

USABILITY TESTING OF FITNESS MOBILE APPLICATION: CASE STUDY ADED SURAT APP

Ryan Alturki and Valerie Gay

School of Electrical and Data Engineering, University of Technology Sydney, Sydney City, Australia

ABSTRACT

Obesity is a major health problem around the world. Saudi Arabia is a nation where obesity is increasing at an alarming rate. Mobile apps could help obese individuals but they need to be usable and personalized to be adopted by those users. This paper aims at testing the usability of a fitness mobile app "Aded Surat", an app in Arabic language. This paper presents an extensive literature review on the attributes that improve the usability of fitness apps. Then, it explains our methodology and our set up of a trial to test the usability of Aded Surat app that is popular in Saudi Arabia. The usability attributes tested are effectiveness, efficiency, satisfaction, memorability, errors, learnability and cognitive load. The trial is done in collaboration with participants from the Armed Forces Hospitals - Taif Region in Saudi Arabia. The results highlight that the app failed to meet with the usability attributes.

KEYWORDS

Usability, Mobile Application, Obesity, User Experience

1. INTRODUCTION

Obesity is defined as an excessive storage of energy in the form of fat [1]. According to the facts provided by World Health Organization (WHO) Media Centre, 13% of world's adult population is considered obese, and 39% of the adult population is believed to be overweight. The prevalence of obesity around the world has doubled between 1980 and 2014 [2]. Saudi Arabia is one such country where obesity is increasing at an alarming rate. The study by Coronary Artery Disease in Saudis (CADISS) in 2005 found that 35.5% of people in the country are obese which means every third person in the country is affected. A National Nutrition Survey of 2007 mentioned that obesity is a significant concern because the prevalence of obesity in the men (14%) and women (23.6%) in Saudi Arabia [3]. Overweight and obesity are considered as major risk factors for various chronic diseases such as cancer, cardiovascular diseases and diabetes [4-8].

The health problems and diseases that result from obesity have encouraged a lot of researchers to discuss how the condition can be overcome or prevented [9-14]. Most of the research work states that obesity can be overcome by increasing physical activity and changing eating behaviour. However, it is sometimes very difficult to motivate obese individuals to change their lifestyle and become involved in physical activity. Research has shown that the most effective behaviour change related to fitness and health occurs through behaviour interventions [15-18]. Mobile technology such as mobile applications (apps) have been found to be a very useful intervention tool for increasing physical activity because through their unique features these apps motivate individuals to achieve their fitness goals [19-22].

Fitness apps are becoming increasingly popular both around the world and in Saudi Arabia. Smartphones and their apps have seen an exponential growth in their usage in Saudi Arabia in

recent times. Researchers ranked the country third overall in terms of global smartphone usage penetration [23]. In 2016, smartphone users in Saudi Arabia are estimated to be near 15.9 million and this figure is estimated to increase to 19.1 million by 2019 [24]. Because of cultural restrictions many people, but especially women, find it easier to interact publically and socially in a virtual environment through mobile apps on smartphones. The increasing ubiquity of smartphone technology provides an opportunity to develop an Arabic app to help fight obesity.

Sometimes apps development can cost millions of dollars but most of the apps fail miserably [25]. Of all the branded apps, 80% are downloaded less than one thousand times and only 1% has been downloaded one million times or more. After downloading, 25% of mobile apps are never used again [26]. Ample market research suggests that the main reason of failure of mobile apps is the lack of usability [25, 27-28]. The usability of mobile apps enhances user experience (UX) and can play a significant part in the success of the mobile apps [29-32].

The population of obese and overweight individuals in Saudi Arabia is increasing at an alarming rate. Therefore, the increasing use of health and fitness apps in Saudi Arabia is an opportunity to introduce a technological solution which involves developing an app will be popular among obese individuals. There are many Arabic health and fitness apps available but to our knowledge none of these apps have been built with the purpose of enhancing the usability of the app to motivate people to lose weight by considering usability attributes and factors. Moreover, there is no research on fitness and health apps that outlines guidelines for usability. Moreover, to the best of our knowledge, there is no Arabic app that uses any specific features that enhances UX for obese individuals. This leads to our research problem:

- How to improve mobile fitness apps usability to help obese users to reach their health and fitness goals?

This paper presents our method and results of the usability of “Aded Surat”, an Arabic-language fitness mobile apps. Seven usability attributes: effectiveness, efficiency, satisfaction, memorability, errors, learnability and cognitive load were tested collaboration with the Armed Forces Hospitals in the Taif Region of Saudi Arabia, which provided the participants. All participants are people suffering from obesity and are motivated to lose weight.

2. RELATED WORK

2.1. Fitness Apps In Saudi Arabia

Alnasser et al. examined 65 Arabic fitness apps to determine the level of adherence for each app to the 13 evidence-informed practices [33]. The Centers for Disease Control and Prevention, National Institutes of Health, the Food and Drug Administration and the US Department of Agriculture determined these 13 evidence-informed practices [34]: 1-BMI is determined and explained; 2-fruits and vegetables are recommended and tracked for daily servings; 3-Physical activities are recommended for daily use; 4-recommendations for drinking water and tracking the daily consumption; 5-recording and tracking the daily consumption of food; 6- a calorie tracker is provided for maintaining calorie balance; 7-advising goal-setting to lose 1 to 2 lb per week; 8-portion control information is provided; 9-Advising users about ways to read and understand nutrition labels; 10- a weight-tracking feature should be provided; 11-physical activities are tracked for daily use; 12-recommending and providing a tool for planning meals; 13-providing a social network among users or allowing users to share via popular social networks, for example Facebook, Twitter, Instagram or Snapchat.

The result of this study stated that there is no app that has more than six evidence-informed practices and only nine apps had between four to six. Therefore, it is clear that there is no Arabic fitness app which successfully adheres to all 13 practices and there is an essential need to address this issue.

2.2. App Selection Process

All of the apps in Google Play or Apple Store are not free. Fitness apps can be divided into three levels:

- At Level 1, apps are free to download but they do not have all the features. The user needs to subscribe and make payments to access extra features;
- At Level 2, the apps are not completely free. The user must pay to download the apps;
- In Level 3 the apps are completely free to be downloaded.

We selected a fitness app to examine its usability and identify how the features in the app affect UX. The app has a high level of adherence to the 13 evidence-informed practices. The app selected is popular in Saudi Arabia and has high ratings on both Google Play and Apple Store so we expect this app to have special features and to be usable. The study of the app will help us determine how the usability in fitness apps can be increased. The app selected is free so that the participants in the usability testing can access them without cost to themselves.

Aded Surat: This app has nine evidence-informed practices out of 13 and has more than 3500 android users. According to Apple store it was ‘the best health and fitness app’ for 2014 and 2015 [35]. It is also compatible for integration with the Health app on the iPhone. On both Google Play and the Apple Store, this app has a rating of more than 4.5. It is free in both stores. It also has one of largest databases of Arab and local food; it is updated daily. It is a very popular Arabic app for monitoring one’s consumption of calories [35-37].

2.3. Usability

Usability is considered one of the main factors that define the success of a smartphone [38]. Usability can be defined as a multidimensional characteristic of any product. International Standards Organization (ISO) standard 9241-11 gives the meaning of usability as 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” [39-40]. This definition of usability has been accepted widely [41]. In 2011 it was replaced by ISO/IEC 25010. This form includes a model of software quality that portrays usability as the degree to which a satisfied user can efficiently and effectively attain certain goals under specific conditions. The term UX is used extensively in contrast to usability. The two terms are used interchangeably. However, UX has a much broader meaning than the term usability [42]. It can say that usability is more concerned with how easy the product and display features are to use. UX includes the user and the product’s complete interaction as well as the thoughts, feelings and perceptions that are the results of this interaction [43].

2.4. Usability Models and Attributes

The usefulness of portable devices has significantly improved over time and operators are now capable of executing numerous tasks using their smartphones. Customers have the option to choose from a range of mobile phones and there are many apps available for them. This has made it necessary for developers to design mobile devices that enhance the usability experience. The definition of usability has evolved over time and the advent of mobile devices has certainly changed the meaning of the term. There are more attributes that have become important and the

traditional models of usability have been challenged and altered by various researchers. Below we discuss some of the usability models and this will show how the perception of usability has changed over time.

2.4.1. Eason Usability Model

One of the very first traditional models of usability was proposed by Ken Eason in 1984. According to Eason, usability is an interaction between three variables: task, system and user, which feed into the user's response to each task episode [44]. Over the course of a number of task episodes, a pattern of response develops, which indicates more or less usability. In the figure below, I can see that the main independent variable is the system. The system is defined in terms of the functions it can perform. The user characteristics show how well the user can use the system. Task characteristics show how easily and regularly it can be performed by the user. Then these three variables decide the usability depending on the reaction of the user. A positive outcome will mean continued user learning [44].

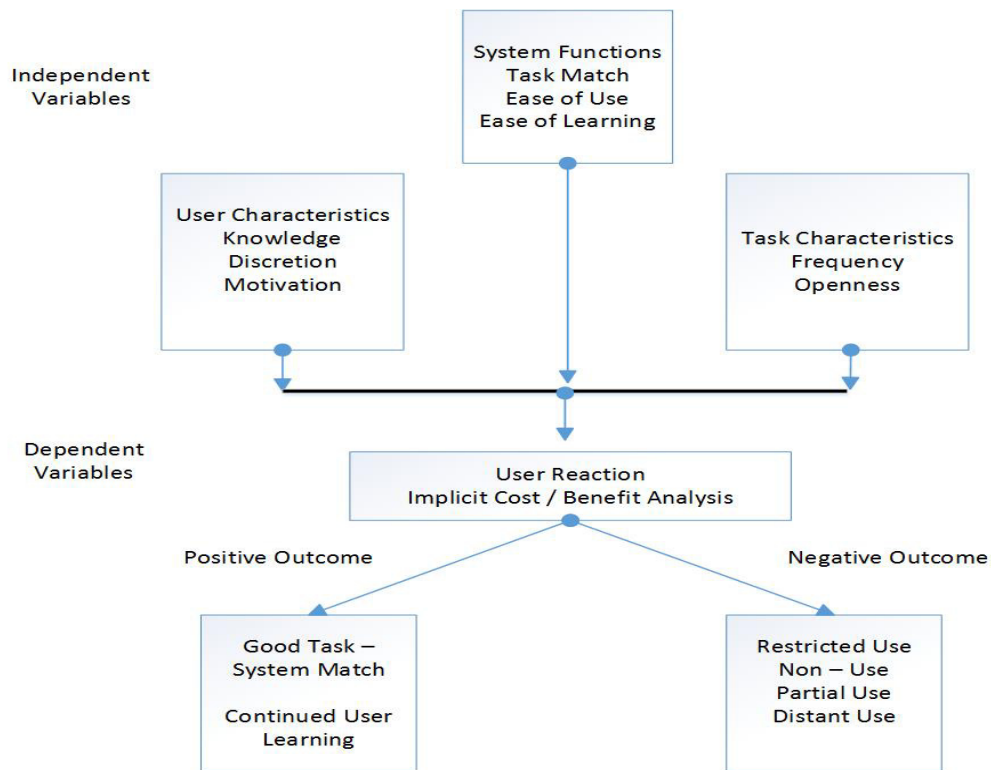


Figure 1. Eason usability model

2.4.2. Dix Usability Model

Dix et al. [45] introduced a usability model, which is very similar to Nielsen [46] model of usability. The findings of Dix et al. include the same attributes as those of Nielsen but their model placed more factors under each attribute. The comparison between the Nielsen and Dix model was conducted by Abran et al. [47] which is summarized below:

Table 1. Comparison between Nielsen’s and Dix’s model.

	Nielsen	Dix
Effectiveness	Percentage of tasks accomplished	Percentage of tasks accomplished
		Ratio on failure of handling
		Percentage of tasks achieved per unit of time
Efficiency	Time to complete one task	Time to complete one task
	Time spent on errors	Time spent on errors
	Error percentage	Error percentage
	Documentation or use frequency of help section	Documentation or use frequency of help section
	Number of failed repeated commands	Number of failed repeated commands
Satisfaction	Number of times that user expresses his frustration	Number of times that user expresses his frustration
		Rating scale for users' satisfaction with functions
		Percentage of users' favourable and unfavourable comments
Learnability	Time to learn	Time to learn

Dix et al. expanded upon the factors in some of the attributes that can affect the usability of the product. However, it can be argued that some of the factors are inherently included in each attribute’s definition [45].

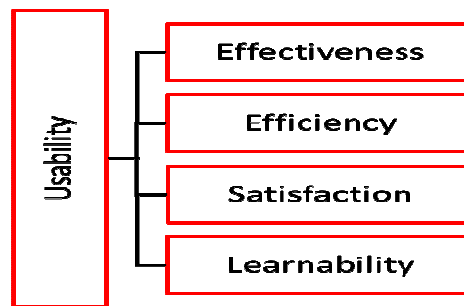


Figure 2. Dix usability model

2.4.3. Norman Usability Model

Norman [48] model of usability aimed at suggesting design principles, which can make the product less complex and easy to use, thus improving its usability.

- Visibility: The visibility attribute should help the user understand the device and give the user alternatives of action.
- A good conceptual model: Designers should design a product that is intuitive and conceptual; the product should be capable of presenting operations consistently.
- Good mapping: Good mapping is an important feature for improving usability because it helps the customer form relationships between actions and results. Good mapping makes

it easy to determine control and their effects and makes it easy to determine the relationship between system state and what is visible.

- Feedback: A good design can enhance the usability by providing the user with continuous feedback regarding results of their actions.

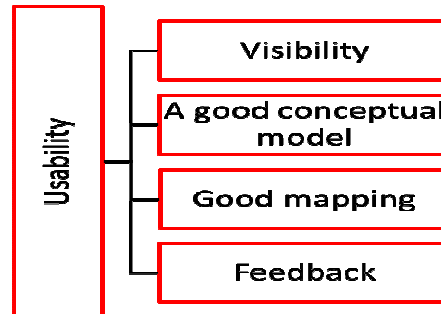


Figure 3. Norman usability model

2.4.4. Shneiderman And Plaisant Usability Model

Shneiderman and Plaisant's model of usability is influenced by Nielsen and the ISO. Their model to a great extent combines and complements these two previous studies [49].

Time taken to learn and speed of performance can be considered part of efficiency in the ISO definition. Time to learn in the Nielsen model can be compared to retention over time. The rate of errors is similar to number and severity of errors. Rate of errors impacts upon the attribute of efficiency in Nielsen and ISO. Subjective satisfaction is similar to the attribute of satisfaction in Nielsen and the ISO [49].

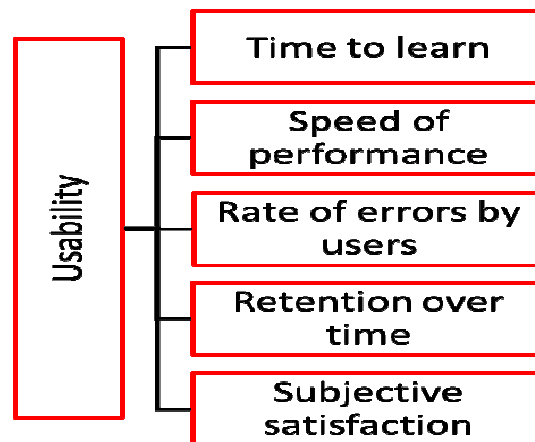


Figure 4. Shneiderman and Plaisant usability model

2.4.5. Yeh Usability Model

Yeh [50] believed in creating a design so that it provides ease, efficiency and enjoyment to enhance the usability of the product. In turn, this makes the product more interactive to use. It is referred to as '3e indicators'.

- Easy: This is the ability of the system to reduce four particular elements, which are physical work, memory work, unnecessary work, and visual work.
- Enjoyable: This is the ability of the system to bring physiological, ideological, sociological or psychological enjoyment.
- Effective: The system should help the user in successfully achieving goals and completing tasks.

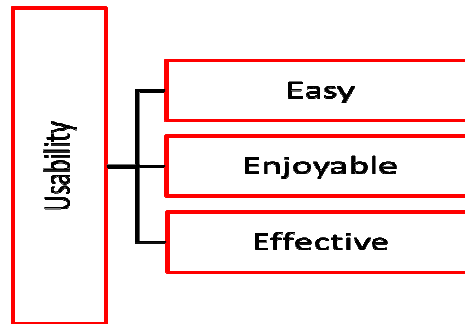


Figure 5. Yeh usability model

2.4.6. Wang And Huang Usability Model

Wang and Huang's [51] study focused on describing usability principles and interface design. They believed that the previous models of usability could be summarised into four principles; these include all the usability attributes defined by Nielsen [46], Norman [48] and Yeh [50]. Their research outlined four principles of usability and sixteen interface attributes. They used these attributes to conduct usability testing and a questionnaire.

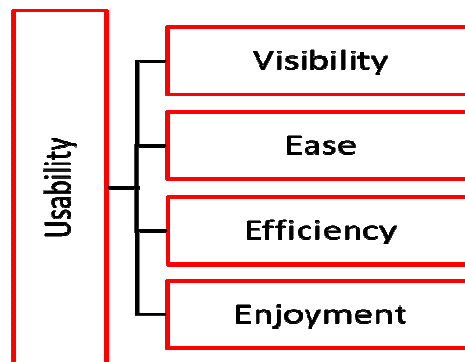


Figure 6. Wang and Huang usability model

3. METHODOLOGY

3.1. Usability Metrics Selection

The ISO/IEC 9126-4 makes the recommendation that any usability metric must make reference to effectiveness, efficiency and satisfaction. Attributes such as memorability, errors, cognitive load and learnability are linked to the efficiency and effectiveness of the app. Whilst they each measure the effectiveness and efficiency of these apps, they do so from a specific perspective. If

an app has less errors, it means that it is effective because the user can perform more tasks in less time without repeating the tasks with errors. Similarly, if an app has better learnability, it helps the user undertake more tasks accurately so it is more effective. The user therefore becomes more efficient in their completion of these tasks. Any of the above features, improve usability and user satisfaction. This is the reason that usability metrics usually include effectiveness, efficiency and satisfaction as the important features for improving usability [52].

3.1.1. Usability Metric For Effectiveness

Effectiveness can be measured using the completion rate of tasks. However, another measurement that can be used is the number of mistakes that users make when trying to finish a task. Effectiveness can therefore be defined as a percentage by utilising the simple equation represented below [53].

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

3.1.2. Usability Metric for Efficiency

Efficiency is used as a tool to measure the time taken to finish a task. It is usually the time taken by participants to complete a task. Efficiency can be calculated using two methods: Overall Relative Efficiency and Time-Based Efficiency [53].

$$\text{Overall Relative Efficiency} = \frac{\sum_{i=1}^R \sum_{j=1}^n n_{ij} t_{ij}}{\sum_{i=1}^R \sum_{j=1}^n t_{ij}} \times 100\% \quad \text{Time Based Efficiency} = \frac{\sum_{i=1}^R \sum_{j=1}^n \frac{n_{ij}}{t_{ij}}}{NR}$$

Where:

- R: number of users
- N: number of tasks.
- n_{ij} : result for task (i) by user (j). if the task is completed successfully, then $n_{ij} = 1$, otherwise $n_{ij} = 0$.
- t_{ij} = time spent by user “j” to complete task “i”. If the user does not complete the task successfully, then the time will be measured until the moment the user gave up from the task.

3.1.3. Usability Metric For Satisfaction

Users’ satisfaction can be determined through standardized questionnaires that measure satisfaction. These can be dispensed after each task or following the usability testing session. Once the user attempts a task, they are given a questionnaire to measure the difficulty of task and the task level satisfaction. Post-task questions can take various forms: ASQ, Subjective Mental Effort Questionnaire (SMEQ), Single Ease Question (SEQ), Usability Magnitude Estimation (UME) etc. From the above list, we will use SEQ as recommended by Sauro [54]. SEQ has the advantage in that it is brief and simple to answer as well as being easy for the experimenter to conduct and then tally the results. The SEQ in this case is “Overall, how easy or difficult did you find this task?”. This SEQ has a rating scale of 7 points where 1 is very easy and 7 is very difficult. The level of satisfaction is found via a formalized questionnaire for users to gain an overall idea of how easy the app is to use. There are different types of questionnaires available however the choice depends on the budget as well as the degree of significance placed upon the user’s perceived level of satisfaction as a factor of the overall project [55].

3.1.4. Usability Metric For Cognitive Load

Cognitive load has been identified as the measure of mental activity on working memory at any particular instance [56]. To determine the app's cognitive load, we will use the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) test. NASA-TLX allows the user to evaluate the situation of the workload after the testing is done. It measures the overall task demands by identifying 3 broad scales, which are task, behaviour and subject-related. Each of the scales has factors. The task-related scale includes mental, physical and temporal demands. The behaviour-related scale includes performance and effort. Subject related includes frustration. A user will need to have description for each of the factors as demonstrated below [57]:

- Mental demand: To what extent did you need to perform mental and perceptual activities (such as thinking and calculating)?
- Physical demand: To what extent did you need to perform physical activities (such as pushing and pulling)
- Temporal demand: To what extent did you feel a time pressure while performing tasks?
- Effort: How hard did you have to work hard (mentally and physically) to perform tasks?
- Performance: How satisfied are you with your performance?
- Frustration level: How stressed or annoyed did you feel while performing these tasks?

The NASA-TLX test contains two stages which are weights and ratings. In the weighting procedure, a user will be required to evaluate the influence of each factor regarding a task. There are 15 potential pairs of factors about which a comparison is made. A user will be giving 15 cards and each card contains a pair of the factors and asked to select the most relevant factor regarding the task. Each time the user selects from a pair, the examiner counts it. The scale for a factor for each user can range from 0 to 15. The total comparisons for all factors should equal 15. In the second stage, a user needs to rate each of the factors above in a scale that is divided into 20 equal intervals and each interval equals 5 points with a total of a 100 on the scale. As it is a post-event test, it is an effective one as it captures the thoughts and interaction of the user.

3.1.5. Usability Metric For Learnability

Learnability is the ability of the interface to help the user accomplish tasks on the first attempt [58]. Learnability can therefore be measured through establishing the task performance of users who have not been exposed to that app before. Another way of looking at usability is through perceiving how usability or task performance has improved after repeated trials.

3.1.6. Usability Metric For Errors

Another usability measurement is measuring the amount of errors made by the user when completing a task. Errors are defined as mistakes that are made by the participant when attempting a task. Counting the errors provides excellent diagnostic information and it should be mapped into usability problems [59].

3.1.7. Usability Metric For Memorability

Memorability measures how easy it is to remember how to perform a task on the app after the casual user returns to the app after a certain period of not using it [60]. Memorability has the same tests of efficiency and effectiveness but these are repeated after some period of time in order to determine whether the user has remembered how to perform the same task; and hence whether this has improved the usability.

3.2. Usability Testing Environment

The tests were conducted in a typical usability test environment. Laboratory settings were controlled in order to ensure that there were no external interruptions such as varying lighting conditions or disturbing noises. Test sessions were completed via Apple's wireless AirPlay technology. A MacBook was used for recording. The first step was to install Reflector, which is a wireless streaming and mirroring receiver that converted a laptop into an AirPlay receiver. This allowed the user to mirror their smartphone's screen onto their laptop. It also eliminated the need to have an external camera to record events. Moreover, it also helped to minimize the distraction for the user. The purpose of using this software and technology was to create the friendly and quiet environment that is essential for usability testing [61-62].

Twenty-six participants tested Aded Surat app. While they tested it, their mobile screens were recorded through Reflector software. All participants were asked to use the app 3 times. Each time, all participants were asked to perform 14 tasks, which were the same for all users. The time difference between the first and second sessions was one hour. Between the second and the third sessions, there was a one-week interval.

The Armed Force Hospital in the Taif Region of Saudi Arabia provided candidates who suffered from obesity and were motivated to lose weight in order to have a healthier life style. The usability test was divided into five phases:

- Introduction: In the first phase, both participants and the examiner introduced themselves. The purpose of the introduction phase was to establish a comfortable interaction between the examiner and participants.
- Warm-up: In this phase, participants were asked to download the app "Aded Surat" and to fill out a brief questionnaire that aimed to collect participants' information such as gender and age.
- Deep focus: During this phase, the examiner gave the users a list of the 14 tasks. The participants used the app with the focus being on what it was doing; how it worked and how the app could be used. The examiner encouraged the participants to think aloud while they were performing the tasks. Moreover, when participants finished a task, they were asked to rate it in an SEQ questionnaire.
- Retrospective: In the penultimate phase, the examiner explained the NASA-TLX questionnaire and asked participants to fill it out.
- Wrap up: In the final phase, the examiner thanked the participants and answered any enquiries.

4. RESULTS

Researchers examined all the videos that were recorded on mobile screens while the participants were performing in the trial. All users who successfully completed a task scored 1 and at the same time we measured how long it took to complete a task. In contrast, users who completed a task in the wrong way or gave up on a task received 0 and the time taken was measured as well. Then the equations for effectiveness, overall relative efficiency and time-based efficiency were applied. Then all errors that participants had made while performing tasks were calculated. Regarding the learnability attribute, we compared participants' performances in the first session with those of the second. Memorability was then measured by comparing participants' performances in the second session with those of the third. Both satisfaction and cognitive loads were applied only in the first session as they measured the performances of participants who had not previously been exposed to an app. If these loads had been applied in the second and third sessions, this condition could not have been met. Next, we examined the data from the SEQ questionnaire that was used

to measure satisfaction. The rating for each user was calculated and then divided by 14 to determine the average satisfaction value for each user. We then examined the data from the NASA-TLX questionnaire and applied the roles to determine the total user score for the cognitive load [57].

Table 2. Participants' information.

Users	Gender	Age group	Occupation	Type of phone
1	Male	35 to 44	Self employed	iPhone 7
2	Male	25 to 34	Teacher at high school	iPhone 7
3	Female	25 to 34	Unemployed	OnePlus 3
4	Female	45 to 54	Government employee	iPhone 6 S
5	Female	25 to 34	Government employee	HTC 10
6	Female	Prefers not to say	Prefers not to say	iPhone 7
7	Female	25 to 34	Accountant in a company	iPhone 7 Plus
8	Female	25 to 34	Receptionist at a hospital	iPhone 6 S
9	Female	Prefers not to say	Prefers not to say	iPhone 7
10	Female	Prefers not to say	Prefers not to say	iPhone 7
11	Male	18 to 24	Student at university	iPhone 6
12	Female	18 to 24	Unemployed	iPhone 7 Plus
13	Female	45 to 54	Unemployed	iPhone 6 S Plus
14	Male	25 to 34	Teacher at university	Galaxy S 7
15	Male	Prefers not to say	Prefers not to say	iPhone 7
16	Male	18 to 24	Student at university	OnePlus 3
17	Male	25 to 34	Unemployed	iPhone 6 S
18	Male	45 to 54	Self employed	iPhone 7
19	Male	55 to 64	Retired	iPhone 7 Plus
20	Male	25 to 34	Employee in a company	iPhone 7
21	Male	25 to 34	Family business	iPhone 7
22	Female	18 to 24	Student at university	iPhone 6 S Plus
23	Male	25 to 34	Government employee	iPhone 7
24	Male	35 to 44	Employee in a company	iPhone 6 S
25	Female	45 to 54	Unemployed	iPhone 7 Plus
26	Female	Prefers not to say	Prefers not to say	iPhone 7

Twenty-six participants, thirteen females and thirteen males, were part of the usability of Aded Surat app. Their information is presented in Table 2.

4.1. Effectiveness

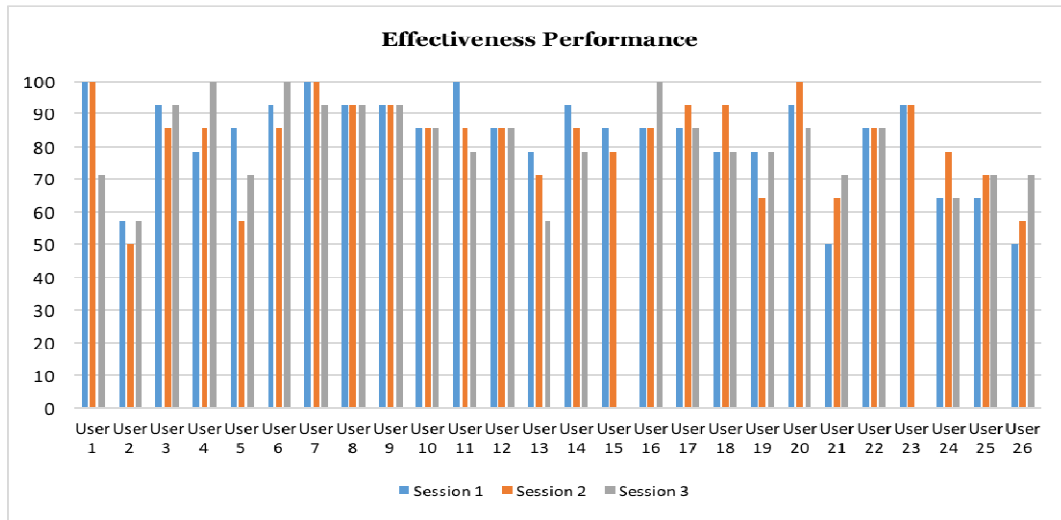


Figure 7. Effectiveness performance

Figure 7 describes each user's effectiveness performance over the course of the three sessions. Out of the 26 participants, only six participants, who are user 4, user 6, user 16, user 21, user 25 and user 26, showed positive progress across sessions and yet they differed in their pattern. While user 4, user 21 and user 26 had the pattern of keep improving over the three sessions, user 6, user 16 and user 25 showed a pattern of overall improvement across the three sessions. Whilst on one hand, user 4, user 21 and user 26 had the value of 78.57%, 50% and 50% respectively in session 1, the value then increased to 85.71%, 64.28% and 57.14% respectively in session 2 and then did so again to 100%, 71.42% and 71.42% respectively in session 3. On the other hand, user 6 had the value of 92.85% in session 1, but this decreased to 85.71% in session 2 and then increased to 100% in session 3. User 16 had the same value in both session 1 and session 2 at 85.71% but this then increased to 100% in session 3. User 25 had the value of 64.28% in session 1 and then increased to 71.42% in both session 2 and session 3.

However, five other participants, who are user 8, user 9, user 10, user 12 and user 22, had the same value over the three sessions. While user 8 and user 9 had the value of 92.85%, user 10, user 12 and user 22 had the value of 85.71%.

Six participants, identified as user 2, user 3, user 17, user 18, user 19 and user 24, had the same value in session 1 and 3. While user 2's, user 3's and user 19's value decreased in session 2 compared to their value in session 1 and session 3, user 17's, user 18's and user 24's value increased in session 2 compared to their value in session 1 and session 3.

In contrast, nine participants, who are user 1, user 5, user 7, user 11, user 13, user 14, user 15, user 20 and user 23, showed negative progress across sessions. Their effectiveness decreased over the three sessions. User 7 had the highest percentage of value each time. In session 1, it is 100% and remained constant in session 2 and yet it decreased in session 3 to 92.85%. Moreover, User 1 showed the same pattern though the value was lower at 71.42% in session 3.

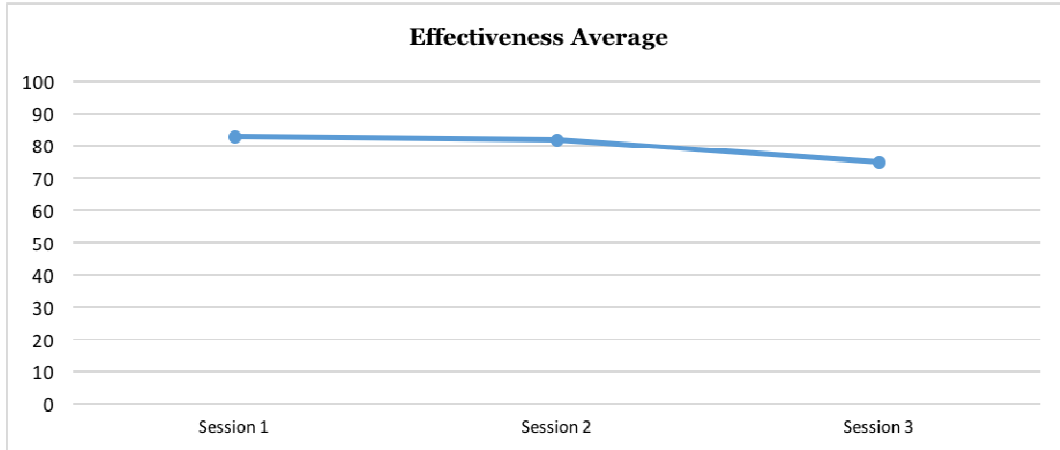


Figure 8. Effectiveness average

Figure 8 shows the effectiveness performance average, which decreased over each session. In session 1 it was 82.86%, then it fell to 81.86% and finally in session 3, it reached to 74.99%.

4.2. Efficiency

4.2.1. Overall Relative Efficiency

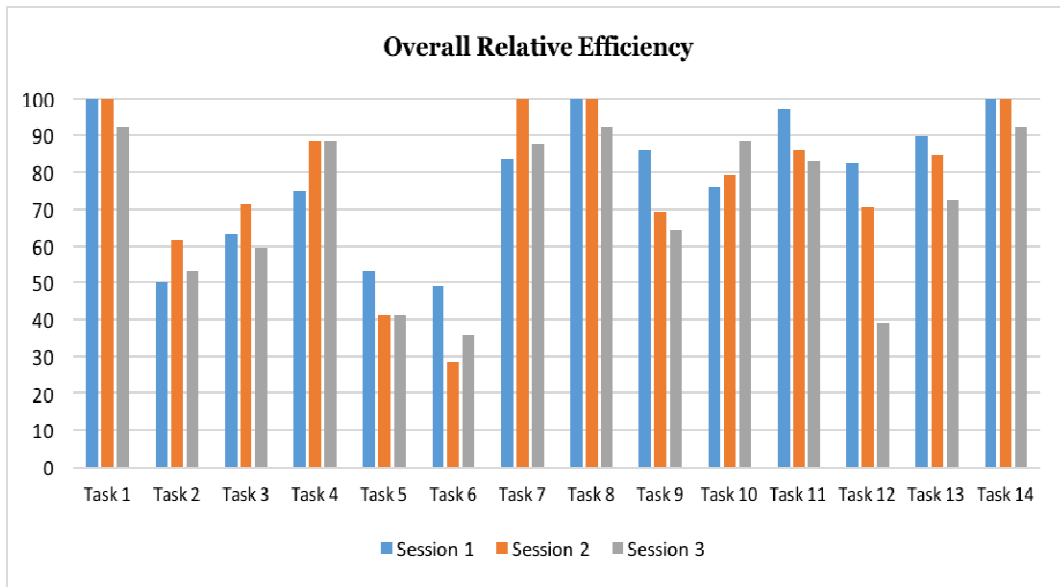


Figure 9. Overall relative efficiency

Figure 9 demonstrates overall relative efficiency for tasks for the three sessions. In session 1, among the 14 tasks, 3 tasks scored 100% whereas in session 2 and session 3 it was 4 and 0 respectively. Only in tasks 2, 4, 7 and 10 did the overall relative efficiency percentage improved over the three sessions. However, all the other tasks dramatically decreased between the first and final sessions.

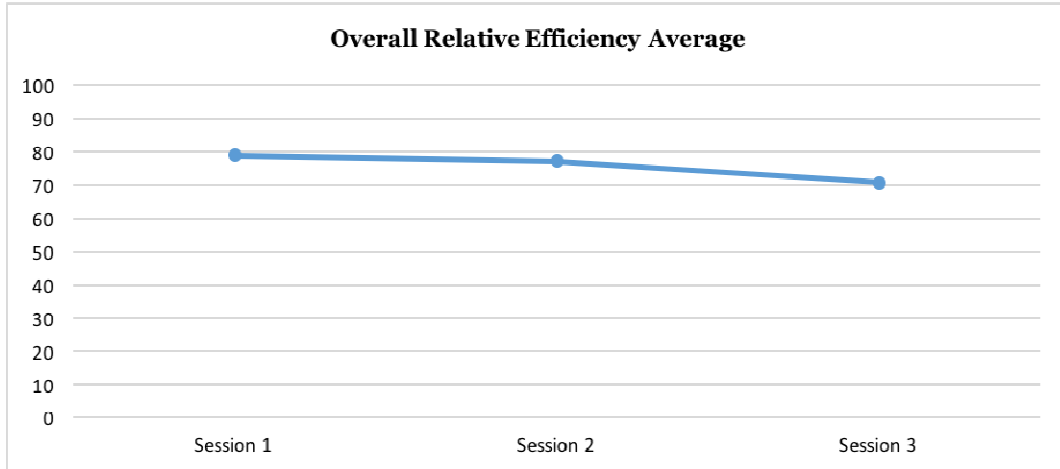


Figure 10. Overall relative efficiency average

Figure 10 shows the overall relative efficiency average, which decreased over each of the three sessions. In session 1 it was 78.84%, then it fell to 77.16% and in the final session it reached 70.67%.

4.2.2. Time-Based Efficiency

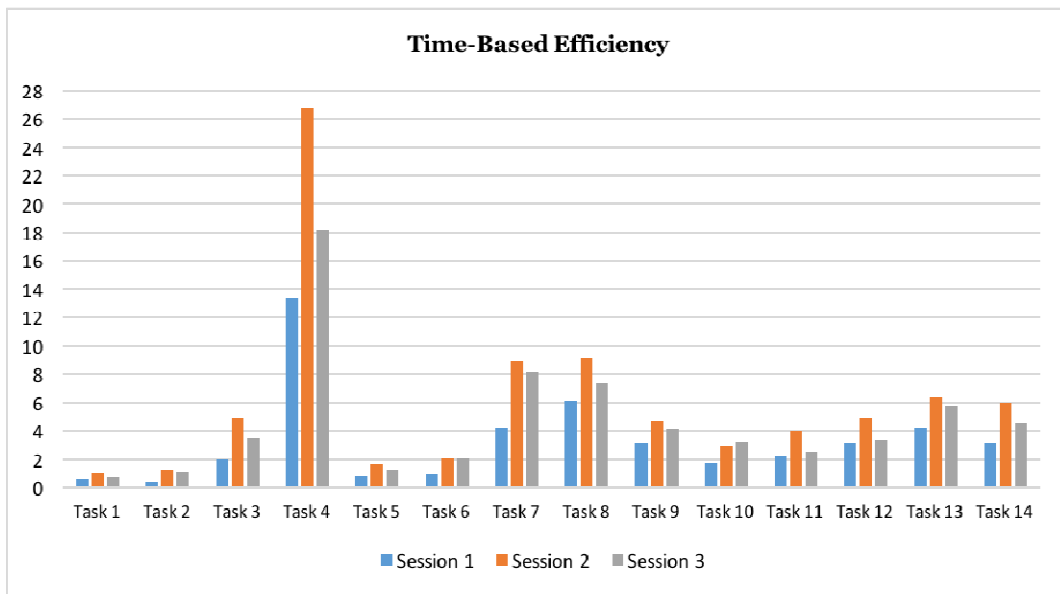


Figure 11. Time-based efficiency

Figure 11 states time-based efficiency for tasks among the sessions. Task 4 had the highest time-based efficiency score among tasks. In sessions 1, 2 and 3 it was 13.35 goals/sec, 26.8 goals/sec, which is the highest value for a task in all sessions, and 18.12 goals/sec respectively. Task 8 had the second highest time-based efficiency score followed by task 7 and task 13. On the other hand, task 1 got the lowest time-based efficiency followed by task 2 and task 5.

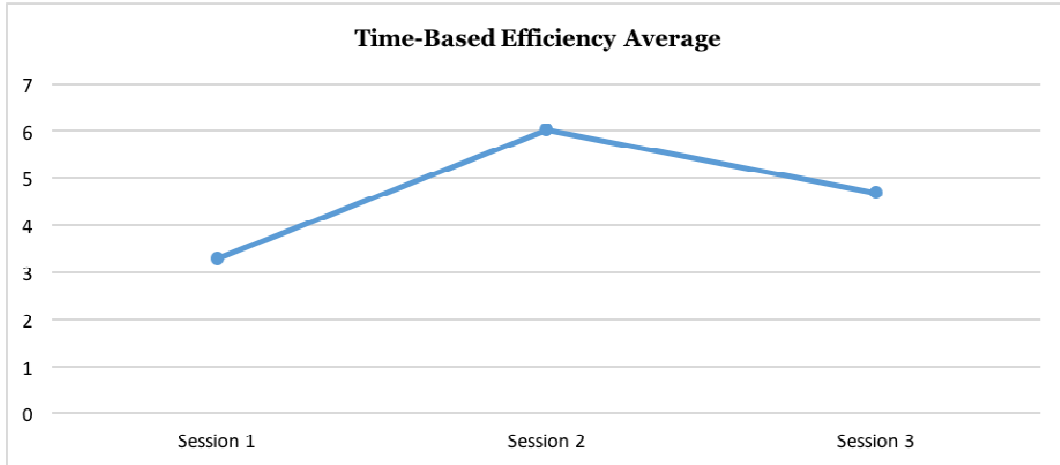


Figure 12. Time-based efficiency average

Figure 12 shows the time-based efficiency average, which fluctuated across sessions. In session 1 it was 3.3 goals/sec, then it increased to 6.03 goals/sec and finally in the session 3 it decreased to 4.7 goals/sec.

4.3. Satisfaction

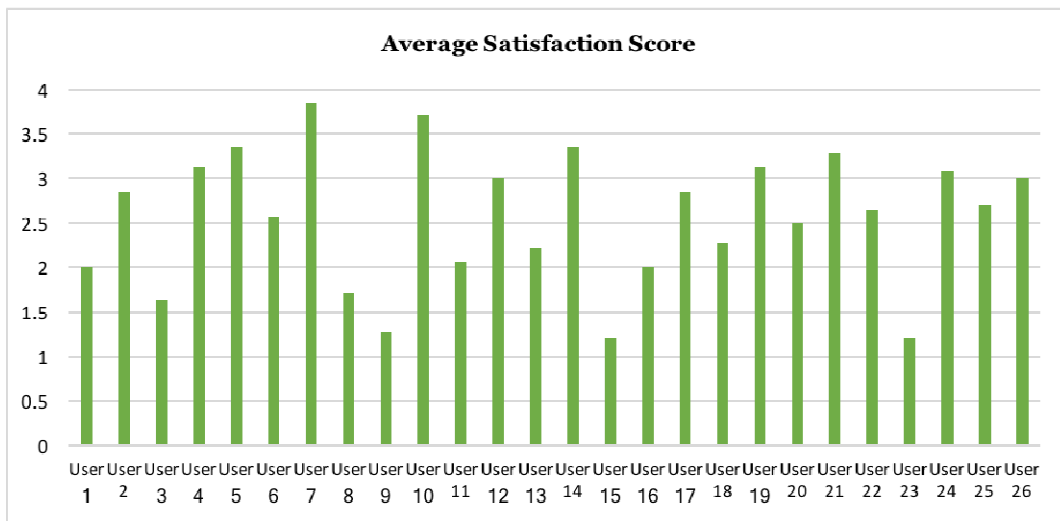


Figure 13. Average satisfaction score

Figure 13 shows each user’s average satisfaction score for all tasks. User 7 had the highest score at 3.85, followed by user 10 at 3.71 and then both user 5 and user 14 scored 3.35. However, User 23, user 9 and user 3 had the lowest score at 1.21, 1.28 and 1.64 respectively.

4.4. Cognitive Load

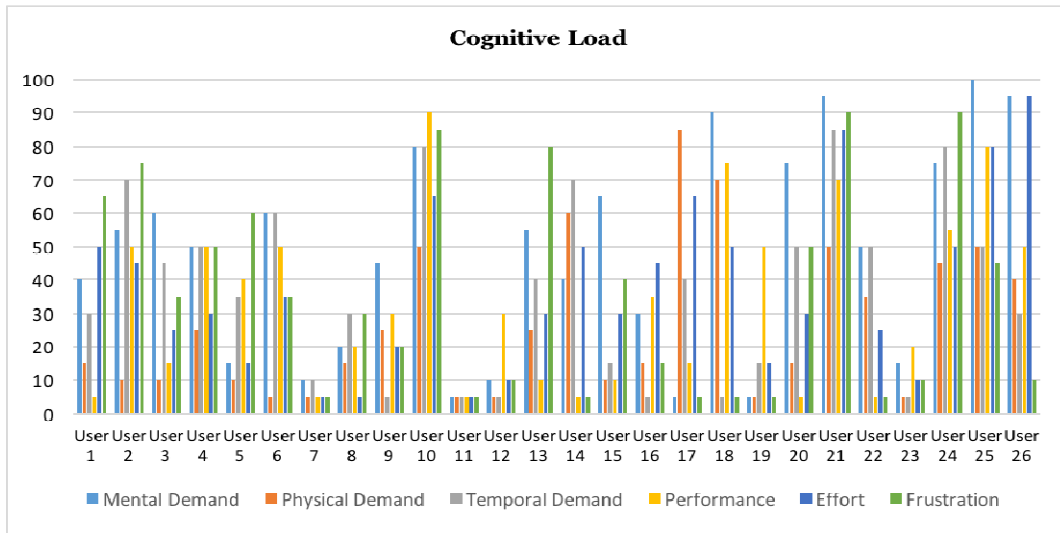


Figure 14. Users' rating for each subscale in cognitive load

Figure 14 shows each user's rating for each subscale in the cognitive load. User 11's cognitive loading is the most consistent. Scores are 5% for all the six subscales. User 7's cognitive loading is the second most consistent. Scores lie between physical demand, performance, effort and frustration at (5%) to mental and temporal demand (10%). However, between user 13, user 17 and user 18 the score gap is too high. Mental demand and effort scored the highest value amongst all the subscales. On the other hand, physical demand and performance scored the lowest values.

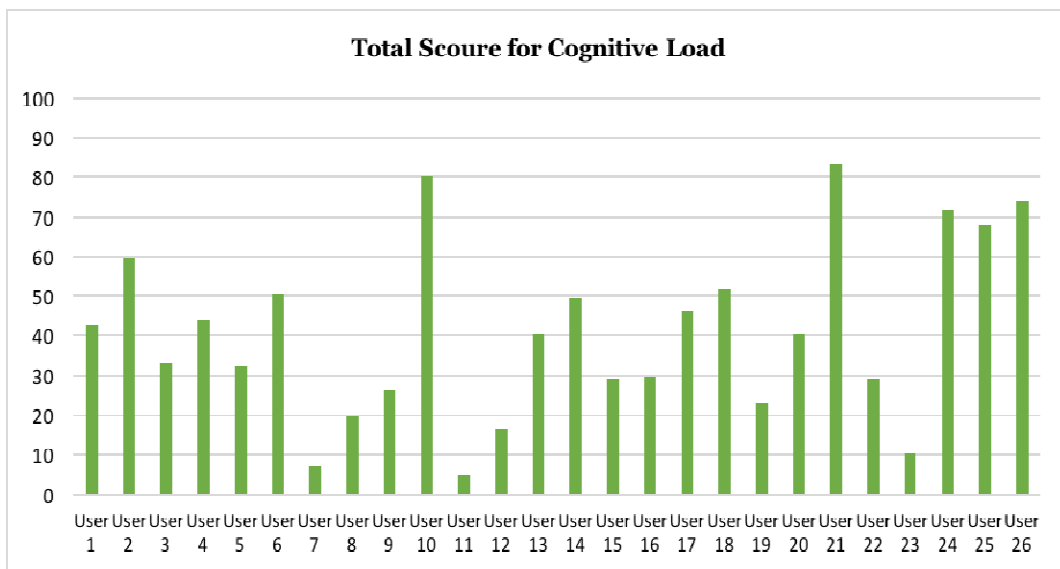


Figure 15. Total score for cognitive load

Figure 15 refers to the total score for cognitive load amongst users. User 21 had the highest value at 83.3%. User 10 and user 26 scored 80% and 74% respectively. However, user 11 had the lowest score at 5%, followed by user 7 and user 23 at 7% and 10.3% respectively.

4.5. Errors

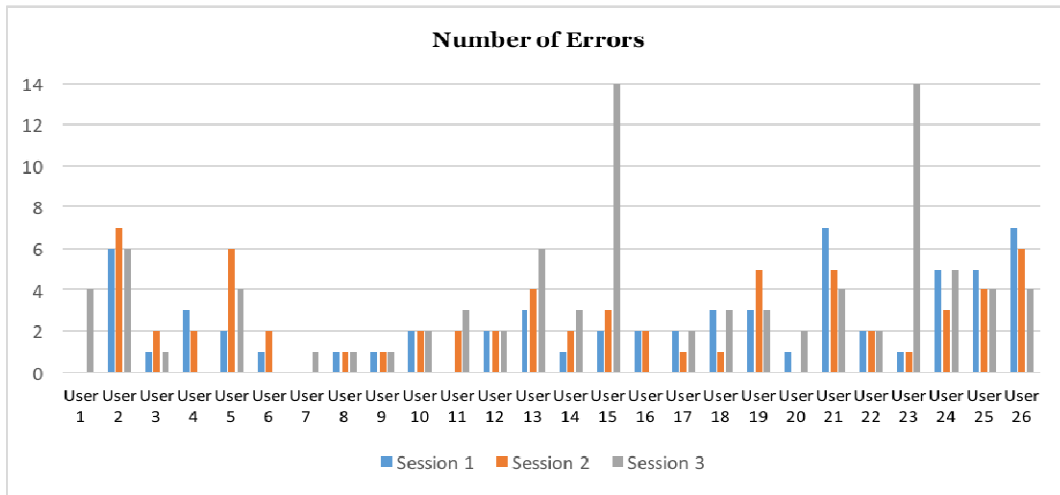


Figure 16. Number of errors

Figure 16 shows the number of errors made by each user. User 7 had the lowest number of errors with just one in session 3. User 6, user 8, user 9 and user 20 had the second lowest number of errors with three errors over the three sessions. User 1, user 3 and user 16 had the third lowest number of errors with four errors over the three sessions. However, User 2 and user 15 had the highest number of errors at 19 over the three sessions.

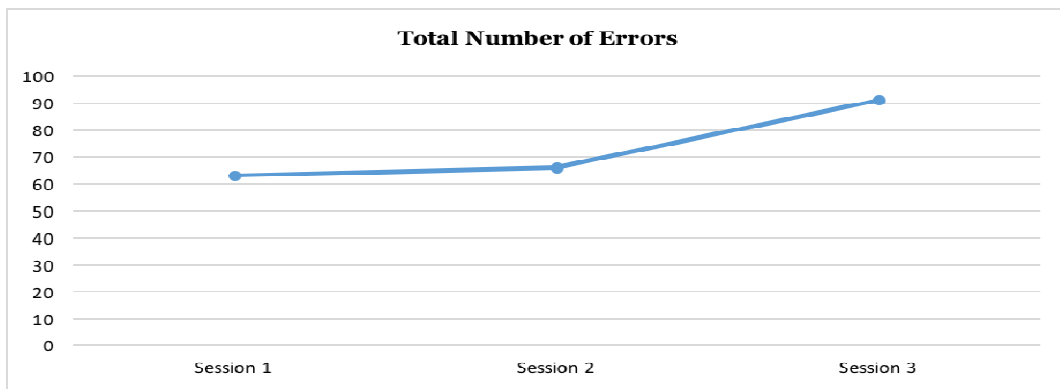


Figure 17. Total number of errors

Figure 17 demonstrates the total number of errors made by all users, which increased over each session. In session 1 it was 63, then it slightly increased to 66 and finally in the third session, it sharply increased to 91.

4.6. Discussion On The Results

One limitation of this study was that users 15 and 23 were not able to participate in the third session as they faced a technical issue with the app. The app did not respond to them when they started performing the first task and after several attempts, they gave up. However, the overall trial for testing the usability of the app succeeded as the level of usability was determined.

Despite the positive increase in the overall score for time-based efficiency between session 1 and session 3, the percentage score for user's effectiveness and overall relative efficiency decreased over time. Moreover, the number of errors increased from the first session to the second session and did so again from the second to third sessions. As a result of this, the app had a negative association with both learnability and memorability attributes. Furthermore, several participants scored a high percentage in the satisfaction questionnaire, which is negative as a high score means it was more difficult and out of the 26 participants, no one rated the whole task as very easy and scored 1 as an average. Besides this, overall cognitive load score was high as the lowest percentage scored by a participant was 14.33%, which means that several participants were not able to perform tasks correctly while doing other activities; for example, speaking to examiners.

The five usability attributes (effectiveness, efficiency, learnability, memorability and errors) did not improve over time. Moreover, both satisfaction and cognitive load scored high percentages because the majority of participants found the app difficult to use. Therefore, the results state that Aded Surat app has a low level of usability, which is expected due to the fact that it was designed and developed without considering usability attributes and factors.

It is recommended that conducting a qualitative study to determine the reasons and factors that negatively affect the level of usability of the Aded Surat app. The qualitative study will also consider the importance of social and cultural norms and how they can be applied to improve the usability of the app. A specific usability guideline for fitness mobile apps will then be created, which will help to develop a fitness app that is specially designed for obese individuals in Saudi Arabia.

5. CONCLUSION

The primary purpose of this paper has been to examine the usability for an Arabic fitness mobile app "Aded Surat". This paper has highlighted the attributes that are considered to be a crucial for improving the usability of fitness mobile apps through presenting an extensive literature review. The paper has presented the methodology and the procedures for testing the Aded Surat app. Seven usability attributes, (effectiveness, efficiency, satisfaction, memorability, errors, learnability and cognitive load) were tested. The trial for the test was done in collaboration with the Armed Forces Hospitals - Taif Region in Saudi Arabia, which provides the candidates. The result from this trial was that Aded Surat app failed to meet with the usability attributes and consequently participants found it difficult to use. Future work will include performing a qualitative study for the app to determine how to improve the level of usability and then create usability guidelines for fitness mobile apps. Based on these guidelines, an app that is specifically designed for obese individuals in Saudi Arabia will be developed. Obesity is a major issue for health departments all over the world. Saudi Arabia is a country where the obesity has reached an alarming rate of 35.5% of the population. Better app usability would help keep these individuals motivated to make necessary lifestyle changes.

REFERENCES

- [1] A. P. SIMOPOULOS and T. B. VAN ITALLIE, "Body weight, health, and longevity," *Annals of internal medicine*, vol. 100, pp. 285-295, 1984.
- [2] W. H. Organization. (2016, 2 October). Obesity and overweight. Available: <http://www.who.int/mediacentre/factsheets/fs311/en/>
- [3] O. R. Center. (2016, 10 October). Obesity in Saudi Arabia. Available: <https://www.obesitycenter.edu.sa/pages/patients.aspx?id=258>
- [4] A. Afshin, M. H. Forouzanfar, M. B. Reitsma, P. Sur, K. Estep, A. Lee, et al., "Health Effects of Overweight and Obesity in 195 Countries over 25 Years," *The New England journal of medicine*, vol. 377, pp. 13-27, 2017.
- [5] K. Singer and C. N. Lumeng, "The initiation of metabolic inflammation in childhood obesity," *The Journal of clinical investigation*, vol. 127, pp. 65-73, 2017.
- [6] K. R. Fontaine, D. T. Redden, C. Wang, A. O. Westfall, and D. B. Allison, "Years of life lost due to obesity," *Jama*, vol. 289, pp. 187-193, 2003.
- [7] J. Stevens, J. Cai, E. R. Pamuk, D. F. Williamson, M. J. Thun, and J. L. Wood, "The effect of age on the association between body-mass index and mortality," *New England Journal of Medicine*, vol. 338, pp. 1-7, 1998.
- [8] E. E. Calle, M. J. Thun, J. M. Petrelli, C. Rodriguez, and C. W. Heath Jr, "Body-mass index and mortality in a prospective cohort of US adults," *New England Journal of Medicine*, vol. 341, pp. 1097-1105, 1999.
- [9] C. Summerbell, E. Waters, L. Edmunds, S. Kelly, T. Brown, and K. Campbell, "Interventions for preventing obesity in children (Review)," *Cochrane library*, vol. 3, pp. 1-71, 2005.
- [10] W. Saris, S. Blair, M. Van Baak, S. Eaton, P. Davies, L. Di Pietro, et al., "How much physical activity is enough to prevent unhealthy weight gain? Outcome of the IASO 1st Stock Conference and consensus statement," *Obesity reviews*, vol. 4, pp. 101-114, 2003.
- [11] O. Bar-Or, "Juvenile obesity, physical activity, and lifestyle changes: Cornerstones for prevention and management," *The physician and sportsmedicine*, vol. 28, pp. 51-58, 2000.
- [12] J. L. Anderson, E. M. Antman, S. R. Bailey, E. R. Bates, J. C. Blankenship, D. E. Casey Jr, et al., "AHA Scientific Statement," *Circulation*, vol. 120, pp. 2271-2306, 2009.
- [13] J. O. Hill and H. R. Wyatt, "Role of physical activity in preventing and treating obesity," *Journal of Applied Physiology*, vol. 99, pp. 765-770, 2005.
- [14] J. O. Hill and J. C. Peters, "Environmental contributions to the obesity epidemic," *Science*, vol. 280, pp. 1371-1374, 1998.
- [15] I. Contento, G. I. Balch, Y. L. Bronner, L. Lytle, S. Maloney, C. Olson, et al., "The effectiveness of nutrition education and implications for nutrition education policy, programs, and research: a review of research," *Journal of nutrition education (USA)*, 1995.
- [16] G. D. Foster, A. P. Makris, and B. A. Bailer, "Behavioral treatment of obesity," *The American journal of clinical nutrition*, vol. 82, pp. 230S-235S, 2005.
- [17] T. A. Wadden and A. J. Stunkard, *Handbook of obesity treatment*: Guilford Press, 2002.
- [18] K. D. Brownell, *LEARN program for weight management 2000*: American Health, 2000.
- [19] J. Yang, "Toward physical activity diary: motion recognition using simple acceleration features with mobile phones," in *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics*, 2009, pp. 1-10.
- [20] T. Denning, A. Andrew, R. Chaudhri, C. Hartung, J. Lester, G. Borriello, et al., "BALANCE: towards a usable pervasive wellness application with accurate activity inference," in *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*, 2009, p. 5.
- [21] S. M. Arteaga, M. Kudeki, A. Woodworth, and S. Kurniawan, "Mobile system to motivate teenagers' physical activity," in *Proceedings of the 9th International Conference on Interaction Design and Children*, 2010, pp. 1-10.
- [22] D. E. Conroy, C.-H. Yang, and J. P. Maher, "Behavior change techniques in top-ranked mobile apps for physical activity," *American journal of preventive medicine*, vol. 46, pp. 649-652, 2014.
- [23] F. Raben and E. Snip, "The MENAP region is developing, but can it keep its promise?," *Research World*, vol. 2014, pp. 6-11, 2014.
- [24] Statista. (2017, 8 March). Number of smartphone users in Saudi Arabia from 2014 to 2021 (in millions)*. Available: <https://www.statista.com/statistics/494616/smartphone-users-in-saudi-arabia/>
- [25] D. LLP. (2012, 2 February). So Many Apps -- So Little To Download. Available: <http://www.mondaq.com/x/192692/IT+internet/So+Many+Apps+So+Little+To+Dow%20load>

- [26] S. Dredge, "Most branded apps are a flop says Deloitte. But why," ed, 2011.
- [27] M. Bhuiyan, A. Zaman, and M. H. Miraz, "Usability Evaluation of a Mobile Application in Extraordinary Environment for Extraordinary People," arXiv preprint arXiv:1708.04653, 2017.
- [28] R. Youens. (2011, 2 February). 7 Habits of Highly Effective Apps. Available: <https://gigaom.com/2011/07/16/7-habits-of-highly-effective-apps/>
- [29] I. Nascimento, W. Silva, A. Lopes, L. Rivero, B. Gadelha, E. Oliveira, et al., "An Empirical Study to Evaluate the Feasibility of a UX and Usability Inspection Technique for Mobile Applications," in 28th International Conference on Software Engineering & Knowledge Engineering, California, USA, 2016.
- [30] H. Hoehle, R. Aljafari, and V. Venkatesh, "Leveraging Microsoft's mobile usability guidelines: Conceptualizing and developing scales for mobile application usability," *International Journal of Human-Computer Studies*, vol. 89, pp. 35-53, 2016.
- [31] S. Pagoto, K. Schneider, M. Jovic, M. DeBiaise, and D. Mann, "Evidence-based strategies in weight-loss mobile apps," *American journal of preventive medicine*, vol. 45, pp. 576-582, 2013.
- [32] A. C. King, E. B. Hekler, L. A. Grieco, S. J. Winter, J. L. Sheats, M. P. Buman, et al., "Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults," *PloS one*, vol. 8, p. e62613, 2013.
- [33] A. A. Alnasser, R. E. Amalraj, A. Sathiaseelan, A. S. Al-Khalifa, and D. Marais, "Do Arabic weight-loss apps adhere to evidence-informed practices?," *Translational behavioral medicine*, vol. 6, pp. 396-402, 2016.
- [34] E. R. Breton, B. F. Fuemmeler, and L. C. Abrams, "Weight loss—there is an app for that! But does it adhere to evidence-informed practices?," *Translational behavioral medicine*, vol. 1, pp. 523-529, 2011.
- [35] A. M. Content. (2016, 5 January). Adad Alsrat from Hmiate Available: <https://itunes.apple.com/us/app/dad-als-rat-alhraryt-w-mraqbt/id908026885?mt=8>
- [36] Hmiate. (2016, 5 January). Adad Alsrat from Hmiate. Available: <http://m-diet.com/Home/Tracker>
- [37] G. Play. (2016, 5 January). Adad Alsrat from Hmiate Available: <https://play.google.com/store/apps/details?id=com.mdiet.tracker&hl=ar>
- [38] R. Baharuddin, D. Singh, and R. Razali, "Usability dimensions for mobile applications—A review," *Res. J. Appl. Sci. Eng. Technol.*, vol. 5, pp. 2225-2231, 2013.
- [39] W. ISO, "9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)," *The international organization for standardization*, vol. 45, 1998.
- [40] S. Ben and C. Plaisant, "Designing the user interface 4 th edition," ed: Pearson Addison Wesley, USA, 2005.
- [41] E. Folmer and J. Bosch, "Architecting for usability: a survey," *Journal of systems and software*, vol. 70, pp. 61-78, 2004.
- [42] D. Saffer, "Designing for Interaction: Creating Smart Applications and Clever Devices," New Riders Press, < <http://www.designingforinteraction.com>, vol. 2, p. 2.1, 2007.
- [43] W. Albert and T. Tullis, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*: Newnes, 2013.
- [44] K. D. Eason, "Towards the experimental study of usability," *Behaviour & Information Technology*, vol. 3, pp. 133-143, 1984.
- [45] A. Dix, S. Cairncross, G. Cockton, R. Beale, R. St Amant, and M. Hause, "Human-Computer Interaction," 1993.
- [46] J. Nielsen, *Usability engineering*: Elsevier, 1994.
- [47] A. Abran, A. Khelifi, W. Suryan, and A. Seffah, "Usability meanings and interpretations in ISO standards," *Software Quality Journal*, vol. 11, pp. 325-338, 2003.
- [48] D. Norman, "The Design of Everyday Things Basic Books New York," 2002.
- [49] S. Ben and C. Plaisant, "Designing the user interface 4 th edition," ed: Pearson Addison Wesley, USA, 2005.
- [50] C. Yeh, "The principles of interaction design in the post-digital age," Taipei: ARTIST-MAGAZINE, 2010.
- [51] C.-M. Wang and C.-H. Huang, "A study of usability principles and interface design for mobile e-books," *Ergonomics*, vol. 58, pp. 1253-1265, 2015.
- [52] R. Alturki and V. Gay, "USABILITY TESTING OF FITNESS MOBILE APPLICATION: METHODOLOGY AND QUANTITATIVE RESULTS," presented at the 7th International Conference on Computer Science, Engineering & Applications (ICCSEA 2017), Copenhagen, Denmark, 2017.

- [53] J. Mifsud. (2015, 3 November). Usability Metrics – A Guide To Quantify The Usability Of Any System. Available: <http://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/>
- [54] J. Sauro. (2010, 9 December). IF YOU COULD ONLY ASK ONE QUESTION, USE THIS ONE. Available: <https://measuringu.com/single-question/>
- [55] A. Garcia. (2013, 18 October). UX Research | Standardized Usability Questionnaire. Available: <https://chaione.com/blog/ux-research-standardizing-usability-questionnaires/>
- [56] J. P. Tracy and M. J. Albers, "Measuring cognitive load to test the usability of web sites," in Annual Conference-society for technical communication, 2006, p. 256.
- [57] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139-183, 1988.
- [58] J. Sauro. (2013, 1 December). HOW TO MEASURE LEARNABILITY. Available: <https://measuringu.com/measure-learnability/>
- [59] J. Sauro. (2011, 4 December). 10 ESSENTIAL USABILITY METRICS. Available: <https://measuringu.com/essential-metrics/>
- [60] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," *Journal of Interaction Science*, vol. 1, pp. 1-16, 2013.
- [61] C. Walsh. (2015, 6 November). A Guide To Simple And Painless Mobile User Testing. Available: <https://www.smashingmagazine.com/2015/12/simple-and-painless-mobile-user-testing/>
- [62] J. Mifsud. (2016, 3 November). Usability Testing Of Mobile Applications: A Step-By-Step Guide. Available: <http://usabilitygeek.com/usability-testing-mobile-applications/>