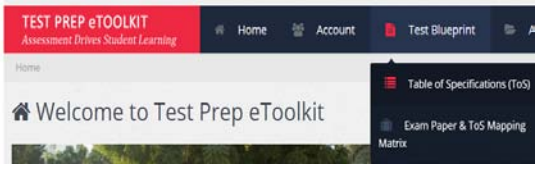

Fig. 2: Test case of this study-test prep etoolkit (http://www.eToolkit.digipowertal.com)

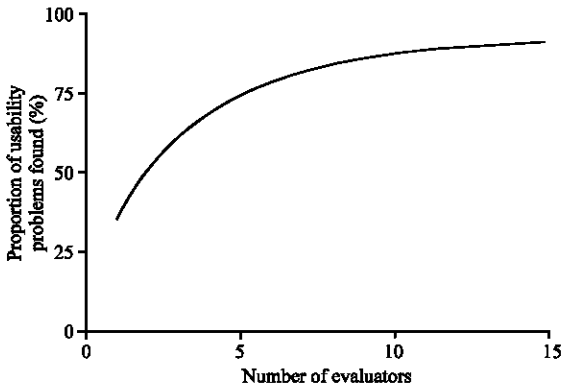

Fig. 3: Usability experts in a usability evaluation

Statement of the problem: This study sought to measure the usability of the web-based test blueprint system using different usability metrics. Specifically, it attempts to measure the extent of compliance of the proposed web-based test blueprint system in terms of the following usability goals: effectiveness, efficiency and satisfaction.

Participants of the study: The participants (N = 5) of this study are double experts. Double experts are those individuals who have multiple areas of expertise. The double experts can effectively determine more usability problems since they are expert users who have strong knowledge base about usability, competence on the use of computers or systems and strong familiarity about the domain-under-study. As illustrated in Fig. 3, it is reasonable to recommend the use of five evaluators in order to identify 75% of the usability problems (Nielsen, 2006). Beyond five creates a decreasing efficiency, since, evaluators will find repeated usability problems in a looping manner.

MATERIALS AND METHODS

This study made use of a descriptive research design using the following data gathering techniques:

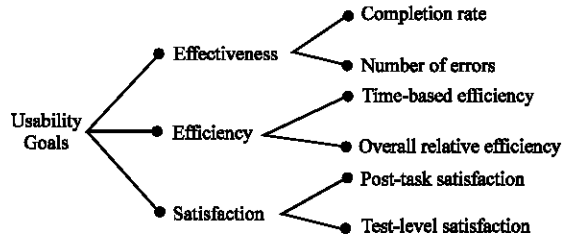

Fig. 4: Usability metrics

Website walkthrough: This is a usability evaluation method in which participants work through the website and perform the required tasks. At the end of each task, a task record had been provided to the participants to key-in their completion rate, the number of errors encountered and the duration of completion.

Standardized satisfaction questionnaire: Two types of standardized satisfaction questionnaire were used in this study: the 1-item post-task questionnaire called Single Ease Question (SEQ) which was completed at the end of each individual task during the website walkthrough and the 10-item test-level satisfaction questionnaire known as System Usability Scale (SUS) questionnaire which was completed by the participants at the end of the usability test session which determined the overall ease-of-use of the system under study.

Data analysis: The level of usability of the web-based test blueprint system can be best measured using observable and quantifiable metrics as illustrated in Fig. 4.

Effectiveness metrics

Completion rate: This is the percentage of completion of specific task. Each successful task gets a binary value of '1' while '0' for an unsuccessful completion of a task. All the binary digits are averaged. The result of the completion rate is dependent on the context-of-use of the task being evaluated. Effectiveness based on the completion rate can be derived using the equation as:

$$\text{Effectiveness} = \frac{\text{Number of tasks completed successfully}}{\text{Total number of tasks undertaken}} \times 100\%$$

Number of errors: Another measurement involves counting the number of errors that the participant makes when attempting to complete a task. Errors may be in a form of a slip, mistake, omission or unintended action that a user makes while attempting complete a task.

Efficiency metrics: Like in the case of effectiveness, efficiency is also affected by the kind of user, task and technology used by the system under study. It can be calculated in one of 2 ways:

Time-based efficiency: This refers to the ratio of effective user's work time to all user's work time:

$$\text{Time based efficiency} = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N}\frac{n_{ij}}{t_{ij}}}{NR}$$

Where:

N = Total number of tasks

R = Number of users

n_{ij} = Result of task i by user j (if the user successfully completes the task, then $N_{ij} = 1$, if not, then $N_{ij} = 0$)

t_{ij} = Time spent by user j to complete task i (if the task is not successfully completed then time is measured until the moment the user quits the task)

Overall relative efficiency: The overall relative efficiency uses the ratio of the time taken by the users who successfully completed the task in relation to the total time taken by all users as shown in the equation as:

$$\text{Overall relative efficiency} = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}t_{ij}}{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}t_{ij}} \times 100\%$$

Satisfaction metrics: Using standardized satisfaction questionnaires, the mean values of the user's post-task satisfaction and user's test-level satisfaction are derived.

Post-task user's satisfaction: The Single Ease Question (SEQ) was used to measure participant's post-task satisfaction. The scale ranges from 1-7 where 1 as the very difficult and 7 as the very easy.

Test-level user's satisfaction: The System Usability Scale (SUS) questionnaire was utilized to determine the overall satisfaction of the participants with a scale of 1-5 where 5 is the most positive response. This 10-item industry standard satisfaction questionnaire is easy, simple and cheap to administer and is even working on a small sample size with reliable and valid results.

To score the SUS, subtract the scale position from 1 on all oddly-numbered items and subtract 5 from the scale position on all evenly-numbered items then multiply the sum of all items by 2.5 to get an overall SUS score that ranges from 0-100. It is noteworthy that the SUS score is not a percentage but rather a normalized score that produces a percentile ranking. The overall SUS score can

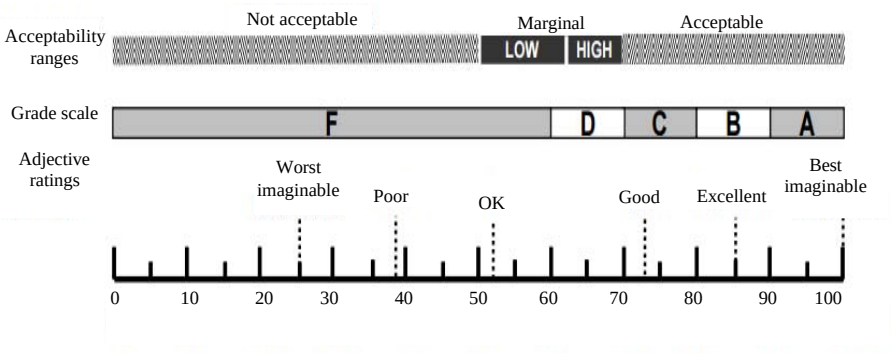

Fig. 5: SUS score interpretations

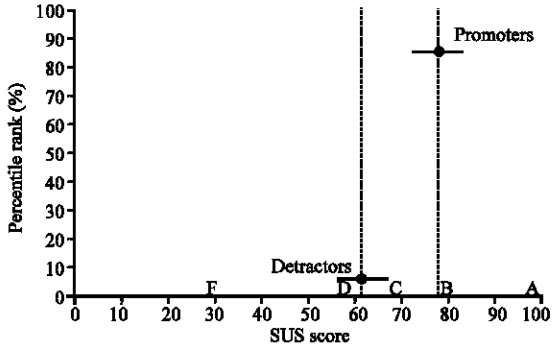

Fig. 6: SUS score promoters and detractors

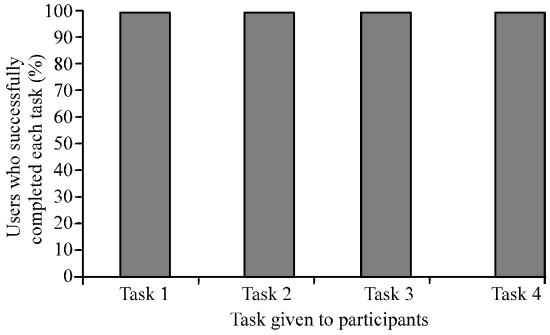

Fig. 7: Completion rate chart

be interpreted as shown in Fig. 5 (Bangor *et al*., 2009). Based on the literature, a SUS score of 68 is considered as above average and anything below 68 is below average (Sauro, 2010, 2011a, b). A SUS score could be interpreted as either a promoter score or a detractor score. As shown in Fig. 6, promoters have an average SUS score of 82 while detractors have an average score of 67 ($p<0.01$). If the system has a SUS score of $\geq$80, it implies that users may most-likely promote the use of the system to other people.

RESULTS AND DISCUSSION

Effectiveness

Completion rate: Throughout the website walkthrough, all the participants (N = 5) had successfully completed all the required tasks (N = 4) namely: preparation of the test blueprint, Undergoing through the approval process, Preparation of the mapping matrix between the test paper and the test blueprint and printing of reports. Using the formula to derive effectiveness, the successful completion rate is tabulated and graphically illustrated in Table 1 and Fig. 7, respectively. The set target is 78% being the threshold rate of completion rate (Sauro, 2011a).

The 100% completion rate fits the aim of teachers (i.e., to accurately produce effective test questions) as they are accountable to the student's learning achievemen. Getting a 100% completion rate implies a high confidence level which can be assigned between 75-100% of all the teachers doing the same tasks.

Table 1: Completion rate

Tasks	Target (%)	Actual (%)
Prepare the test blueprint	78	100
Undergo through the approval process	78	100
Prepare the mapping matrix between the exam paper and the test blueprint	78	100
Print reports (test blueprint and mapping matrix)	78	100

Table 2: Number of errors

Tasks given to users	No. of distinct error per task
Prepare the test blueprint	1
Undergo through the approval process	0
Prepare the mapping matrix between the exam paper and the test blueprint	2
Print reports (test blueprint and mapping matrix)	0

Number of errors: Normally, users of any interactive system may encounter errors (Sauro, 2011a, b). As revealed in Table 2, the participants had encountered distinct errors in Task 1 (N = 1) and Task 3 (N = 2). The average number of errors for all the tasks is 0.75. Half of the given tasks did not yield to any error. The result shows that the web-based system is simple and intuitive.

Table 3: Overall relative efficiency

			Participants				
Tasks	Target	Criteria	1	2	3	4	5
1	1800 sec	n_{ij}	1	1	1	1	1
		t_{ij}	2400 sec	1800 sec	2400 sec	2100 sec	2400 sec
2	0.1 sec	n_{ij}	1	1	1	1	1
		t_{ij}	0.1 sec	0.1 sec	0.1 sec	0.1 sec	0.1 sec
3	900 sec	n_{ij}	1	1	1	1	1
		t_{ij}	1200 sec	900 sec	1200 sec	900 sec	1200 sec
4	0.1 sec	n_{ij}	1	1	1	1	1
		t_{ij}	0.1 sec	0.1 sec	0.1 sec	0.1 sec	0.1 sec

Overall relative efficiency (%): 100.00

Table 4: Post-task satisfaction results using SEQ

	Participants							
Tasks	1	2	3	4	5	Target	Mean	SD
Overall, this task is…								
1	5	5	4	5	5	4.8	4.80	0.447
2	7	7	7	7	6	4.8	6.80	0.447
3	4	3	4	4	4	4.8	3.80	0.447
4	7	7	7	6	7	4.8	6.80	0.447

Overall mean: 5.55

Table 5: Test-level satisfaction results using SUS

Items	Normalized mean	SD
I think that I would like to use this system	4.00	0.000
I found the system unnecessarily complex	3.00	0.000
I thought the system was easily to use	4.00	0.000
I think that I would need the support of a technical person to be able to use this system	2.80	0.447
I found the various functions in the system were well-integrated	4.00	0.000
I thought there was too much inconsistency in this system	3.40	0.548
I would imagine that most people would learn to use this system very quickly	4.00	0.000
I found the system very cumbersome to use	3.60	0.548
I felt very confident using the system	3.80	0.447
I needed to learn a lot of things before I could get going with this system	3.40	0.548

Overall SUS score: 90.00

Efficiency

Overall relative efficiency: The participants completed task 1 within 30-40 min while task 3 within 15-20 min. On the other and tasks 2 and 4 were completed in a split of 0.1 sec. The results show that expert users are task-oriented who wanted to complete tasks in a highly-efficient manner. As reported in Table 3, the overall relative efficiency is 100% which implies that the web-based test blueprint system maximizes the time resources and effort of the teachers which in turn, enhances the classroom assessment processes of the institution. It also implies that teachers can complete each task within the target time duration, thus be able to meet deadlines.

Satisfaction

Post-task satisfaction: As reported in Table 4, all the targets are met except for task 3. Participants had slight difficulty in preparing the mapping matrix which requires a practical knowledge on pedagogy and classroom assessment. Despite of this, the standard deviation shows that responses do not have any major gaps of differences. The overall mean is 5.55 which is within the positive stream of the scale which implies that users experienced a perceived ease-of-use while using the web-based test blueprint system.

Test level satisfaction: As presented in Table 5, all the participant's raw responses had been normalized to form a percentile mean. Interestingly, the responses for all the oddly-numbered items range from 3.80-4.00 which belong to the positive stream of the scale. The overall SUS score is 90 which implies that users are satisfied on the web-based test blueprint system. The SUS score of 90.00 is far above the average SUS score of 68.00 and it is considered excellent with a grade rating of B which is further interpreted as acceptable. It denotes that users feel the system's ease-of-use which prompts them to promote the use of the system to their colleagues.

CONCLUSION

Usability evaluation should be encouraged as an organizational quality practice to periodically evaluate existing information systems. The usability metrics and the research methodologies used in this study are recommended for usability evaluation of any system, since, these are cost-effective, quick and simple to use which yield to valid and reliable results.

SUGGESTIONS

It is recommended that usability metrics should be integrated as a built-in functionality of any system. The acceptability or usability level of a system by the users should always be maintained in order to leverage effectiveness, efficiency and satisfaction. The top management of an organization or institution should strongly support technology integration with usability as the end in mind.

REFERENCES

Bangor, A., P. Kortum and J. Miller, 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. J. Usability Stud., 4: 114-123.

Bevan, N., J. Carter and S. Harker, 2015. ISO 9241-11 Revised: What have we Learnt About Usability Since 1998?. In: Human-Computer Interaction, Kurosu, M. (Ed.). Springer, Berlin, Germany, ISBN:978-3-319-20900-5, pp: 143-151.

Nielsen, J., 2006. Quantitative studies: How many users to test?. Nielsen Norman Group, Fremont, California.

Sauro, J., 2010. What to do with task times when users fail a task. Usability LLC, USA. http://www. measuringu. com/blog/failed-times.php.

Sauro, J., 2011a. A Practical Guide to the System Usability Scale (SUS). Measuring Usability LLC, USA. http://www.measuringu.com/sus.php.

Sauro, J., 2011b. What is a good task-completion rate?. Usability LLC, USA. http://www.measuring usability.com/blog/task-completion.php.