Journal of Communications Technology, Electronics and Computer Science, Issue 6, 2016
ISSN 2457-905X

Evaluation and Comparison of Software Packages in Business Environment Using Scenario-Based Approaches

Mahasen Anabtawi[1], Ibrahim O. Adwan[2], Salaheddin Odeh[1]
Electronic and Computer Engineering Master Program, Faculty of Engineering
[1]Al-Quds University, Abu Dies, Jerusalem, Palestine
[2]Unrwa, Administration Department, Ramallah, Palestine
emanabtawsodeh@eng.alquds.edu, i.othman@unrwa.org

***Abstract*—The competition between companies to produce software (SW) packages and the wide spread of international standards cause users' expectations to become increasingly high. Organizations need to measure and evaluate the usability of SW packages. Usability engineering indicates that we must offer better means for structuring, managing and developing Software packages. In this study, there are two systems: first the employed, commercial system and second the developed prototype with computer graphics and multimedia which was designed to overcome the drawbacks of the commercial system. Both systems are used in a business environment to make various transactions as: ordering supplies; purchasing orders; controlling and editing stocks with their balances on shelves. This study is a scenario-based one that tests and evaluates the performance management standards by doing an objective evaluation and assessment between the two systems: the employed commercial system and the prototype. Scenario-based approaches are becoming ubiquitous in system analysis and design. This study uses discrete-event simulation to analyze the user interface UI of both systems. It analyses and evaluates the main components and elements of the interface and its measurements in the computerized systems. The dependent variables are: effectiveness, efficiency, satisfaction, transparency and consistency. The independent variables are: the user interfaces of both systems and the three scenarios.**

Keywords— Comparative evaluation; prototype; scenario-based; software; usability engineering and testing; user-interface UI.

Introduction

In modern management, change is the only constant factor in modern activities. It is worth mentioning that the employment of new systems (transfer from traditional systems to computerized systems) should be a part of the organization, strategic planning mainly because of globalization and the appearance of multi-national companies. Before making any changes, it is necessary to evaluate the existing systems [1]. The effective computerized administration system requires a thorough analysis of both computers and their users, so the interaction between the user and the interface can be evaluated. This will result in designing effective systems that significantly consider the human behavior and business requirements.

The good computerized systems would achieve: general acceptance, validity, clarity, efficiency, satisfaction and easy exploration. User-interface engineering is considered one of the companies' priorities in their development, that affect the productivity, and improve the responses of customers. The study focuses on user-interface applicability as a part of a computerized system. It forms a new trend that was applied in industrial technology. The application of this new prospect made companies realize that the user-interface applicability experiments are extremely significant to discover the bugs and defects; yet they need time and money. The study would be justified even when the system is already in service. As an example, the British Ministry of Defense received 6600 requests for modification of an employed system. One major reason for such modification was the negligence of the human factor in the design. It should be noted that usability tests for evaluating interactive software systems are beneficial and profitable, as after the well-known rule of thumb that every $10 invested in usability will achieve a return of $100.[2]

In this study, we applied usability engineering to evaluate user interface in business administration and to compare two computerized systems. The first system is a commercial software system which serves as a reference, the second one is a prototype designed with enhanced computer graphics and multimedia issues to bypass drawbacks of the commercial system. The prototype development followed an iterative design process which used the Norman's model as a start point, then enabled human computer interaction, HCI aspects, this included using the ISO standard number TR18529 which includes ergonomic aspects of human-system interaction with software systems [3]. Both systems enable carrying out different business administrative tasks as: ordering supplies, issuing purchase orders, controlling stocks, etc.

In order to carry out the tasks in the study, users receive training on both systems. Three well-trained and educated scholars assisted the investigators in monitoring users. The users who used the commercial software system were 27 users. In our study, 6 users were trained to execute three scenarios (clarified in section III) on both systems: The commercial and the prototype. Using the commercial system as a reference, it was possible to compare and analyze the statistical results using SPSS, we used t-test instead of ANOVA, because the number of subjects was less than 30[4]. The comparison was investigated by giving tasks to the users on both systems, then measuring their satisfaction and ability of recognizing the system.

The study took into account the statistics to evaluate users' interactions with the two systems using different evaluation criteria. The independent variables are: the user interface (UI) of the commercial system and the prototype and the sequence of the three suggested scenarios. The dependent variables are: efficient, error rate, satisfaction, effectiveness, stress, navigation, confidence, transparency, consistency, etc.

I. Designing a Prototype and Stages of Testing

Socio-Technical systems are composed of three subsystems: individual persons, organization and technology. These subsystems affect each other and any change in any one of them will affect the rest. Design rules for the relation between individuals and technology should fit individual's skills and experiences on one hand and foster learning by offering flexible modes of actions to increase in competence on the other hand. For strengthening the individual-organization relation, individuals must be able to make full range of their skills and development and have an impact on the work processes. Finally, the design rule for the organization-technology relation is that the technology should fit organizational structure of the company.

Simulation is the imitation of the operation of a real-world process or system over time. Whether done by hand or on a computer, simulation involves the generation of an artificial history of a system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system [5]. The behavior of a system as it evolves over time is studied by developing a simulation model. This model usually takes the form of a set of assumptions concerning the operation of the system. These assumptions are expressed in mathematical, logical, and symbolic relationships between the entities, or objects of interest of the system. Once developed and validated, a model can be used to investigate a wide variety of "what if" questions about the real-world system. Potential changes to the system can first be simulated, in order to predict their impact on system performance. Simulation can also be used to study systems in the design stage, before such systems are built.

Thus, simulation modeling can be used both as an analysis tool for predicting the effect of changes to existing systems and as a design tool to predict the performance of new systems under varying sets of circumstances. In some instances, simple models can be developed using mathematical methods as the use of differential calculus, probability theory, algebraic methods, or other mathematical techniques. However, many real-world systems are so complex that models of these systems are virtually impossible to be solved mathematically. In these instances, numerical, computer-based simulation can be used to imitate the behavior of the system over time. From the simulation, data are collected as if a real system was being observed. This generated data is used to estimate the measures of performance of the system.

The Prototype was developed and implemented using the iterative design process which consists of the following 4 stages:

1. Studying the users and their tasks as part of task analysis.

2. Making a prototype early in the design phase and then reviewing it with expert users.

3. Testing the prototype usability with typical users (comparative usability testing).

4. Correcting the prototype, testing it again and so until the desired result is achieved (see Fig. 1).[6]

The following chart clarifies the four stages:

In this paper, we'll concentrate on clarifying the third step of the design process; namely, usability testing, which builds an important ingredient of the iterative design process. It is an objective evaluation method that reveals how much the desired requirements of a new prototype are fulfilled. Unlike the subjective evaluation techniques accomplished through questionnaires and surveys, objective evaluation techniques, that are exclusively applied in this comparative usability testing, are achieved in an automated manner.

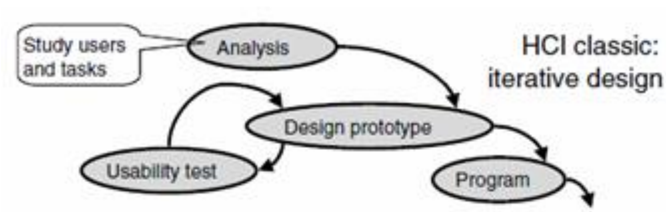

Fig. 1. Iterative design process [7].

For instance, objective evaluation measures users' motor-task performance like the time needed or the number of actions necessary to solve an offered scenario. Users' motor-task performance leans on predicting key-stroking or pointing times. The following are the criteria that have been adapted to investigate and compare the differences between the two systems objectively: time to complete a task, percent of time completed, percent of task completed per unit time, ratio of success to failures, time spent in errors, percent of number of errors, percent of number of competitors better than it, number of commands used, frequency of help and documentation, percent of favorable/unfavorable user comments, number of repletion's of failed commands, number of runs of successes and of failures, number of times interface misleads the user, number of good and bad features recalled by users, number of available commands not invoked, number of regressive behaviors, number of users preferring your system, number of times users need to work around a problem, number of times the user is disrupted from a work task, number of times user loses control of the system, and number of times user expresses frustration or satisfaction. As previously stated, t-test is employed in this phase of the study as the sample size is less than 30.

A. Scenarios and Interdisciplinary Learning

Scenario-based engineering calls for interdisciplinary development with intensive and continuous developer–customer or developer–user interaction. This is attributed to the following reasons:

1. Writing scenarios requires detailed knowledge that is only provided and validated by domain experts, in this case experts are specialized in business administration. The scenarios are developed in an evolutionary manner, brainstorming for potential scenarios led to the topic-centered interviews at individual workplaces for elaborating individual scenarios in detail. In most projects, pre-structured interviews are essential to keep focused and short discussions with the domain experts

2. Developers who are specialized in computer engineering commonly describe current and future system usage scenarios using the language of the problem domain which is vital to establishing good communication with nontechnical domain experts.

3. Customers and users prefer to talk about concrete scenarios rather than abstract models, since the scenarios match their perception of the problem in a better way (except if formally defined). The intensive cooperation during scenario creation and usage ensures the early contribution of customers to the developers' work.

B. Establishing Scenarios

Scenario-based methods for both statistical experiment design and analysis are not only beneficial for descriptions of people using technology in order to reshape their activities; but might be of great significance before a system is built and its impacts are felt.[8],[9] During a usability testing session, the usability expert explains to the user all operations related to the experiment in very simple and clear way. Following are the scenarios used in these experiments:

- First Scenario: The user is given a draft and is requested to process it to both systems. This scenario evaluates the process from the moment of switching on the PC until the end of the process.

- Second Scenario: The user is requested to alter the percent of the financial code after completing the first scenario to both systems.

- Third Scenario: The user is requested to change the currency of the transaction to both systems.

It should be noted that the researcher has done permutations among the users to avoid statistical errors caused by learning effects [10].

C. Statistical Hypothesis

In order to discuss the results achieved by the comparative evaluation, it is necessary to formalize and define statistical hypotheses for the two systems. The dependent variables are: effectiveness, efficiency, satisfaction, transparency, consistency. The independent variables are both the sequence of the scenarios and the user interface (UI) of the two business administration systems. For compensating the influence of such variables on the evaluation results, we have to permutate their sequences. The null and alternative hypotheses (H0 and H1) can be used to test whether there are significant differences between the mean values. The complete hypotheses' formulizations for the investigation are described [11].

D. Design and evaluation process

An iterative design process was used during the development of a graphical user interface (GUI) for an interactive application. The design process consisted of three phases: 1) observation phase in which an understanding of the problem was obtained through structured interviews with end-users, 2) visualization phase during which the GUI was simulated according to a scripted scenario, and 3) an evaluation phase during which feedback on the interface was obtained through three different evaluation phases: expert reviews, user reviews and usability testing.

Following is an explanation of the three evaluation phases:

Expert Reviews: As its name implies, expert reviews are conducted in the presence of human factors specialists – Psychologists - and consists of a combination of standard inspection methods (in this case, heuristic evaluation, cognitive and pluralistic walkthroughs, consistency and standards inspections) all bundled into one inspection session.

On an interface that was simulated, tasks were performed according to a scripted scenario while the experts observed and evaluated the heuristics and components of both the interface and the scenario.

User Reviews: Tests were conducted during a roleplay interaction. The tester assured that all participants understood that they were the office experts to use this new system. As subjects explained how to complete each task, the experimenter navigated through the interface. When an impasse was met (i.e., it just didn't work the way the participant expected), the participant explained how it "should" work and the simulation was modified in real-time. In a few cases, the testers used to take notes about the user favorable and unfavorable reaction when they are in impasse. After completing the task, the tester used to discuss the gaps and the performance of the system with the participants.

Usability Testing: Seven measures were used to evaluate and compare the prototype and the commercial system:

- Percent of number of errors (Table I).
- The number of commands used (Table I).
- Time to complete a task (Table II).

- Number of repetitions of failed commands (Table III).
- Number of times interface misleads the user (Table IV).
- Number of regressive behaviors (Table V).
- Number of favorable/unfavorable user comments (Table V).

II. Results and discussion

Objective evaluation techniques are achieved in an automated manner. Usability of both systems is checked and evaluated objectively. This objective testing decides whether the user interface of the commercial system lies in the acceptance region (if the significance of the commercial system is less than 0.05, it is unacceptable, if the significance of the prototype is greater than 0.05 then it is recommended that the commercial system is replaced by the prototype). T-Test is employed in this phase because the number of population is less than 30.

Table I measures the effectiveness of both systems. Effectiveness is defined as the accuracy and completeness with which specified users can achieve specified goals in particular environment [12].

The results of the test revealed that the cognitive walk-through technique resulted in the greatest quantity of feedback during this iterative design process. As an example; we can see that the number of commands used in the first scenario decreased from 30 to 17 while in the second scenario decreased from 19 to 5 and in the third from 21 to 5 commands. But we can't conclude anything of replacing the commercial system because the SD=0 thus T was not calculated because it is not identified, so further comparisons must be made.

Table II measures the efficiency of the both systems. Efficiency is defined as the recourses expanded in relation to the accuracy and completeness of goals achieved; or the quality of doing something characterized with no waste of time or money [11]. Table II shows the feedback obtained from all expert and user evaluation sessions. From the table below, we see the times of using the mouse in the first scenario decreased from 44 times to nearly 20; while in the second scenario, it has decreased from nearly 27 to nearly 9; while in the third scenario, it is decreased from 29 to nearly 9. On this basis, evaluation is decided to improve the structure of the interfaces and to add several improved features. Again, T values are not significant here and further investigations must be made.

Table III measures the satisfaction of both systems. Satisfaction is defined as a tool to measure user's subjective satisfaction with the computer interface.[13] It contains an overall measure of satisfaction and measures user satisfaction in four specific interface aspects: Screen factors, terminology and system feedback, learning factors, and system capabilities.

Table III shows the feedback obtained from all expert and end user evaluation sessions. From the table; we can see that the number of favorable user comments decreased from 1.33 to 0.50 in both the first and the second scenario; in the third; it is decreased from 1.17 to 0.50.

Table IV measures the transparency of both systems. Transparency is defined as a change in a computing system, such as new features or new component, is transparent if the system after changing adheres to previous external interface as much as possible while changing its internal behavior.

Table IV shows the feedback obtained from all experts and end user evaluation sessions. From the table; we can see the No. of times interface misleads the user decreased from 1.33 to 0.17 in the first scenario; while in the second scenario decreased from 1.33 to zero and in the third from 1.17 to 0.17. It is noted that the significance of the commercial system is 0.01, 0.001, 0.013:all are less than 0.05 so the commercial system is rejected, while the significance of the prototype is .363 which is greater than .05 and the prototype lies in the acceptance region.

Table V measures the consistency of both systems. Consistency is defined as a consistent theory that does not contain a contradiction. The lack of contradiction can be defined in either semantic or syntactic terms. The semantic definition states that a theory is consistent if and only if it has a model, i.e. there exists an interpretation under which all formulas in the theory are true. This is the sense used in traditional Aristotelian logic. The syntactic definition states that a theory is consistent if and only if there is no formula P such that both P and its negation are provable from the axioms of the theory under its associated deductive system.

Table V shows the feedback obtained from all experts and end users evaluation sessions. From the table below, we can see that the number of favorable comments increased from 0.83 to 1 and the number of unfavorable commends decreased from 1.33 to 0.67, the significance of the commercial system is 0.043 which is again less than .05 telling that it lies in the unacceptable region, while of the prototype is 0.102 meaning that it lies in the acceptable region. Thus, it is recommended that the commercial system to be replaced by the prototype.

III. Conclusions

There are several factors that influence the business-administrative tasks carried out through computerized systems accessed by graphic user-interfaces such as: time to complete a task and learn, speed of performance, extending global reach, maximizing impact and integration, responding to demand, retention over time, the rate of errors, and subjective satisfaction. This study aimed to investigate the effect of using computerized business administration on the achievement of users dealing with various business-administrative tasks such as purchase, transactions etc.

The raw data of the experiments were handled statistically by using the Student's test (t-test); then analyzed by SPSS. Through the usability test, it was found that the suggested prototype system is superior to the commercial one because of various modern computer graphics and multimedia issues that were taken into consideration to cover the drawbacks of the commercial system. The two systems were compared by means of an experimental testing. Usability criteria such as: effectiveness, efficiency, satisfaction, transparency and consistency were used. The investigation has shown that the employment of user-centered, user-interfaces in business administration has significant effects on staff achievements in an enterprise. For further investigation of the usability testing, it is recommended that more experimental studies on the role of the user-centered computer systems in business administration are carried out.

It is necessary to note that, before the employment of any product, the organization should take into consideration the usability and applicability of this product; statistical methods are good indicators for the purpose of evaluating by reasoning and guiding the purposeful collection and analysis of data towards the continuous improvement of any process. Results and guidelines achieved through this evaluation help and orient software system developers and user-interface designers in their tasks of both developing of newer computerized systems for business administration with user-centered user-interfaces or optimizing the existing ones.

REFERENCES

[1] Ian Sommerville, “Software Engineering”, Global Edition, Pearson, 2016

[2] R. G. Bias and D. J. Mayhew, Cost-Justifying Usability: An Update for the Internet Age. Morgan Kaufmann, Elsevier, 2011

[3] ISO 13407: Human-centered design processes for interactive systems [Online]. Available: http://www.usabilitynet.org/tools/13407stds

[4] J. Pallant, SPSS Survival Manual: A Step By Step Guide to Data Analysis Using SPSS for Windows, Open University Press, 2013.

[5] Sam Skyborne and Eddie Byrne, ‘Simulation: The Dawn of a Superhero.”, DukeBox.life, 2016

[6] Chris Birchall, “Re-Engineering Legacy Software”, Manning Publications, 2016.

[7] S.Lauesen, User interface design:Asoftware Engineering Perspective Harlow: England:Pearson-Addison-Wesley,2005,pp.41-66.

[8] M. B. Rosson, S. Maass, W. A. Kellog, "The designer as user: Building requirements for design tools from design practice," Communications of the ACM, vol. 31, no. 11, pp. 1288–97, 1989.

[9] K. Weidehhaupt, K. Pohl, M. Jarke and P. Haumer, "Scenarios in system development," Current practice. IEEE Software, vol. 15, no. 2, pp. 34–45, 1998.

[10] Tomayess Issa and Pedro Isaias, “Sustainable Design: HCI, Usability and Environmental Concerns (Human Computer Interaction”, springer, 2015

[11] I. Adwan, “Applying and developing usability engineering for the administration of Procurement and Warehousing Processes in "UNRWA" by Using the Software package - (Reality),” MBA thesis, Institute Business and Economic Studies, Al-Quds Univ., Abu Dies, Jerusalem, 2006.

[12] E. A. Hartmann and A. Westerwick, “Shopfloor Systems Based on Human Skill and Experience,” in Proc. 7th IFAC/IFIP/IFORS/IEA Symposium on Analysis; Design and Evaluation of Man-Machine Systems, Kyoto, Japan, Sept. 16-18, 1998.

[13] B. Shneiderman, C. Plaisant, Designing the User Interface: Strategies for Effective Human-Computer Interaction. 5th Edition, Addison Wesley Longman, 2009.

TABLE I
EVALUATION OF USABILITY FOR THE COMMERCIAL SYSTEM AND THE PROTOTYPE **(EFFECTIVENESS)**

Criteria	Commercial system				Prototype	
	Mean	SD	T	Sig	Mean	SD
First Scenario						
Number of Errors	0.50	0.837	1.464	0.203	0.00	.000(a)
Number of Used Commands	30.00	.000(a)			17.00	.000(a)
Second Scenario						
Number of Errors	.00	.000(a)			.00	.000(a)
Number of Used Commands	19.00	.000(a)			5.00	.000(a)
Third Scenario						
Number of Errors	.00	.000(a)			.00	.000(a)
Number of Used Commands	21.00	.000(a)			5.00	.000(a)

TABLE II
EVALUATION OF USABILITY FOR THE COMMERCIAL SYSTEM AND THE PROTOTYPE (**EFFICIENCY**)

Criteria	Commercial system				Prototype			
	Mean	SD	T	Sig	Mean	SD	T	Sig
First Scenario								
Times of use of mouse	44.00	7.950	13.557	.000	19.83	2.137	22.734	.000
Time to execute the task/minutes	11.17	2.483	11.015	.000	4.83	.753	15.727	.000
Second Scenario								
Times of use of mouse	26.67	7.448	8.771	.000	9.17	1.835	12.237	.000
Time to execute the task/minutes	4.67	1.862	6.139	.002	2.67	.816	8.000	.000
Third Scenario								
Times of use of mouse	29.17	5.115	13.967	.000	9.33	1.966	11.626	.000
Time to execute the task/minutes	6.50	1.871	8.510	.000	1.83	.408	11.00	.000

TABLE III
EVALUATION OF USABILITY FOR THE COMMERCIAL SYSTEM AND THE PROTOTYPE (**SATISFACTION**)

Criteria	Commercial system				Prototype			
	Mean	SD	T	Sig	Mean	SD	T	Sig
First Scenario								
No. of favorable user comments	1.33	.816	4.000	.010	.50	.548	2.236	.076
No. of repetitions of failed commands.	1.50	1.975	1.861	.122	.17	.408	1.000	.363
No. of times the user expresses satisfaction	1.17	.408	7.000	.001	.83	.408	5.000	.004
Second Scenario								
No. of favorable user comments	1.33	.516	6.325	.001	.50	.837	1.464	.203
No. of repetitions of failed commands.	.67	.816	2.000	.102	.17	.408	1.000	.363
No. of times the user expresses satisfaction	1.00	.000(a)			1.00	.000(a)		
Third Scenario								
No. of favorable user comments	1.17	.408	7.000	.001	.50	.548	2.236	.076
No. of repetitions of failed commands.	.83	.983	2.076	.093	.17	.408	1.000	.363

No. of times the user expresses satisfaction	1.17	.408	7.000	.001	.50	.548	2.236	.076

TABLE IV
EVALUATION OF USABILITY FOR THE COMMERCIAL SYSTEM AND THE PROTOTYPE (**TRANSPARENCY)**

Criteria	Commercial system				Prototype			
	Mean	SD	T	Sig	Mean	SD	T	Sig
First Scenario								
No. of times interface misleads the user.	1.33	.816	4.000	.010	.17	.408	1.000	.363
Second Scenario								
No. of times interface misleads the user.	1.33	.516	6.325	.001	.00	.000(a)		
Third Scenario								
No. of times interface misleads the user.	1.17	.753	3.796	.013	.17	.408	1.000	.363

TABLE V
EVALUATION OF USABILITY FOR THE COMMERCIAL SYSTEM AND THE PROTOTYPE (**CONSISTENCY)**

Criteria	Commercial system				Prototype			
	Mean	SD	T	Sig	Mean	SD	T	Sig
First Scenario								
No. of favorable user comments	.83	1.169	1.746	.141	1.00	.000(a)		
No. of unfavorable user comments	1.33	1.211	2.697	.043	.67	.816	2.000	.102
No. of regressive behavior	1.50	.837	4.392	.007	.50	.548	2.236	.076
Second Scenario								
No. of favorable user comments	1.00	.894	2.739	.041	.17	.408	1.000	.363
No. of unfavorable user comments	1.67	.816	5.000	.004	.33	.516	1.581	.175
No. of regressive behavior	1.33	.816	4.000	.010	.33	.816	1.000	.363
Third Scenario								
No. of favorable user comments	.83	.753	2.712	.042	1.00	.632	3.873	.012
No. of unfavorable user comments	1.00	.632	3.873	.012	.50	.548	2.236	.076
No. of regressive behavior	1.17	.408	7.000	.001	.50	.548	2.236	.076