IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 6, 2022, accepted April 4, 2022, date of publication April 12, 2022, date of current version April 22, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3166893

Mobile Application Usability Evaluation: A Study Based on Demography

WAHAJ ALI[1], OMER RIAZ[1], SHAHZAD MUMTAZ[1], AMJAD REHMAN KHAN[2], (Senior Member, IEEE), TANZILA SABA[2], (Senior Member, IEEE), AND SAEED ALI BAHAJ[3]

[1]Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[2]Artificial Intelligence and Data Analytics Laboratory, CCIS, Prince Sultan University, Riyadh 11586, Saudi Arabia
[3]MIS Department, College of Business Administration, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

Corresponding author: Omer Riaz (omer.riaz@iub.edu.pk)

This work was supported by Prince Sultan University, Riyadh KSA [by paying the article processing charges (APC) of this publication].

ABSTRACT An application or product is considered ''usable'' if it is pleasing, easy to use, and works as expected user interface. Most companies majorly focus on the application's functional requirements but put minimal effort into user experience (usability). Consumers' adaptation of these applications depends on the number of features and user interface. In this work, an android based application, ''Houzcalls'' is used for usability study using PACMAD usability model. This work is focused on the variations in PACMAD attributes based on the participants' education and age. Participants are segregated into two major groups FG1 and FG2 based on their education. All the participants with more than 10 years of education are in FG1 while others in FG2. Each focal group is divided into four subgroups Under 25, 25-35, 36-45, and Over 45 based on their ages. The results have shown participants in FG1 have shown more Effectiveness, Efficiency, Satisfaction, Learnability, and Memorability. In contrast, they have committed fewer Errors and shown less Cognitive Load during usability testing as compared to FG2. These variations can also be seen age-wise as generally ''Under 25 and 25-35'' subgroups have shown better results than other subgroups. It is inferred from the study that application usability and acceptability can be increased by considering the general population during development which includes all groups of people based on education and age.

INDEX TERMS HCI, usability testing, attribute-based evaluation, PACMAD usability model, UI, test design, test execution, focus groups, age, education, mobile application, requirement engineering.

I. INTRODUCTION

Over the last decade, there has been a tremendous growth of software applications for mobile devices. Smartphones have exponentially gained computational frequency and storage efficiency while being portable and low-cost devices [1]. As a result, a growing large community of wide range of mobile users has been established due to increasing affordability. The estimated number for users of smartphone devices by 2020 was 5.65 billion [2] and an annual financial market share of around 15 billion USD per year over the last few years [3]. The focus of mobile application development has attracted to develop applications that are useful for personal use and cover other areas like business management, health care management, education service delivery, electronic media, etc. This diverse application domain requires more attention from developers to develop easy-to-use and reliable. Developing an easy to use application further requires understanding the target users in terms of their requirements, objectives, and intellect. If these points are considered thoroughly then the usability (in terms of effectiveness, efficiency, satisfaction, etc.) of these applications will result in higher acceptance from users and attract more users.

It is a well-known fact that when searching for an application with a specific functionality using a search engine,

The associate editor coordinating the review of this manuscript and approving it for publication was Giuseppe Desolda.

the applications are usually ranked in order considering the number of downloads and positive users' reviews. Nowadays due to busy lifestyles and short attention span application users don't properly vet applications (i.e. they get agitated by poorly designed and implemented software). User's search for applications according to their need while selection is mostly based on other users' feedback. One of the main adaptation criteria after functionality is the "Look and Feel" of the application which in terms called Usability [4].

Software usability is a component of software development that guarantees the applications are usable with the target users. Usability testing is a practice that attempts to ensure that an application can be used effectively. Usability testing is a practice that attempts to ensure that accomplish one or more defined tasks. In other words, it refers to evaluating the ease with which users can learn to use a product [4]. Usability evaluation practices are commonly classified into categories as follows:

A. CHECKLIST BASED EVALUATION [5]

This method is based on a predefined checklist [6] which then passes on to multiple experts, who provide their opinion about the product by checking the correct boxes according to themselves on the checklist. It is also called Heuristics based evaluation [7].

B. ATTRIBUTE BASED EVALUATION [4]

In this approach, usability is divided into Attributes and evaluated using experimental methods called "Models." Overall usability is calculated by contribution of each attribute. This is also called User testing [8]. User testing is further divided into Scenario based (i.e. in person or remote) testing, subjective questionnaires based and moderator based evaluations etc.

The adaptability of usability in computer applications has given birth to a new field of Computer Science (CS) termed Human-computer Interaction (HCI) [9]. Although usability engineering is a part of the software development process, it comes just after the development/implementation phase in the software development life cycle (Figure 1).

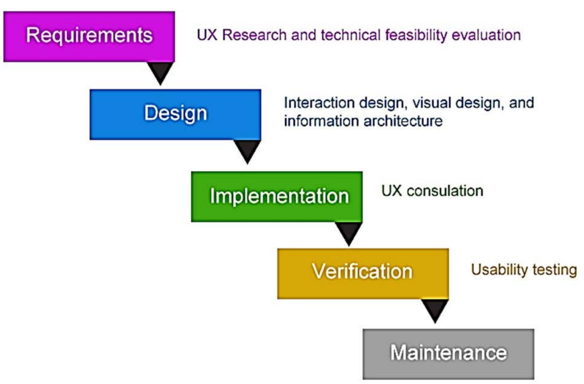


FIGURE 1. Integration of usability testing in SDLC *[47]*.

Mobile devices work differently from computer systems in terms of task performance, device design, and device operation [10]. The usability metrics for traditional computer design principles and guidelines are far apart from defining the mobile interface features of mobile phones [11]. As the features in mobiles vary from device to device, these pose a major challenge during the designing of usability tests. In short, the first mobile usability tests were derived and modified from the usability tests of desktop systems environment [12].

This research focuses on android application's usability using a scenario-based approach by selecting major usability attributes. The participants were selected on base of their education and age. In this work impact of education and age on the usability score of individual attributes are observed. This paper is divided into following sections.

Section II contains discussion about already present usability models in chronological order. Section III explains step by step approach of methodology adopted during evaluation. Section IV demonstrate research results and their discussions. Finally, Section V concludes this research.

II. BACKGROUND STUDY

This review aims to study the state of current evaluation methods on usability, to select an appropriate usability model that provides us with a better selection of usability attributes and guides us to develop improved GUI for a mobile application. The usability evaluation methods are selected in chronological order from oldest to recent. For example, prior to Shackle's model reported in 1991 only the usability attributes were reported in unstructured manner.

Shackel in 1991 proposed a set of 4 attributes (Effectiveness, Learnability, and Flexibility & Attitude) to evaluate usability. Shackel's idea heavily emphasizes on consideration of environments in which the system is used. This model was said to be used with any evaluation process (i.e. Expert Review, Simulation trials & Task-Based Evaluation) [13]. Chen *et al.* [14] conducted a study for library practitioners, taking a strong-technology-centered approach in defining usability, which was highly in favor of Shackel's approach. Koohang [15] proposed his usability attributes based on shackel's feature set.

In continuation of Shackel's model, Nielsen [4] in 1993 offered his definition of usability which initially comprises of 4 attributes (Learnability, Effectiveness, Efficiency, and Satisfaction), later he removed "Effectiveness" and added "Memorability and Errors" hence 5 attributes. This definition was well accepted and received more attention from the HCI community due to the use of attributes for users' perception of the system and Recall (Satisfaction and Memorability). Muqtadiroh *et al.* [16] performed a web usability study using Nielsen's usability attributes and found the approach satisfactory. Farahani and Khajouei [17] performed a usability study by combining ISO and Nielsen model to evaluate HIS (Hospital Information System). In 1994, Nielsen [5] proposed a process for heuristic (checklist) evaluation which was also perceived well by the Software Engineering community and was documented in [18], [19].

With the acceptance of Nielsen's model, the ISO (International Standard organization) in 1998 presented their definition for usability evaluation "ISO 92411-11" [20] in the hope of making it a global standard. This definition consists of 3 attributes (Effectiveness, Efficiency, and Satisfaction). This attribute selection was the already familiarity with usability attributes in the research community. As ISO presented the model, it was well accepted by the community. Farahani *et al.* [17] and Moumane *et al.* [21] performed usability evaluation on mobile applications using this model, while Mkpojiogu *et al.* [22] and Pradnyana *et al.* [23] performed usability evaluation on web-based applications using this model.

In 2001 [24] ISO revised their previous model (9241-11) into "ISO 9126-1" and also updated their attribute set from "Effectiveness, Efficiency, and Satisfaction" to "Functionality, Maintainability, Portability, Reliability, Efficiency and Usability" for usability evaluation to keep up with the continuously changing field of software development. To date, this model is widely used in the HCI community for both general and web-based applications in terms of attribute-based evaluation. The authors in [21], [25]–[29] performed usability evaluation of mobile (general) applications using this model, while Suwawi *et al.* [30] performed usability evaluation of academic website using this model.

In 2013 Harrison *et al.* [31] presented their framework of usability model called People At The Center Of Mobile Application Development (PACMAD), which consists of 7 attributes "Effectiveness, Efficiency, Satisfaction, Learnability, Memorability, Errors and Cognitive load," the main objective of this model was to extend the existing models (ISO 9241-11 Nielsen [4] and [20]) to the environment of mobile applications usability (i.e. to allow the addition of extra services during the development process to facilitate user). The main contribution of this model was the inclusion of "Cognitive load (context in which application is being used)" as a measuring attribute. The author's defined new definitions and the measuring criteria for each attribute especially "Cognitive load" as for the first time, it's used as a dependent variable in conjunction with other attributes [31]. Dalal *et al.* [32] presented some guidelines while testing mobile applications in extension with the PACMAD model. Saleh *et al.* [33] conducted a review, based on current practices in usability and found out PACMAD suitability with mobile applications, also the authors presented the GQM (Goal Question Metrics) approach for PACMAD (an extension) as if PACMAD was to be considered as Heuristics based evaluator instead of Attribute. Kasali *et al.* [34] performed a usability study using the feature (Attribute) selections from the PACMAD model & Integrated Measurement Model (IMM) on mobile health application. Zahra *et al.* [35] conducted a mobile E-Marketplace application usability study using the feature set of the PACMAD model.

Table 1 shows usability evaluation models along with their used attributes and definitions. It also shows how each attribute variates in terms of definitions (measuring criteria).

In this work we've gathered data by using 3 techniques (i.e. Questionnaires, Recordings and Oral feedbacks), to study the effect of user's age on mobile-application usability by segregating users into 2 major focus groups based on their education. These effects were further explored in terms of users' age by segregating each focus group into 4 sub-groups. Further the application usability is calculated by using users' opinions, recordings, feedbacks, age and education in terms of each PACMAD usability attribute individually and collective effect.

III. METHODOLOGY

This study follows the steps identified for usability evaluation in "Handbook of Usability" [36] (Figure 2).

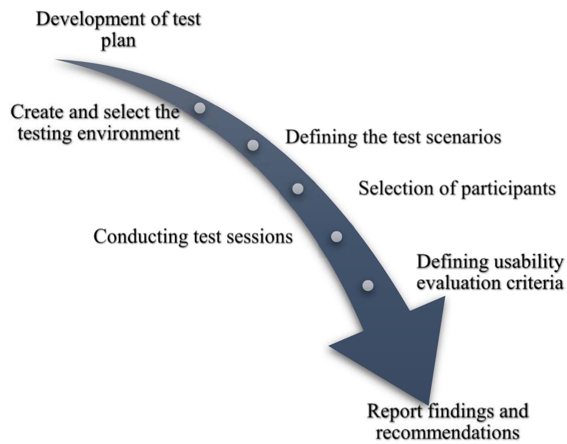


FIGURE 2. **Steps involved in methodology.**

A. DEVELOPMENT OF TEST PLAN

The test plan has two stages which are explained as follows.

1) SELECTION OF TESTING MOBILE APPLICATION

We selected the android mobile application named "Houzcalls" for this study. This app was loaned to us by a Company named "System Junctions Pvt Ltd." The app is a working prototype. The company required us to test out their app by conducting user reviews in a proper testing environment. To properly test the App first, we've to know about the application's working.

Houzcalls is an android-based application aimed to develop a connection between service providers and customers. The basic theme is to rapidly connect a customer with a service provider, from a medical doctor, plumber, carpenter, etc.

2) SELECTION OF USABILITY MODEL

In this study "PACMAD Usability Model" [31] framework has been adopted as it incorporates the usability attributes of "Nielsen" [4] and "ISO 9241" [20] and more importantly it includes "Cognitive load." Cognitive load refers to the amount of cognitive processing required by the user to use the application. In addition, it keeps in checks the emotion user

TABLE 1. Usability evaluation models.

Usability Evaluation Models	Attributes	Definitions
Shackel (1991)	Effectiveness	System's performance is better than the requisite level, by required percentage of the specified target range of users, within some required portion of the range of usage settings.
	Flexibility	Positive changes in the system to the existing ones (Updation)
	Learnability	Training of users after some time with the installed system.
	Attitude	Acceptance of users towards the system.
Nielsen (1993)	Efficiency	An efficient system results in high throughput.
	Learnability	The system should be easy to learn and understand.
	Satisfaction	A user's pleasant feeling after using the system is termed as likeability.
	Memorability	Termed as the return of user to the system without starting again from the beginning.
	Errors	Recovery of a system from error, it describes the error rate should be less.
ISO-9241-11 (1998)	Effectiveness	Accuracy and completeness with which users achieve specified goals.
	Efficiency	Resources expended in relation to the accuracy and completeness with which users achieve goals.
	Satisfaction	Freedom from discomfort, and positive attitudes towards the use of the product.
ISO-9126-1 (2001)	Portability	A set of attributes that bear on the ability of software to be transferred from one environment to another.
	Functionality	A set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied needs.
	Reliability	A set of attributes that bear on the capability of software to maintain its level of performance under stated conditions for a stated period of time.
	Usability	A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.
	Maintainability	A set of attributes that bear on the effort needed to make specified modifications.
	Efficiency	A set of attributes that bear on the relationship between the level of performance of the software and the amount of resources used, under stated conditions.
PACMAD (2013)	Effectiveness	The completion of a required task by the user.
	Efficiency	The completion of tasks by the user with speed and accuracy.
	Satisfaction	Pleasantness the user feels when use.
	Learnability	The ease with which user can gain proficiency in using the software.
	Memorability	The ease with which user can recall the previous use of the software.
	Errors	The use of software without running into errors. Low errors mean satisfactory performance.
	Cognitive Load	The mental capability required to do the task. This factors in with the user's current condition along with environmental factors.

TABLE 2. Intersection of usability attributes among authors.

Attributes	Author							
	[24]	[4]	[38]	[20]	[39]	[40]	[41]	[31]
Efficiency	X	X		X	X	X	X	X
Effectiveness		X		X		X	X	X
Learnability		X			X	X	X	X
Satisfaction		X		X	X		X	X
Memorability		X			X	X	X	X
Errors		X	X					X
Cognitive Load							X	X
Accessibility			X				X	
Enjoyment							X	
Content			X				X	
Acceptability					X		X	
Aesthetics							X	
Utility						X	X	
Flexibility							X	
Playfulness							X	
Functionality	X		X					
Portability	X							
Usability	X		X					
Reliability	X					X	X	
Maintainability	X		X					

feels while participating in tests. Table 2 and Weichbroth [37] show- Positive changes in the system to the existing ones (Updation).

B. DEFINING THE TEST SCENARIOS

Defining testing scenarios (tasks) has been always important while evaluating usability. Because the usability of application is measured by evaluating user performance in each task. In general, the tasks should cover all the functionality of the application. According to [4] one should remember the following points while defining the tasks:

- ❖ Make the task realistic.
- ❖ Make the task actionable.
- ❖ Avoid giving clues and describing the steps

We've prepared 5 scenarios for users, covering almost all major functionalities from the customer perspective of the evaluated application. Following are the tasks:

1) REGISTER YOURSELF

This task requires users to register themselves in the application for first-time use. The working layout of the task as follows:

Open Application ➔ Select Pakistan Mobile Code from the dropdown ➔ Enter Mobile No ➔ Check I agree Box ➔ Enter code sent by Application to mobile into required fields.

2) MAKE A CLEANING APPOINTMENT

This task requires users to make a cleaning appointment by filling the required spaces carefully. The working layout of the task as follows:

Open Application ➔ Select Cleaning Icon ➔ Write "Cleaning House" in Task Description Box ➔ Select Add-must have ➔ Write "Own broom," "Own buckets" & "Own chemicals" separately by clicking add button ➔ Select the address by using location icon ➔ Select Due date for the task ➔ Enter budget for the task.

3) MAKE A COPUTER SERVICE APPOINTMENT

This task requires users to make a computer service appointment by following the rules give to them. The working layout of the task as follows:

Open Application ➔ Select Computer Icon ➔ Write "Reset my system" in Task Description Box ➔ Select remotely task radio button ➔ Select Due date for the task ➔ Enter budget for the task.

4) MAKE A MEDICAL SERVICE APPOINTMENT

This task requires the user to carefully make a doctor's appointment by specifying the details provided. The working layout of the task as follows:

Open Application ➔ Select Medical Services Icon ➔ Select Doctor Visit ➔ Icon Select Add new for creating a new patient history ➔ Fill the required fields for profile creation & press save & continue ➔ Select newly created profile ➔ Select Injuries icon ➔ Select Multiple injuries ➔ Write "Full body fracture" in additional disease box ➔ Select address by choosing a location.

5) UPDATE EXISTING CLEANING APPOINTMENT

This task requires the user to update the existing cleaning appointment which they added in the first task. The working layout of the task as follows:

Open Application ➔ Select My History Tab ➔ Select Cleaning Task ➔ Change Due date ➔ Change budget Save/Update.

C. TESTING ENVIRONMENT

The testing environment plays a major role in any study. The main reason behind using a defined testing environment is to provide usability evaluators a relaxed and focused space.

We opted for "Laboratory testing" as a testing environment as according to [42] and [43]. "Laboratory testing" is more efficient and easier than field testing and the same issues can be found in laboratory testing as of field testing also with less external complexities.

For testing, we've selected a conference room because it's a secluded and dedicated space with a preinstalled high-definition camera. Figure 3 shows the furniture and camera setting for test conduction.

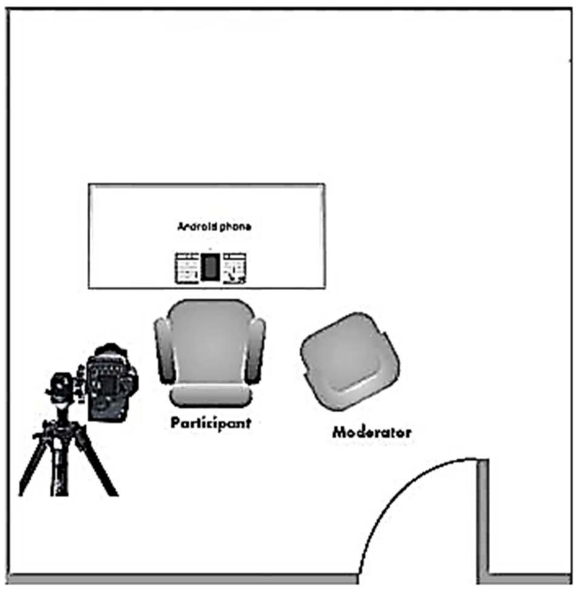


FIGURE 3. Testing environment setting.

D. SELECTION OF TESTING PARTICIPANTS

The selection of appropriate users/evaluators is important because we've to depend on the feedback/findings provided by them. This study is performed in Bahawalpur, Pakistan. The **76** selected participants were primarily divided into two groups differentiated based on their educational qualifications. For example, **Focus Group 1** (FG1) has participants having more than 10 years of education while **Focus Group 2** (FG2) participants have less than or equals **10 years of education**. The experiments were performed to identify which group perceived well by the application, the discerning factors, how to resolve these factors, and the similarity trends.

Each Focus Group participant is further divided into 4 subgroups based on their **age**. These groups are participants with **age less than 25** (under 25), **ages between 25 to 35** (25-35), **ages in between 36 to 45** (36-45), and finally participants with **age greater than 45** (over 45). This allowed us to study the impact of age in these two Focus Groups. All the subgroups have 10 participants except the FG2 with 45+ age which have only 6, because of the availability of very few android smartphone users (having age >= 45 && under educated) during testing in Bahawalpur.

E. CONDUCTION OF TEST SESSIONS

The usability evaluation comprises of three sessions, where session 1 and session 2 are performed on the same day while session 3 is performed after 5 days. Each session took maximum of 1 hour 30 minutes with 15 minutes break.

In session 1, participants were given orientation about testing and application tasks "Orientation (Welcome)," 2nd "Pre Test Questionnaire (participant information form and Non-Disclosure (Anonymous) forms, from both Participants and Moderator)" filled by the participants, 3rd "Explanation of Tasks (what to do)," 4th "Performing of Tasks by Participants" and 5th "Post Test Questionnaire (Satisfaction questionnaire and NASA TLX questionnaire)" got filled by the participants.

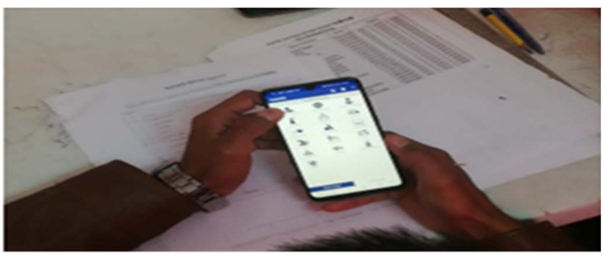
In sessions 2 and 3, 1st we did "Re orientation (again welcome)," 2nd "Emulating of tasks by participants," 3rd "Feedback" about tasks (application) using "Think aloud method" and in written form in suggestion section on satisfaction questionnaire.

Questionnaires both pre and posttest were used only in session 1. Also, an explanation of tasks to be performed was provided to participants in session 1 only.

Overall, the study comprised a total of 78 days. A maximum of 8 participants were called a single day.

Below Figure 4 (a, b) showing the test session in progress.

(a)



(b)

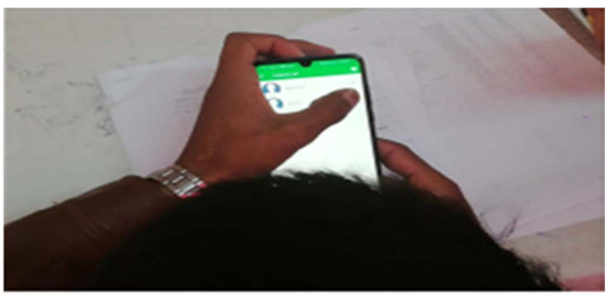


FIGURE 4. Usability test sessions in progress (a and b).

F. USABILITY EVALUATION CRITERIA

Firstly, each attribute of PACMAD is evaluated individually and later all these individual results are used to calculate overall usability of the application.

1) EFFICIENCY

It is measured in terms of task time, the time it took a user to complete a task. Where:

$$Efficiency = \frac{\sum_{i=1}^{N}\sum_{j=1}^{R} n_{ij}t_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{R} t_{ij}} \times 100\%$$

N = Number of tasks.

R = Number of users.

n_{ij} = the result of task by user; if the user completes the task, then is 1 else0.

t_{ij} = the time spent by user to complete task. If the task is not completed, then time is measured until the user quits the task.

We opted for relative efficiency because the tasks in our study are lengthy and require the above normal use of brainpower, especially for aged persons. Therefore, we picked relative efficiency to allow all the users the same benefit of task completion.

2) EFFECTIVENESS

It is a standard usability attribute that represents the completion rate of a task. If a user completed its task there assigns a value 1 if not then the value is 0.

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

According to a study conducted by [44], the average task completion rate was 78% but it was said that when conducting effectiveness, one should hope for a 100% success rate.

3) SATISFACTION

It is calculated using questionnaires which are filled by the users after session 1. We developed a hybrid questionnaire which comprises of 20 questions based on Likert 5 [45] scale ratings (Strongly Disagree (1) - Strongly Agree (5)). This questionnaire is filled only once in session 1 as these have to be applied to users who have not used the application before.

4) LEARNABILITY

According to [4] it can be calculated by comparing the results of multiple sessions of the same tasks. For this, we conducted two sessions (session 1 and session 2) of repetitive tasks on the same day to identify the user's learning curve regarding the application.

5) MEMORABILITY

It can be calculated by comparing the results of the same repetitive task sessions over some time. For this, we conducted session 3 five days after session 1 and session 2. In session 3 same tasks as of session 1 and session 2 are performed without elaboration to evaluate the memorability of the user regarding this application.

6) ERROR OCCURRENCE

Number of errors the user makes during tasks. A screen recording application records tasks performed by the user. These recordings are further reviewed to calculate critical errors, noncritical errors, and wrong turns, etc. that occurred during performance.

7) COGNITIVE LOAD

This is calculated by applying the NASA-TLX [46] questionnaire at the end of session 1 only because it provides better results when the user who is answering the questionnaire hadn't done the task before. It's a 6 questions form which inquiries about:

1. Mental Demand
2. Temporal Demand
3. Performance
4. Physical Demand
5. Effort
6. Frustration

It is to be said that all the above attributes have been collected, processed, calculated, and evaluated for each participant individually keeping their education and age in consideration. Each focus group participant's attribute scores are added together to calculate overall usability of the mobile application based on education and age. The obtained data was calculated with SPSS version 11.0. Descriptive statistics such as frequencies, percentages, means, variance and standard deviations are calculated.

Following are the features, which makes this study distinct:

- Different mechanisms (questionnaires, screen recordings and oral feedback) of data collection to infer better and more evident results.
- A full approach on Usability evaluation of mobile application (top-bottom) through PACMAD model.
- Division of participants (users') in terms of education and age.
- 3 separate usability evaluation sessions were conducted in order to thoroughly evaluate usability and also to produce more accurate results.
- Exploration of results calculated from experiment in terms of individual attribute and overall application usability. Both individual and overall results were formulated and presented on the basis of education and age.
- Reporting of issues facing the evaluated application in terms of critical and noncritical.

IV. RESULTS AND OBSERVATION

Till date there is no documented usability evaluation reported in which the above mentioned methodology was used. During evaluation users were selected according to their age and education

The results are compiled by an extensive study of data gathered during the usability evaluation (questionnaires, screen recordings, and oral feedback). The participants who completed the given task scored 1 and at the same time, we measured the time needed to complete the activity. In contrast, users who completed the task incorrectly (skip details) or gave up scored 0. Time was also measured for incomplete (given up) tasks. All the timings were measured through session recordings. The equations for Effectiveness, Relative Effectiveness, Task based Efficiency, and Relative Efficiency were utilized. After that, all errors made by participants during the tasks were counted. For the Learnability attribute, we compared the participants' results in session 1 with the results in session 2. Memorability was then measured by comparing the participant's results in session 2 and session 3. Both satisfaction and cognitive load were calculated using questionnaires submitted only during session 1. Finally, we identified the overall usability of the application concerning focus groups and each age group in terms of comparison.

TABLE 3. Effectiveness results for FG1 and FG2.

	Age Groups	Session 1	Session 2	Session 3	Average (Age wise)
FG 1	Under 25	94%	99%	78%	90.3%
	25-35	96%	99.2%	84%	93.1%
	36-45	94%	99%	75%	89.3%
	Over 45	71%	83%	55%	69.7%
	Average (Session wise)	88.9%	95.0%	73.0%	85.6%
FG 2	Under 25	90%	97%	67%	84.7%
	25-35	90%	96%	70%	85.3%
	36-45	70%	87%	52%	69.7%
	Over 45	52%	67.3%	29%	49.3%
	Average (Session wise)	75.4%	86.8%	54.5%	72.2 %

A. EFFECTIVENESS

The overall effectiveness of the application is shown in Fig 5, where FG1's accumulative effectiveness is 85.6% while FG2 effectiveness is 72.2%. For FG1 and FG2, session 2 has shown the highest effectiveness score among three sessions while session 3 has the lowest. Despite the low effectiveness in session 3 the overall effectiveness of application remained 79% which is a good outcome as compared to [44]'s 78% mark.

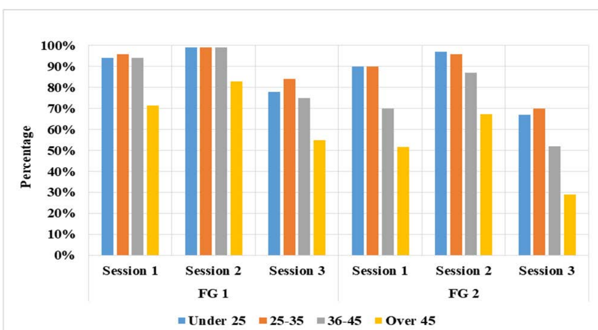


FIGURE 5. Effectiveness variation (session wise) among FG's.

According to Table 3. and Figure 5, the highest scoring age group was "25-35" in both "FG1 (93.1%) and FG2 (85.3%)," the age group "Under 25" scored second highest effectiveness in both "FG1 (90.3%) and FG2 (84.7%)." Age group "36-45" in "FG1 (89.3%)" also highly favored the application towards effectiveness but on the FG2 side, the effectiveness was less than 70% showing low favorability towards application.

The age group "Over 45" in both FG1 (69.7) and FG2 (49.3) was least effective (difficult to use) towards the application, therefore the highest task incompletion rates. It is also said that none age group from FG1 and FG2 had scored an ideal percentage value of 100% because not a single age group could complete their task with full effectiveness.

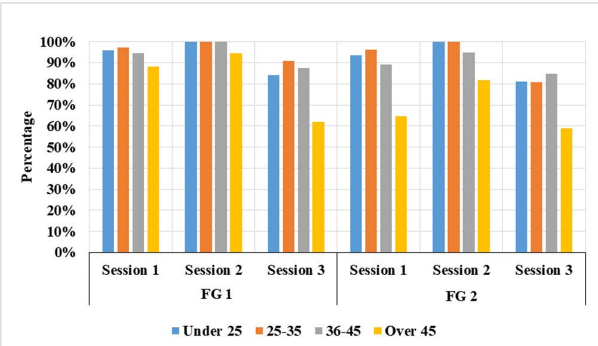


FIGURE 6. Efficiency variation (session wise) among FG's.

TABLE 4. Number of incomplete tasks in each session.

Age	FG1 S1	FG1 S2	FG1 S3	FG2 S1	FG2 S2	FG2 S3
Under 25	0	0	5	2	0	8
25-35	0	0	3	2	0	7
36-45	0	0	8	4	0	10
Over 45	8	2	18	10	4	13
Total	8	2	34	18	4	38

According to Table 4, it is clear in FG1 and FG2 the highest numbers of uncompleted tasks were in session 3. In FG1, 34 out of 44 uncompleted tasks were during session 3 and in FG2, 38 out of 60 uncompleted tasks were during session 3. The reason behind a decline in task completion (effectiveness) in session 3 is due to the fact that session 3 was held five days after session 2, some participants forgot about what tasks were and how to do them. From Table 4 it is evident that the "Over 45" group had maximum of incomplete tasks.

B. EFFICIENCY

In this work relative efficiency is calculated instead of time-based efficiency because this study compared four age groups from youngsters to elders. However, youngsters have an advantage in time-based efficiency because of their quick reflexes and proper knowledge of smartphones.

TABLE 5. Efficiency results for FG1 and FG2.

	Age Groups	Session 1	Session 2	Session 3	Average (Age-wise)
FG 1	Under 25	95.74%	100%	84.27%	93.3%
	25-35	97.11%	100%	90.69%	96%
	36-45	94.55%	100%	87.46%	94%
	Over 45	88.24%	94.60%	61.81%	82%
	Average (Session wise)	94%	99%	81.1%	91.3%
FG 2	Under 25	93.45%	100%	80.97%	91.5%
	25-35	96.24%	100%	80.72%	92.3%
	36-45	89.18%	94.94%	84.94%	89.7%
	Over 45	64.57%	81.94%	59.02%	68.5%
	Average (Session wise)	86%	94%	76%	85.3%

TABLE 6. Satisfaction results for FG1 and FG2.

	Scales	Under 25	25-35	36-45	Over 45	Overall Average
FG 1	Strongly Disagree	0.5	1	0.5	6.5	2.12
	Disagree	1	2.5	8	21.5	8.25
	Neither Agree nor Disagree	14.5	11.5	26	30	20.50
	Agree	43.5	47.5	47	31.5	42.38
	Strongly Agree	40.5	37.5	18.5	10.5	26.75
FG 2	Strongly Disagree	2	2.5	4.5	12	5.25
	Disagree	15	12	16.5	26	17.38
	Neither Agree nor Disagree	31.5	35.5	38.5	27	33.12
	Agree	37	35	30.5	29	32.88
	Strongly Agree	14.5	15	10	6	11.38

Table 5 contains the efficiency evaluation results for FG1 and FG2. The last row represents the average of each session for FG1 and FG2 while 6th column shows the average of different age groups for FG1 and FG2 respectively. Each participant performed 5 tasks in each session.

The overall efficiency results for FG1 and FG2 are 91.3% and 85.3% (Table 5) respectively. In both groups the efficiency pattern has remained the same (Figure 6), the highest efficiency is in session 2 (99% and 94%) while the lowest is in session 3 (81% and 76% respectively). "25-35" subgroups in both focus groups have secured the highest efficiency. Only the "Over-45" group in FG2 has found the tasks difficult as their efficiency remained under 80%.

TABLE 7. Satisfaction scale score by FG1 and FG2 among age groups individual and overall (%).

Age Groups	FG1		FG2	
	Mean	Variance	Mean	Variance
Under 25	4.21	0.61	3.41	1.09
25-35	4.18	0.65	3.47	1.00
36-45	3.75	0.78	3.25	1.03
Over 45	3.60	0.80	2.90	1.27

C. SATISFACTION

Satisfaction was obtained using a "Hybrid" questionnaire based on "Likert 1-5 score (1 being the lowest and 5 being the highest)."

Table 7 shows the mean and variance of satisfaction results for each age group in FG1 and FG2. The value of mean is cumulative and proportional to participants' satisfaction. The value of variance represents how many participants have similar opinions in subgroups, the lower value of variance means more participants have similar level of satisfaction.

In Table 7, the highest satisfaction score is observed for "Under 25 and 25-35" subgroups in FG1, and also low spread in participants' opinion is evident by low variance. On the other hand, "36-45 and Over 45" subgroups in FG1 have lower satisfaction mean and opinions have a wider spread. Overall satisfaction for each subgroup of FG2 remained less than any subgroup in FG1, it can be seen that variances are higher as well. It indicates that FG2 participants had mixed satisfaction feedback and variations in feedbacks become wider with the increase in age. Generally, for FG1 the variance of subgroups remained under 0.8 which shows most of the participants have an almost identical opinion about the application usability. Whereas in FG2 the variance remained more than 1 which shows that participants have varying opinions about application usability in each group.

Table 6, shows how the participants' opinion variates according to the Likert scale. For "Under 25 and 25-35" subgroups for FG1 more than 80% remained Strongly Agreed and Agreed. For "36-45 and Over 45" subgroups for FG1 opinions start to gradually shift from Strongly Agreed to Disagree. For "Under 25 and 25-35" subgroups for FG2 opinions, 70% lies in the Agreed and Neutral scale. With the increase in age, FG2 participants' opinions shifted from Agreed to Disagreed. Therefore, it can be concluded that this application is better suited for users with a higher education.

D. LEARNABILITY

Learnability was measured using effectiveness. The reason for choosing effectiveness as a point to identify learnability, as effectiveness was calculated by keeping an accurate check of each participant's errors while performing tasks. So, to work out accurate learnability, we've to base our comparisons on the results having definitive accuracy which was effectiveness.

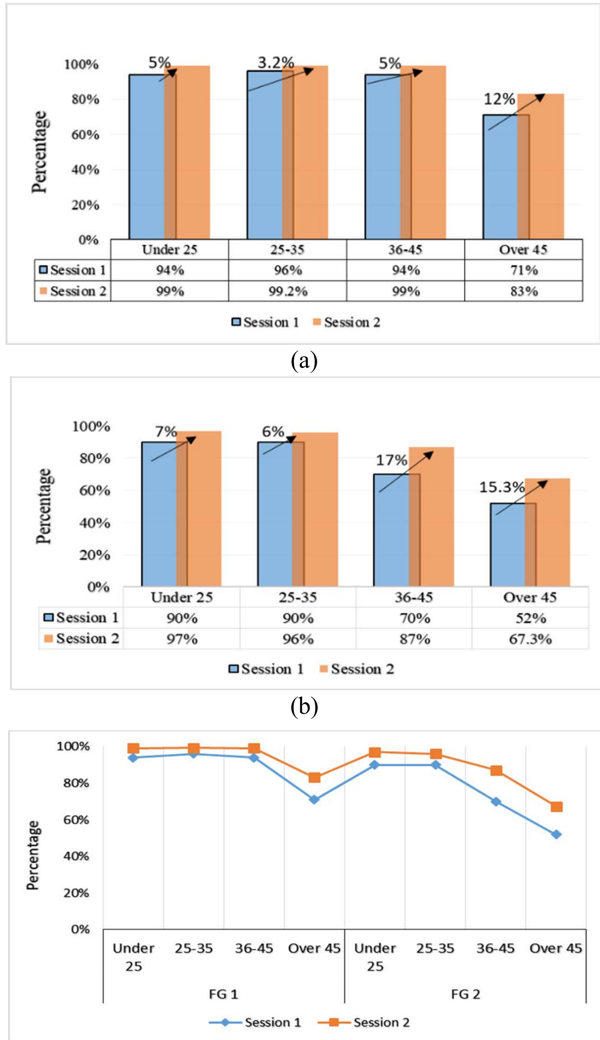
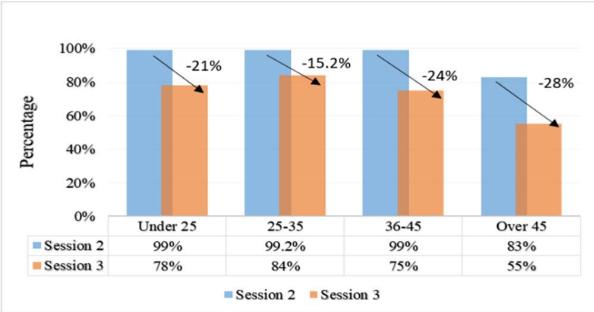


FIGURE 7. (a) FG1 Learnability for Session 1&2, (b) FG2 Learnability for Session 1&2, and (c) Learnability trend of FG1 and FG2.

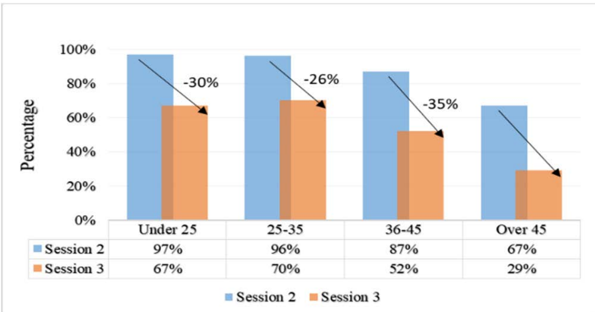
TABLE 8. Error classification.

Critical Errors	**1.** Server Times out mostly. **2.** UI font colors and panes not equipped for sunlight. **3.** Map API doesn't respond well. **4.** Medical Service too much waiting while booking an appointment.
Non Critical Errors	**1.** The description of buttons/menus isn't defining. **2.** The calendar's elements fade under sunlight. **3.** Add must-haves menu isn't defined well. **4.** No Country name with dialing code. **5.** Medical terms aren't well defined. **6.** Too many items to select while booking an appointment with the doctor.

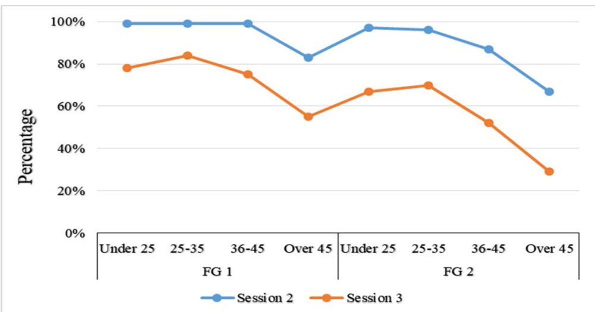
Learnability was measured by comparison of session 1's effectiveness to session 2's effectiveness. Figure 7 (a, b) shows FG1 and FG2 learnability, and (c) shows trend in Learnability for both focus groups.



(a)

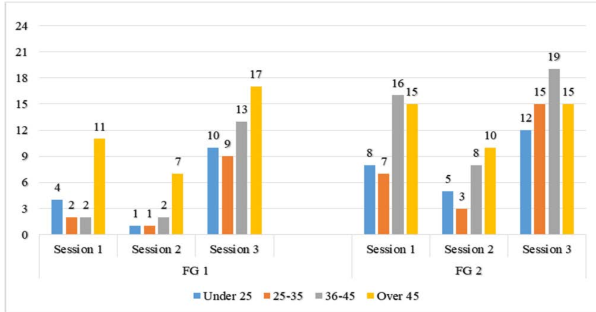


(b)

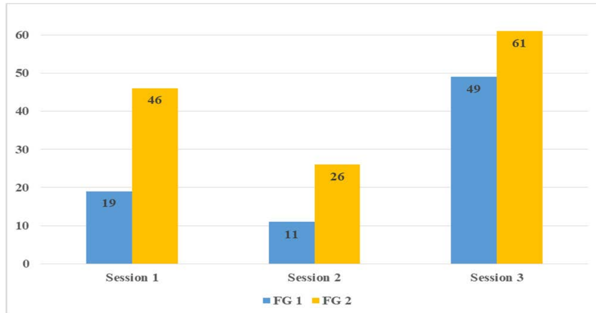


(c)

FIGURE 8. **(a) FG1 Memorability for Session 2&3, (b) FG2 Memorability for Session 2&3, (c) memorability trend of FG1 and FG2.**



(a)



(b)

FIGURE 9. **(a) Session wise error comparison FG1 and FG2. (b) Error count for FG1 and FG2.**

"Over 45" have attained a gain of 12% learnability in FG1 and 15.3% in FG2, whereas, overall FG1 has shown better learnability. The age group "36-45" in FG2 has got the highest effectiveness gain of 17% in session 2. On the other hand, the effectiveness gain values of age groups "Under 25, 25-35 and 36-45" in FG1 and "Under 25 and 25-35" in FG2 are low because their respective session 1 scores were very high hence making the ideal effectiveness gain of almost 100%. Overall, learnability gain for both FG1 and FG2 remained very high still FG1 showed better learnability as compared to FG2.

E. MEMORABILITY

Memorability is a recalling power of a user which gradually decreases with respect to age and time. It is measured by comparing session 2's effectiveness and session 3's effectiveness. In an ideal case, there should be no decline in session 3's effectiveness. However, Neilson [5] identified that if memorability declined almost 30% from the previous result, it is still considered a good recall for the product.

Figure 8 (a) shows memorability score of all the age groups in FG1. Every age group has favored memorability because their individual decline is less than 30%. Figure 8 (b) shows memorability score of all age groups in FG2 where subgroups "Under 25" and "25-35" have acceptable recalls, all the other groups have more than 30% decline rate. In both groups "Over 45" have the highest decline in memorability due to the age factor. Figure 8 (c) shows the trend of memorability among FG1 and FG2 as both groups follow the same pattern of decline in age groups in terms of highs and lows.

F. ERRORS

Figure **9** (a) shows a session wise cumulative error comparison between FG1 and FG2 and Figure 9 (b) shows the error rate for all subgroups in FG1 and FG2. All the subgroups have encountered one or more errors while performing their tasks. The graph has demonstrated that the less educated group FG2 has encountered a large number of errors as compared to FG1. The error occurrence trend among the sessions has remained similar in each focus group, and most errors have occurred in session 3. In FG1, over 45 has committed the most 35 errors in all three sessions, and 25-35 has committed the least, 12 errors. For FG2, 36-45 has committed 43 errors and over 45 has committed 40 errors. The FG2 under 25 and 25-35 groups have the least 25 error encounters. The results have shown that the more educated group has performed fewer errors and over 45 participants in both focus groups have committed many errors.

During testing, participants indicated errors, mistakes, and shortcomings in the tested application, which affected usability. The errors are classified as Critical (errors which heavily

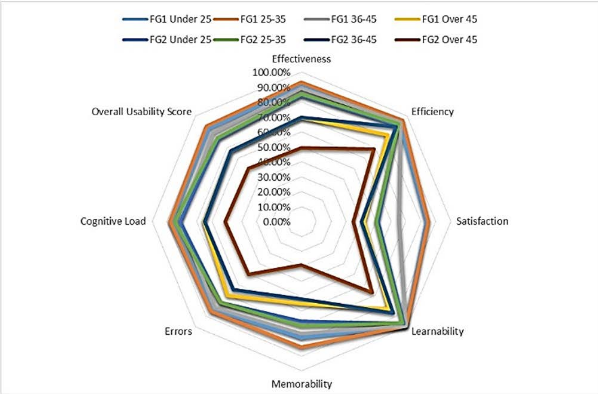
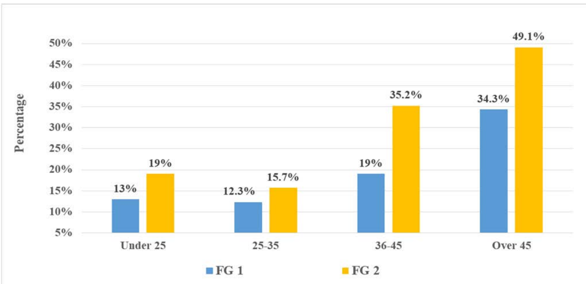
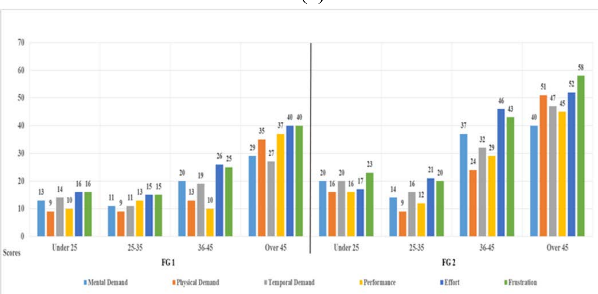


FIGURE 10. Complete results.



(a)



(b)

FIGURE 11. (a) Cognitive load subscales for FG1 and FG2. (b) Cognitive load for FG1 and FG2 are groups.

affect usability (needs to resolve on priority)) and Non critical (errors which don't directly affect usability (can be resolved later if not now)). Some of these errors are mentioned in Table 8.

The application was perceived well and tasks were completed efficiently due to well defined input fields, detailed menus. The major error occurrence was recorded in session 3 because users found the terminologies and options menus of the application to be difficult to use after a long time. Also, the application didn't provide any error recovery features (i.e., undo, redo, etc.), the use of simple language for application GUI that has better understandability for all types of users. This application should provide multiple languages support.

TABLE 9. Complete usability (individual + overall).

Usability Attributes	FG1			
	Under 25	25-35	36-45	Over 45
Effectiveness	90.3%	93.1%	89.3%	69.73%
Efficiency	93.3%	96%	94%	82%
Satisfaction	84%	85%	65.5%	42%
Learnability	99%	99.2%	99%	83%
Memorability	78%	84%	75%	55%
Errors	83.3%	85.3%	82%	70%
Cognitive Load	87%	88%	81.2%	65.3%
Overall Usability	87.84%	90.09%	83.71%	66.72%
	FG2			
Effectiveness	84.7%	85.3%	69.67%	49.3%
Efficiency	91.5%	92.3%	89.7%	68.5%
Satisfaction	51.5%	50%	40.5%	35%
Learnability	97%	96%	87%	67.3%
Memorability	67%	70%	52%	29%
Errors	76.7%	76.7%	64.7%	50%
Cognitive Load	81.3%	84.7%	64.8%	51.2%
Overall Usability	78.53%	79.29%	66.91%	50.04%

G. COGNITIVE LOAD

Recalling "Usability Evaluation Criteria:" participant's input was obtained on the application's task load by using a "NASA's TLX" questionnaire [46], through which the tasks load scores for each participant based on subscales and workload is determined. "Subscales" & "Workload" results are discussed here for a better conclusion. Figure 11 (a) shows that the overall task load outcome of the application was quite biased based on education.

Figure 11 (b) compares age wise cognitive load for both focus groups. "25-35" age group has the lowest cognitive load and "Over 45" age group has the highest load in FG1 and FG2. Age groups "35-45" and "Over 45" for FG2 have high cognitive load. The cognitive load is inversely proportional to the participant's satisfaction, if the cognitive load approaches 50 % then the user is considered unsatisfied. It is observed that despite having a prominent difference in workload in FG1 and FG2 participants, feedback was in favor of the application. It can be concluded that FG1 (more educated) participants were more satisfied than FG2 participants.

Therefore, by carefully considering results from subscales and workload, we can say that despite the overall highest scores of the age group "over 45" in FG1 and FG2 the remaining groups still favored the application in terms of ease of use, etc. It is also said during cognitive load study it was clear that without proper education one can't fully comprehend the advancements of technology, because many participants felt the need of using much more brainpower in understanding the basic "English written terms" or "Performing task" especially in FG2 as of their counterparts in FG1.

Figure 10 and Table 9 contains usability data for all attributes of the PACMAD model along with overall application usability.

V. CONCLUSION

Our study has shown that education directly affects the usability of applications while age has inversely affected. This pattern can be seen in all PACMAD's attributes. It is observed that for both FG1 and FG2, behavior of different age groups have remained the same. "25-35" subgroups in both FGs have shown the best usability which is followed by "Under 25" subgroups. "36-45 and Over 45" for FG2 and "Over 45" of FG1 have lowest usability scores. "25-35" subgroup is the best FG2 but it has performed less than "36-45" subgroup in FG1.

"Over 45" from both FGs and "36-45" from FG2, have obtained Usability scores of under 68 and their satisfaction score remained under 50%. It can be inferred that more useable applications can be developed by keeping "36-45" and "Over 45" subgroups in focus which'll result in development of more usable software for users of all age. To make this statement more acceptable further research on mobile application usability is required using a diverse set of applications (website, distributed systems, expert systems, etc.) along multiple demographic attributes of participants.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present study and all authors contributed equally scientifically.

REFERENCES

[1] C. Hope. *Computer vs. Smartphone*. Accessed: Feb. 8, 2019. [Online]. Available: https://www.computerhope.com/issues/ch001398.htm

[2] W. Gadzama, B. Joseph, and N. Aduwamai, "Ownership, internet usage and their impacts on humans," *J. Commun. Netw.*, vol. 1, no. 1, pp. 1–10, 2017.

[3] T. Counterpoint. *Global Smartphone Market Share: By Quarter*. Counterpoint. Accessed: Apr. 30, 2021. [Online]. Available: https://www.counterpointresearch.com/global-smartphone-share/

[4] J. Nielsen. (1994). *Usability 101: Introduction to Usability*. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability/

[5] J. Nielsen. *10 Usability Heuristics*. Accessed: Apr. 24, 1994. [Online]. Available: https://www.nngroup.com/articles/ten-usability-heuristics/

[6] S. H. Mohammed, F. W. Ahmad, and S. Suziah, "Usability evaluation of the smartphone user interface in supporting elderly users from experts' perspective," *IEEE Access*, vol. 6, pp. 22578–22591, 2018.

[7] R. Parente Da Costa, E. D. Canedo, R. T. De Sousa, R. De Oliveira Albuquerque, and L. J. Garcia Villalba, "Set of usability heuristics for quality assessment of mobile applications on smartphones," *IEEE Access*, vol. 7, pp. 116145–116161, 2019.

[8] K. Moran. *Usability Testing 101*. Nielsen Norman Group. Accessed: Dec. 1, 2019. [Online]. Available: https://www.nngroup.com/articles/usability-testing-101/

[9] B. A. Myers, "User interface software technology," *ACM Comput. Surv.*, vol. 28, no. 1, pp. 189–191, Dec. 1996.

[10] R. B'Far, *Mobile Computing Principles: Designing and Developing Mobile Applications with UML and XML.* Cambridge, U.K.: Cambridge Univ. Press, 2005.

[11] H. Jeongyun, H. Dong-Han, P. Sanghyun, S. Chiwon, and Y. Wan Chul, "A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors," IN *Interacting with Computers*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 263–275.

[12] D. Zhang and B. Adipat, "Challenges, methodologies, and issues in the usability testing of mobile applications," *Int. J. Hum.-Comput. Interact.*, vol. 18, no. 3, pp. 293–308, Jul. 2005.

[13] B. Shackel, "Usability–context, framework, definition, design and evaluation," *Interacting Comput.*, vol. 21, nos. 5–6, pp. 21–37, 2009.

[14] Y.-H. Chen, A. Rorissa, and C. A. Germain, "Usability definitions in a dynamically changing information environment," *Portal: Libraries Acad.*, vol. 15, no. 4, pp. 601–621, 2015.

[15] A. Koohang, "Expanding the concept of usability," *Int. J. Emerg. Transdiscipline*, vol. 7, pp. 129–141, Jan. 2004.

[16] F. A. Muqtadiroh, H. Astuti, E. W. T. Darmaningrat, and F. Aprilian, "Usability evaluation to enhance software quality of cultural conservation system based on nielsen model (wikibudaya)," *Proc. Comput. Sci.*, vol. 124, pp. 513–521, Jan. 2017.

[17] R. Khajouei and F. Farahani, "A combination of two methods for evaluating the usability of a hospital information system," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–10, Dec. 2020.

[18] E. Gonzalez-Holland, W. Daphne, L. Moralez, and M. Mouloua, "Examination of the use of Nielsen's 10 usability heuristics & outlooks for the future," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2017, vol. 61, no. 1, pp. 1472–1475.

[19] E. Wong. (Jul. 2020). *Heuristic Evaluation: How to Conduct a Heuristic Evaluation*. Accessed: 2020. [Online]. Available: https://www.interaction-design.org/literature/article/heuristic-evaluation-how-to-conduct-a-heuristic-evaluation

[20] *Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs)—Part 11: Guidance on Usability*, Standard ISO 9241-11:1998, 1998.

[21] K. Moumane, A. Idri, and A. Abran, "Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards," *SpringerPlus*, vol. 5, no. 1, p. 548, Dec. 2016.

[22] A. Hussain, E. Mkpojiogu, and H. Zakaria, "Usability evaluation of a web-based health awareness portal on Smartphone devices using ISO 9241-11 model," *Jurnal Teknologi*, vol. 77, no. 4, pp. 1–5, 2015.

[23] I. K. R. Arthana, I. M. A. Pradnyana, and G. R. Dantes, "Usability testing on website wadaya based on ISO 9241-11," in *Proc. J. Phys., Conf.*, vol. 1165, Feb. 2019, Art. no. 012012.

[24] *Software Product Evaluation-Quality Characteristics and Guidelines for Their Use*, Standard ISO 9126, 2001.

[25] B. Chua and L. Dyson, "Applying the ISO 9126 model to the evaluation of an e-learning system," in *Proc. 21st ASCILITE Conf.*, Perth, WA, Australia, 2004, pp. 1–7.

[26] K. Toshihiro, "Usability evaluation based on international standards," *NEC Tech. J.*, vol. 3, no. 2, pp. 27–32, 2008.

[27] G. Wang, D. Bernanda, J. Andry, A. Fajar, and S. Sfenrianto, "Application development and testing based on iso 9126 framework," in *Proc. J. Phys., Conf.*, 2019, Art. no. 012011.

[28] B. Zeiss, D. Vega, I. Schieferdecker, H. Neukirchen, and J. Grabowski, "Applying the ISO 9126 quality model to test specifications–exemplified for TTCN-3 test specifications," in *Software Engineering 2007–Fachtagung des GI-Fachbereichs Softwaretechnik*. Bonn, Germany, Gesellschaft für Informatik e. V, 2007.

[29] S. Fahmy, N. Ngah, W. Othman, and Z. Fariha, "Evaluating the quality of software in e-book using the ISO 9126 model," *Int. J. Control Automat.*, vol. 5, no. 2, pp. 115–122, 2012.

[30] D. D. J. Suwawi, E. Darwiyanto, and M. Rochmani, "Evaluation of academic website using ISO/IEC 9126," in *Proc. 3rd Int. Conf. Inf. Commun. Technol. (ICoICT)*, Nusa Dua, Bali, Indonesia, May 2015, pp. 222–227.

[31] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: Literature review and rationale for a new usability model," *J. Interact. Sci.*, vol. 1, no. 1, pp. 1–16, 2013.

[32] N. N. Dalal and P. Patel, "Usability evaluation of mobile applications," *Int. J. Eng. Res. Technol. (IJERT)*, vol. 2, no. 11, pp. 299–302, 2013.

[33] I. Roesnita, F. Norasikin, and S. Ashraf, "Extension of PACMAD model for usability evaluation metrics using goal question metrics (GQM) approach," *J. Theor. Appl. Inf. Technol.*, vol. 79, no. 1, pp. 90–100, 2015.

[34] F. Kasali, O. O. Olaniyan, I. O. Akinyemi, O. B. Alaba, A. Oludele, and S. O. Kuyoro, "An enhanced usability model for mobile health application," *Int. J. Comput. Sci. Inf. Secur. (IJCSIS)*, vol. 17, no. 2, pp. 20–29, 2019.

[35] H. M. Az-zahra, N. Fauzi, and A. P. Kharisma, "Evaluating E-marketplace mobile application based on people at the center of mobile application development (PACMAD) usability model," in *Proc. Int. Conf. Sustain. Inf. Eng. Technol. (SIET)*, Lombok, Indonesia, Sep. 2019, pp. 72–77.

[36] J. Rubin, D. Chisnell, and J. Spool, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*, 2nd ed. New York, NY, USA: Wiley 2008.

[37] P. Weichbroth, ''Usability of mobile applications: A systematic literature study,'' *IEEE Access*, vol. 8, pp. 55563–55577, 2020.

[38] L. Lockwood and L. Constantine, ''Collaborative usability inspections for software,'' Miller Freeman, San Francisco, CA, USA, Tech. Rep., 2003. [Online]. Available: https://www.semanticscholar.org/paper/Usability-by-Inspection-

[39] S. Lauesen and H. Younessi, ''Six styles for usability requirements,'' in *Proc. Requirements Eng., Found. Softw. Quality, 4th Int. Workshop (REFSQ)*, Pisa, Italy, 1998.

[40] H. Sharp, J. Preece, and Y. Rogers, ''Interaction design,'' in *Interaction Design: Beyond Human-Computer Interaction*. New York, NY, USA: Wiley, 2002.

[41] C. Coursaris and D. Kim, ''A meta-analytical review of empirical mobile usability studies,'' *J. Usability Stud.*, vol. 6, no. 3, pp. 117–171, 2011.

[42] J. Kjeldskov, C. Graham, S. Pedell, F. Vetere, S. Howard, S. Balbo, and J. Davies, ''Evaluating the usability of a mobile guide: The influence of location, participants and resources,'' *Behaviour Inf. Technol.*, vol. 24, no. 1, pp. 51–65, Jan. 2005.

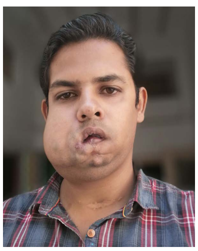
[43] A. Kaikkonen, T. Kallio, A. Kekäläinen, A. Kankainen, and M. Canka, ''Usability testing of mobile applications: A comparison between laboratory and field testing,'' *J. Usability Stud.*, vol. 1, no. 1, pp. 4–16, 2005.

[44] S. Jeff and L. R. James, ''Correlations among prototypical usability metrics: Evidence for the construct of usability,'' in *Human Factors in Computing Systems*. Boston, MA, USA, Assoc. Comput. Machinery, 2009.

[45] V. Preedy and R. Watson, ''5-point Likert scale,'' in *Handbook of Disease Burdens and Quality of Life Measures*. New York, NY, USA: Springer, 2010.

[46] T. Janet and A. Michael, ''Measuring cognitive load to test the usability of web sites,'' in *Proc. Annu. Conf.-Soc. Tech. Commun.*, 2006, pp. 256–260.

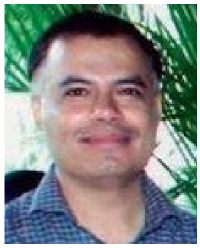
[47] Z. Hussain, W. Slany, and A. Holzinger, ''Current state of agile user-centered design: A survey,'' in *HCI and Usability for e-Inclusion* (Lecture Notes in Computer Science), vol. 5889. Berlin, Germany: Springer-Verlag, A. Holzinger and K. Miesenberger, Eds. 2009, pp. 416–427.



WAHAJ ALI received the M.S. degree in computer science from The Islamia University of Bahawalpur. He has worked on multiple projects involving requirement engineering in terms of user experience. He is a Lecturer in computer science with the Higher Education Department, Government of Punjab. His current research interests include software requirements engineering, user experience (Ux), and usability testing along with human aspects of software engineering (HCI).

OMER RIAZ received the B.S.C.S. degree from the University of Engineering and Technology Lahore, in 2004, the master's degree in system level integration from The University of Edinburgh, in 2005, and the Ph.D. degree in high performance computing from the University of Strathclyde, U.K., in 2014. His research interests include the study of challenges involve implementing time consuming algorithms on advance computer architectures (shared memory, distributed memory, and GPU). His research interests also include numerical analysis and machine learning algorithms.

SHAHZAD MUMTAZ received the Master of Computer Science degree from The Islamia University of Bahawalpur, Pakistan, in 2005, and the Ph.D. degree in computer science from Aston University, U.K., in 2015. He has also worked as an Assistant Director (Computer) at the National Highway Authority, Pakistan, from December 2005 to October 2007. His research interests include in the areas of machine learning and data mining and their application to health informatics domains but recently he has worked in other areas, such as natural language analytics and high-performance computing. His research projects include Probabilistic Modeling of Blood Glucose Through Eye Parameters, An Analysis of Protein Family of Major Histocompatibility Complex, Predictive Modeling of Accidents and Emergency Arrivals and Admissions, Patient Specific Recommendation Systems for HIP Joint Patients, and Predictive Modeling of Extreme Content from Twitter in the Context of Afghanistan.



AMJAD REHMAN KHAN (Senior Member, IEEE) received the Ph.D. and postdoctoral degrees (Hons.) from the Faculty of Computing, Universiti Teknologi Malaysia, with a specialization in forensic documents analysis and security, in 2010 and 2011, respectively. He is currently a Senior Researcher with the Artificial Intelligence and Data Analytics Laboratory, CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He is also a PI in several funded projects and also completed projects funded from MOHE, Malaysia; and Saud Arabia. He is the author of more than 200 ISI journal articles and conferences. His research interests include data mining, health informatics, and pattern recognition. He received the Rector Award for 2010 Best Student from Universiti Teknologi Malaysia.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently working as an Associate Professor with the College of Computer and Information Sciences, Prince Sultan University (PSU), Riyadh, Saudi Arabia. She has published more than 100 publications in high ranked journals. Her research interests include bioinformatics, data mining, and classification. She was awarded the Best Research of the Year Award at PSU, from 2013 to 2016. Due to her excellent research achievement, she is included in Marquis Who's Who (S & T) 2012. She won the Best Student Award from the Faculty of Computing, UTM, in 2012. She is also an editor of several reputed journals and on panel of TPC of international conferences.

SAEED ALI BAHAJ received the Ph.D. degree from Pune University, India, in 2006. He is currently an Assistant Professor with Prince Sattam Bin Abdulaziz University. His research interests include artificial intelligence, information management, forecasting, information engineering, big data, and information security.