Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

USABILITY EVALUATION REVEALS WHAT IS MISSING IN E-COMMERCE MOBILE APPS

QURAT-UL-AIN[1], MALIK MUHAMMAD SAAD MISSEN[2*], ASAD ALI[3] and AMNAH FIRDOUS[4]

[1, 2, 3]Dept. of Information Technology, The Islamia University of Bahawalpur.
[4]Dept. of Computer Science and Information Technology, The Government Sadiq College Women University Bahawalpur.
* Corresponding Author Email: saad.missen@iub.edu.pk

Abstract

Usability is one of the most important characteristics of software applications especially when it comes to mobile shopping applications. There is a great deal of shift from traditional shopping to online shopping because it benefits both parties i.e. customer as well as businessman. A customer just has to download an app to access a wide variety of products and product is at his home within span of two days at maximum while businessman benefits from millions of online shoppers who other- wise are not able to a physical store for any reason. The number of app downloads the number of active users and the number of online transactions is directly related to annual revenues of business. In such scenario, usability factor can play a very vital role for business industry. If a client stops using a mobile shopping app because it is not user friendly, it can badly damage the annual revenues especially when hundreds of alternatives are available and there is a tough competition. Therefore, to keep existing customers intact and to attract new customers, it is very important to provide a user friendly and confusion-less mobile app to customers. Any minor de-sign flaw can result in million-dollar loss in the long run. In this scenario, it becomes important to identify the real usability problems in shopping apps and to propose domain-specific design heuristics for mobile apps developers so that better mobile e- commerce apps could be developed to keep their customers engaged in the long run. There is a large variety of online customers with diverse requirements. Background and constraints and there is a need to evaluate usability of existing mobile ecommerce apps with diverse type of customers to identify the actual problems people are facing with existing applications. In this paper, we conduct usability evaluation experiments on large scale (i.e. 1080 total experiments) with a diverse category of participants on a sample of mobile shopping apps. We identify the actual usability problems people are facing in existing apps along with proposal of domain-specific heuristics for mobile ecommerce apps. Besides this, we provide a detailed systematic review of mobile ecommerce apps.

Keywords : Sections; lists; figures; tables; mathematics; fonts; references; appendices

1. INTRODUCTION

A good interaction between technologies and its users does not only improve user satisfaction, but also helps in fostering effective communications between them (2). Failure in satisfying its users or customers will create dissatisfaction among them eventually leading towards quitting the system. Hence, the provision of an effective, user-friendly and interactive system design for customers is vital to prolong the business (1). This is where the concept of Usability Engineering comes in. Usability engineering[1] is a field that is concerned generally with human–computer interaction and specifically with devising human–computer interfaces that have high usability or user friendliness.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

Technology has seen lot of progress during last few decades. Few years back the power of Internet and social media was a debatable but with the emergence of mobile devices, the real strength of Internet and social media is no longer a debatable topic but it has become a fact. Taking example of e-commerce, it is well known that many questions had been raised on emergence of the concept of e-commerce. The most common among these questions were," Would people adopt it?" "How would they behave?" "What did they want?" etc (3). But today when mobile Internet is the most popular form of Internet among users, the concept of e-commerce has taken the more popular name of m-commerce. Just like e-commerce, m-commerce offers a huge potential to its customers as well as business owners.

While m-commerce is making a well-paced progress and lots of mobile shopping apps are available to customers, several issues need attention and, in this work, our target is one issue which has been largely neglected among researchers as well as developers i.e., usability and the user interface experience. It is an accepted fact that usability is the biggest source of frustration for Internet users (3). Therefore, it is to be understood that in m-commerce user experience of mobile apps is directly related to annual revenues because a user quitting using a mobile shopping app because of his poor user experience will directly affect annual income of the business owner.

Mobile apps usability has attracted lot of attention of researchers recently and many have worked on proposing usability guidelines and usability models (4; 5; 9; 11). However, in this work, we intend to propose usability guidelines for mobile shopping apps by performing a very detailed usability evaluation using a variety of participants. The real intent is to find actual usability problems by exploiting the user experience of variety of customers especially for participants with physical tactile deformities and illiterates. The major research objectives of the presented work are described below:

- To identify the actual usability problems in mobile shopping apps by performing a detailed user-based evaluations,
- to evaluate the user experience of a wide variety of participants from participants with physical deformities to illiterates to reflect a real-world evaluation of the mobile apps,
- To identify the gender-based usability problems in mobile shopping apps,
- To identify which usability factors are more important for users and which are being neglected in design of mobile shopping apps,
- To propose heuristic guidelines in light of the conclusions drawn from detailed usability evaluation.

This paper is organized as follows: In next section, we describe some related work. In experimental section, we describe our experimental settings in detailed manner and then results of the experiments are discussed. In next section to this, we propose heuristic guidelines for the mobile shopping apps.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

2. RELATED WORK

2.1. Usability Evaluation Methods

Heuristic Evaluation is a method to find out the usability issues in design interface by the users. Heuristic evaluation includes small number of evaluator that analyzes the interface and judge according to identified principles of usability (heuristic). Generally heuristic evaluation is complicated can't be done by the one person because one person may not interface all issue as everyone has different mentality and understanding level so everyone interfaces different issue from the same source. This was identified by the experience from various projects that different peoples find different issue of usability of the same source. Therefore, this study concluded that improvement can be possible in effectiveness of given methods through different evaluators (21). When a software company launches the product, before providing it into the market, it should evaluate usability of this product to avoid the possible problems costumers may face. So usability evaluation is essential for the conformation of the product efficiency, effectiveness and satisfaction level of user. There are three widely used methods; Think Aloud (TA), heuristic Evaluation (HE) and Cognitive Walkthrough (CW).

2.2. Think Aloud (TA)

It is commonly used method for usability evaluation. There are various TA methods; users are advised to think aloud at their workplace while there is another method which evaluates the usability of user in lab-based equipment. This study worked on the TA method with the combination of lab-based user testing. In this process users used the products and continuously thinking out loud and their behavior along with their verbal would be notice and recorded by the experimenter in the lab. This method is very good for small number of test user. It provides good quality but user's behavior can be influence due to environment i.e. laboratory.

2.3. Heuristic Evaluation

It is usability inspection method. In this method users when interface the usability problems, these were examined by the evaluators and judged by them with the well-known usability principles. This method is easy to use, cheap and provides fast result analysis. But this method may provide too specific and low priority usability problems which might be not real problem.

2.4. Cognitive Walkthrough

It is the theory-based method. In this method usability evaluation was done by the proper stepwise scenario-based task and focus on the usability problems that effects the learning. This method requires the extensive knowledge of cognitive psychology and through this evaluator can find mismatch between users and designers task conceptualization (22). Molano et al in 2015 conducted a survey to find out the various aspects of measuring usability of mobile app. They mentioned that software is created at

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

all levels of human activities so that these software through the applications fulfill the requirements of users according to the quality criteria. There are variety of methodologies for measuring the usability of desktop application like size screen, touchoperation, voice navigation, probability and use of space type different options. It is interesting in this that they worked on methodologies for analysis and measurements of mobile application usability. They considered several methods and metrics which are currently used for testing the usability of application and focus on the mathematical approach for usability measurement of mobile applications (23).

2.5. Usability In Various Fields:

Mobile commerce applications have been considered necessity and becoming popular among the Arab population for example for performing the online transaction while buying goods. It is necessary to build M-commerce application for usability evaluation and avoiding the usability problems faced by the population like native languages. Ara- bic language brings challenge in design usable for user interface (UI) as set of heuristic evaluating the M-Commerce. In this study heuristics were divided into six categories including Search, findability, visual design, translatability, consistency, adaptability, user control and freedom. Heuristic were developed by the process including three stages, 1st thoroughly studying the literature draft of heuristic were generated, 2nd stage include the review by three experts of HCL and 3rd stage after the expert review final modifies heuristic were used for the final evaluation by usability expert community. There was online survey conducted to collect the data from the community. At the end problem faced by the user were highlighted and marked in comments which were finalized by the proposed heuristic evaluation (16).

The work design evaluate and improve search engine for the Visual impaired (VI) users for efficiently use the search engine for performing activities. Formal concept analysis (FCA) was the conceptual modeling technique used for data analysis. This concept combine work with the interactive navigation called as interactive search engine. These two methods interfere to decrease the time and effort vy the VI use to browser the result of search material. InteractSE was evaluated by using Nielsens heuristic and Web Content Accessibility Guideline (WCAG) by the expert to measure its accessibility and usability. The usability problem was evaluated by using WCAG, these were found to be minor some of the problems might were ignore as either of them was used alone (17).

Student Information system (SIS) offers many functions to sustain academic work- flow. Literature reported that 324 undergraduate students as a user represented various information system usability by using Turkish version of System Usability Scale (SUS-TR). This study resulted in the existed statistically significant relationship between age group of the users and categorized SUS-TR score group. This research concluded that there is a great value of perceived usability variation among the different groups of user, by focusing on these variation and fixing it, can raise the experiences of user usability (18).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

Heuristic evaluation is also commonly used for software usability evaluation like inspection method. Software usability was assessed through traditional heuristic evaluation methods; it was modified and extended so that it can be applied to the medical devices. This software was useful to evaluate the safety of patients while using these devices by identifying the usability problem design. Paper compared the results of the using modified heuristic evaluation method of two 1-channel volumetric infusion pump. This paper reported that pump 1 had more usability problem reported as 192 heuristic violations were categorized for identifying 89 usability problems for Pump 1 as compare to pump 2 with representing 121 violations categorized for 53 usability problem. Pump 1 also reported more severity with more usability issues which were more likely to induce the medical error than pump 2. This paper concluded that heuristic evaluation after modification for medical devices is useful, efficient and low cast method to evaluate the safety of patient by identifying the usability problems and their severities through medical devices (19).

The complexity of medical technology is important criteria to select the new equipment for purchase. Selection of equipment requires the understanding of the features which would provide largest component of the usability level for user. This study con- ducted a survey through distributing questionnaire to investigate the use of medical technology providing components of usability. There were five basic components investigated in this study; easy to learn, efficient to use, easy to remember, difficult to make errors and satisfaction level. They resulted in 30 percent of overall usability with the component “difficult to make errors” while 20 percent of overall usability with each component “easy to learn” efficient to make error” and easy to remember. Only 10 percent of overall usability reported satisfaction level. There were four methods evaluated according to the validity, reliability, cast effectiveness and clarity were; hierarchical task analysis, cognitive walks through, heuristic evaluation and usability tests. From these four methods usability tests were recommended to be primary method to evaluate usability by fulfilling the criteria at hospital level similarly hierarchical task analysis and cognitive walk through also fulfill the same criteria. They concluded that higher level of understanding should be there to choose that which usability evaluation method is most suitable regarding medical technology and design process at hospital (20).

Quality of life (QoL) is the very common variable in health circumstances especially mental health. This can be determined through researches through measuring the specific domain of life covering the human mental health. Digital app helps to advance the QoL in individual through digital accessibility and personalization. This study focused on developing the visual personalized and innovative QoL assessment in people with the following 3 groups of people with severe mental health problem like psychiatric problem, forensic Psychiatry and homeless by observing 59 participants with the development of QoL ME. It comprises of three different stages i.e. brainstorming stage, design stage and usability stage. This Application concluded that usability evaluation revealed from good to excellent. This study recommended that future researches should evaluate the Psychometric quality of QoL Me and investigate either it is useful is practice or not (21).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

3. EXPERIMENTS

3.1. Participants Based Usability Evaluation

The process of testing an application's interface and interactivity, by involving the real time users is called Participation-based usability testing (8). In this method of usability testing, a group of users are assigned with a number of tasks while being watched by an observer. The tasks performed by the user are closely monitored under a close environment, typically a lab with fitted cameras. The purpose of creating this entire scenario is to gather subjective/qualitative and objective/quantitative data that could help in usability evaluation and to figure our problematic areas with respect to the interfaces. Hence, overall user satisfaction is measured through this technique (12).

3.2. Participants Selection

The participant selection process in our case is a complicated process because of the nature of the problem and questions we want answers for. The nature of our problem is such that we have to target a variety of participants for this purpose. Globally, we need following types of participants systematically selected:

(1) Educational Literates

(2) Educational Illiterates

(3) Digitally literates

(4) Digitally illiterates

(5) Participants with physical deformities (hands)

It is also mandatory that we choose equal number numbers of participants from both genders for all these categories of participants for a fair evaluation. We need to be very careful while choosing the number of total participants because a smaller number of participants may not represent opinion of the whole population while an unnecessarily large number will increase our burden. However, it is mandatory that we choose enough number of participants from each category of participants as listed above. Referring to the literature for a benchmark in number of selected participants, we decide to select 6 participants for each category which makes total number of 30 participants.

3.2.1. Gender Consideration

We decided to keep 50% gender distribution among all categories i.e. we will choose 3 male participants and 3 female participants from each category.

3.2.2. Age Binning

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

We set 3 age boundaries for each category i.e. we choose equal number of participants from each age boundary for each category.

- Group A (20 ≤ Age ≤ 35)
- Group B (36 ≤ Age ≤ 50)
- Group C (51 ≤ Age ≤ 65)

We select Total of 30 people are used to perform the usability testing. The reason for selecting this number of participants is based on previous papers' references. For instance in '(7) total of 5 people were used to perform usability testing. Similarly, in the paper (6) 8 people were chosen to perform the usability analysis. Based on the previous work we chose 6 participants from 5 different categories of participants and hence making it total 30 participants. For a fair gender-wise comparison, we keep 50 percent females in each category. Furthermore, participants of age greater than 30 and less than 55 were considered, so to make sure the analysis could be performed accurately.

3.2.3. Selection of Mobile Devices

For a fair selection of mobile devices, we define a set of features comprising of features most commonly found in popular mobile devices.

- Weight
- Screen Size
- RAM
- Processor

We select one Android device and one iPhone meeting nearly these criteria. Selecting phones meeting this minimum criterion will make sure we have almost similar types of devices for a fair evaluation of applications on both platforms.

3.2.4. Mobile Application Selection

For usability analysis of mobile applications, we decide to select 20 percent of the total applications as selected for systematic review of mobile commerce apps. The decision of selecting only 10 percent apps is not a random decision but, in the literature, there exists such precedence (13; 14). However, we make sure that we select those apps that contain all relevant functions to represent their population in a true sense. From the collected data of 180 apps (120 apps in android, 60 apps in iOS), we select total 36 apps. Out of 36 apps, 24 apps are selected from android and 12 apps are selected from iOS i.e., 20 percent of total apps from each platform are selected.

- Apps are relevant i.e., they are truly a commerce application and used by general audience and not by a specific group of users,
- The major language of the Apps is English,
- Apps contain maximum number of functions as selected in systematic review process.

Based on the criteria defined above, we choose following apps for the usability review as shown in table 1.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

3.2.5. Test Material Preparation

- **Pre-Test Questionnaire**

A pre-test questionnaire generally consists of question related to participant. It may contain some personal questions about the participant (avoiding privacy) along with some questions asking about his/her personal experience with smart phones and smart phone apps. Each pre-test questionnaire is made part of post-test questionnaire.

- **Post-Test Questionnaire**

The posttest questionnaire is based on usability criteria. We create this set of usability criteria which is based on the study of existing literature (15) and our observation of the online shopping applications. This usability criteria, description of its attributes, related prepared questions and metrics to be used for answering these questions are presented in table 2 (15).

Table 1: Usability Characteristics, Related Questions and Metrics

Usability Characteristics	Description of Attribute	Questions	Metrics
Mobile shopping Features	These attributes will generally analyze the important features available in apps.	The app facilitates the easy registration process.	Dichotomous Scale (Yes, No)
		The app provides the effective product searching	Dichotomous Scale (Yes, No)
		The app provides a good product view & review	Dichotomous Scale (Yes, No)
		The app provides enough information about deals and discounts	5-point Liker Scale
		The app gives the facility to communicate with buyers	Dichotomous Scale (Yes, No)
		The app provides an effective way to place and finalize orders	5-point Liker Scale
		The app automatically filled the address by tracing the location	Dichotomous Scale (Yes, No)
		The app gives the facility to track, return & refund the product	
Effectiveness	The attributes of usability checked in this area will be how much application is interactive, the navigation feature and multimedia usage, and availability of various	Is it easy to interact with the UI?	Count the number of mistakes done to complete a task?
		Are options easy? To use for customers?	Number of mistakes in selecting options
		Is it easy for customers to navigate? Across the UI?	Number of mistakes during navigation
		The navigation keys are well understandable?	Rating scale for navigation

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

	interaction types with apps.	Does UI indicate easy scrolling if a lot of information is present?	Rating scale for long contents information
		Does UI provide easy main menu for navigation?	Success/Failure rate to use main menu
		Is the main menu or Home Page button available on all subsequent screens?	Dichotomous Scale (Yes, No)
		Is the screen positioning of UI effective for customers?	Rating scale for screen orientation
		The navigation between screens presents both up and down side?	Dichotomous Scale (Yes, No)
		Does UI provide a visual display to show the loading process?	Rating scale for loading application
		Does application provide audio instructions?	Rating scale for voice instruction
Usability	This attribute of usability checks various input/ output methods.	Does the input method provided on the interface is easy to understand and easy to use for customers?	Number of mistakes made to enter an input
		Is the output easily understandable	Rating scale for ease to understand output
Efficiency	This usability attribute measures the time and effort required to interact with apps and performing individual tasks.	Is the application takes extended load time?	Measuring load time of an application
		Does the App crash, hang or freeze.	Dichotomous Scale(Yes, No)
		Does time given to the user to respond is appropriate?	Measuring time to respond on a scale
		How much time is required in completing individual tasks?	Measuring task completion time.
		How much effort is required in completing individual tasks?	Measuring task effort on a rating scale.
		Can the App be easy to use for people with various physical deformity issues?	Dichotomous Scale(Yes, No)
		Error message is easily understandable in case of wrong Input.	5-point Liker Scale
Learnability	This UI attribute explores the help and support function as well as how much apps has learning potential to make the app easy to use in subsequent times	Does appropriate help is provided in UI where needed?	Measuring through Rating scale for help's usefulness
		Is the app contains help icon which is visible and understandable?	Measuring through rating scale for the icon visibility
		Are the icon used in the UI relates to the task?	Measuring through rating scale

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

		Can the customers recognize the functions and their corresponding actions?	Measuring number of icons unrecognized at the starting attempt.
		Is the UI using familiarized terms and easy language?	Measuring easy term usage through rating scale.
		Is proper information provided various functions?	Use Rating scale to get the result.
		Is UI co-relates with other apps and hence easy to learn?	Use Rating scale to get the result.
		Does proper and multiple presentation is provided to display the results?	Use Rating scale to get the result.
		Does app provide easy ways to step back to previous activity?	Number of success vs. failure times stepping back was possible.
Operability	This UI attribute explores the ability to personalize or customize an application.	Does the app offer the ability to change color of products?	Measuring Success rate
		Does the font used in the app is appropriate and readable?	Use Rating scale to get the result.
		Does the app interface is no too cluttered?	Yes , No
		Does the app provide options to search by voice?	Measuring Success rate
		Does the main menu button easily operable?	Measuring Success rate
		Does the main menu contain link to all useful tasks?	Use Rating scale to get the result.
		Does the size of the icons are set appropriately to be operable easily?	Use Rating scale to get the result.
		Does the app provide easy access to the home screen of the mobile?	Measuring Success rate
Satisfaction	This attributes of UI will check the engagement level of customers and also how attractive a UI is?	Does the UI interface is appealing to the customers?	Use Rating scale to get the result.
		Does the UI provide features to engage the customers?	Use Rating scale to get the result.
		Is the overall layout effective to be used for long time?	Use Rating scale to get the result.
		Are all the screens consistent?	Use Rating scale to get the result.
		Are the users happy with the App layout?	Use Rating scale to get the result.
		Are the users willing to use the app multiple/regular basis?	Rating scale for engagement

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

We prepare two types of posttest questionnaire from 2 i.e., one for the participants (Questionnaire A) while the other for the moderator of the test (Questionnaire B). The idea behind distributing questions among participants and moderator is that there are some questions where moderator is in a better position to answer. We do not want participants to get overburdened by the questions.

We prepare a list of tasks to be performed by participants for each application to be tested (see table 3). Each main task consists of some sub-tasks to be performed. The purpose of preparing this list of tasks is to make participants try and test each possible feature of the application. Moderator keenly observes the test participants during task-based evaluation of mobile apps and fills out the task relevant questions (added to Questionnaire B) as shown in screen shots of Questionnaire B. Mobile video cameras are used for recording usability testing process for moderators' observations.

Table 2: List of Tasks for Task-Based Evaluation

Check for function Availability	Check for product search function Check for products view and review function Check for discount/offer function Check for messaging function Check payment options function Check for order tracking function Check for return & refund function
Check for function Interactivity	Time required to search a function Time required to complete a function Mistakes did to complete a function
Navigation Check	Availability of Home Screen button Check Main menu availability Check scrolling available for long contents Test main menu reachability from within the task
Help Check	Availability of help within a task Check for help contents understandability Meaningfulness of help icon Tutorial available for the app
Feedback	Check if feedback is provided for a task Check if feedback is understandable
Error Messages	Check for the error message available where needed
Customizing UI	Change the app color Change input/output method Change font size Change the app language

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

Table 3: List of Selected Mobile Shopping Apps and Their Categories

No.	App Name	Category	Platform
1	Alibaba.com	All in one	iOS
2	Daraz online shopping	All in one	Android
3	Amazon shopping	All in one	Android
4	Ali express	All in one	Android
5	ebay	All in one	Android
6	11 Street	All in one	iOS
7	Banggood	All in one	Android
8	Econox	All in one	iOS
9	Naheed.pk	All in one	Android
10	Negative Apperal	fashion shopping	iOS
11	Namshi	fashion shopping	Android
12	Clicky online shopping	fashion shopping	Android
13	Elo	fashion shopping	Android
14	Patpat	kids and baby clothing	fashion shopping iOS
15	Meesho	fashion shopping	Android
16	Karma	fashion shopping	Android
17	Unze	fashion shopping	iOS
18	Fashion Nove	fashion shopping	Android
19	airlift	Grocery and electronics	Android
20	Maf Carrefour	Grocery and electronics	iOS
21	Metro Online	Grocery and electronics	Android
22	Talabat	Grocery and electronics	iOS
23	Grocer app	Grocery and electronics	Android
24	Noon	Grocery and electronics	iOS
25	Alfatah	Online Grocery and electronics	Android
26	Gahak	Grocery and electronics	Android
27	Krave Mart	Grocery and Electronics	Android
28	Outfitters	Clothes brand	Android
29	Zara	Clothes brand	iOS
30	Limelight	Clothes brand	Android
31	Khaadi	Clothes brand	iOS
32	Ethinic by Outfitters	Clothes brand	iOS

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

33	Beechtree	Clothes and brand	Android
34	Bonanza Satrangi	Clothes brand	Android
35	J.	Clothes brand	Android
36	Maria .B	Clothes brand	Android

4. USABILITY EVALUATION PROCESS

Usability evaluation is a time-taking process with several tests to be conducted on several application using selected participants. In our case, we have total 1080 (36 x 30) tests to be performed. We have to test 36 applications using 30 selected participants. Keeping in mind the easiness of the participants and time it takes to conduct a test, we plan to conduct tests in several sessions. After detailed discussion with participants, we decide to conduct all experiments in four sessions for each participant. The details of each session are given below in table.

It is unanimously decided with participants to complete all 4 sessions within two days. We decide to allocate morning and evening timings for separate sessions. Even few participants expressed their willingness to complete test within a day but we did not want to let their fatigue overcome during usability testing process and hence we requested for 4 separate sessions. We also made sure that participants are invited on those days when they are totally free. We keep a participant engaged even when he/she is not performing usability test to keep him thinking in same space. We take following precautions for all tests.

- All participants are given a briefing about usability testing before the test starts.
- All tests are conducted in a controlled environment where proper lighting arrangements are present along with WIFI facility in case it is needed.
- All mobile apps are installed on corresponding mobile phones already. Participants are made familiar with mobile phones to be used for testing. They are informed about phone's virtual keyboards key patterns, working of back and home buttons, etc.
- Time of start and end of each test is noted down,
- We made sure that no participant has used any of the selected apps already to avoid any learning effects on the results,
- All participants sign a "letter of consent" before their first session starts.

4.1. Usability Data Analysis

4.1.1. Gender-Wise Evaluations

In this sub-section, we compare the gender-wise evaluation of mobile commerce apps. For this purpose, we provide three types of analysis in the form of line graphs.

- Gender-wise Overall Comparison: In this line graph (figure 4), we compare overall evaluation by males and females participants for all the selected mobile commerce apps.
- Gender-wise Completion time Comparison: In this graph (figure 4), we compare task

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

completion time comparison for all selected tasks on all mobile commerce apps.

- Gender-wise Committed Mistakes Comparison: In this line graph (figure 4), we compare the mistakes committed by males and females participants.

A careful analysis of these graphs (figure 4) reveals that we cannot see significant differences between evaluations of males and females for mobile commerce apps. While we can see gaps between lines over all list of applications denoting little differences on Likert scale but we can see synchronous crests and troughs between them showing similar evaluations. However, to draw a significant conclusion from the given graphs, one can conclude that female's participants are more prone to mistakes during task completion.

4.1.2. Participant Category wise Comparison

In this sub-section, we will compare the evaluations of different categories of participants who took part in usability evaluations of mobile commerce apps. Following categories of participants took part in evaluations:

- Educated participants
- Digitally literates
- Digitally illiterates
- Illiterates
- Physically disabled (e.g., thick thumb, small fingers)

We provide several analyses on the evaluations provided by these categories in the form of line graphs.

- Overall comparison of evaluations by all participants categories (overall usability as well as according to different usability factors)
- Comparison of task completion time among participants categories,
- Comparison of committed mistakes by all participants categories,
- Comparison of overall evaluations by participants age categories
- Comparison of committed mistakes by participants age categories

In overall evaluation of all apps set with respect to different categories of participants (see figure 6), slight variations among participant categories have been observed. It is obvious from the graph shown in figure that illiterates have given the lowest ratings on Likert scale with participants with physical deformities follow them in ratings. However, it is interesting to note that there is high level of coherence in line patterns which shows that evaluations have been very fair in their reliability. We can observe peaks on apps like Daraz, meesho, J. etc while lesser ratings are also reflected in graphs for apps like Econox, karma, unze, gahak etc.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

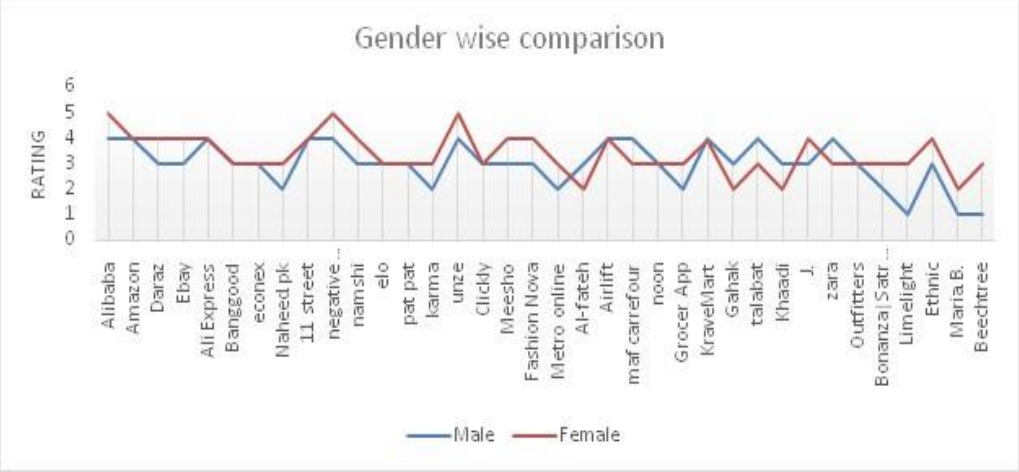
In this set of graphs, we highlight the evaluations by category of participants for different usability factors. We combine the evaluations of participants for questions relating to these usability factors and display it on graph

The overall comparison of usability evaluation among participant categories shows that ratings given by educated participants are relatively positive than participants of other categories. Obviously, the apps are not well designed for illiterates (either education wise or digitally) as well as for people with tactile deformities. The overall design of all apps has been rejected by all other types of participants. Similarly, the usability factor-wise evaluations also show the same trends.

The participant task completion time comparisons show that illiterates and digitally illiterate participants take more time to complete the task as the apps are not affordable for them.

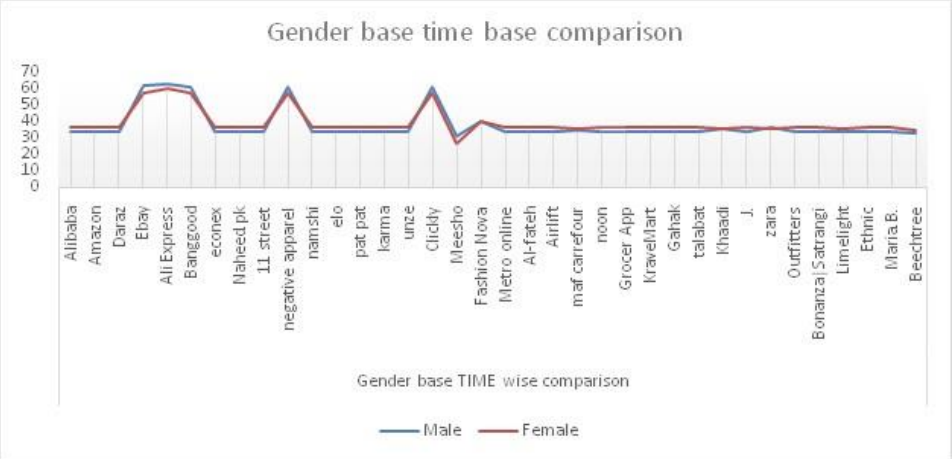
Figure 1: Gender-wise comparison of participant's evaluations

a) Gender-wise Overall Comparison



b) Gender-wise Completion time Comparison

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ



c) Gender-wise Committed mistakes Comparison

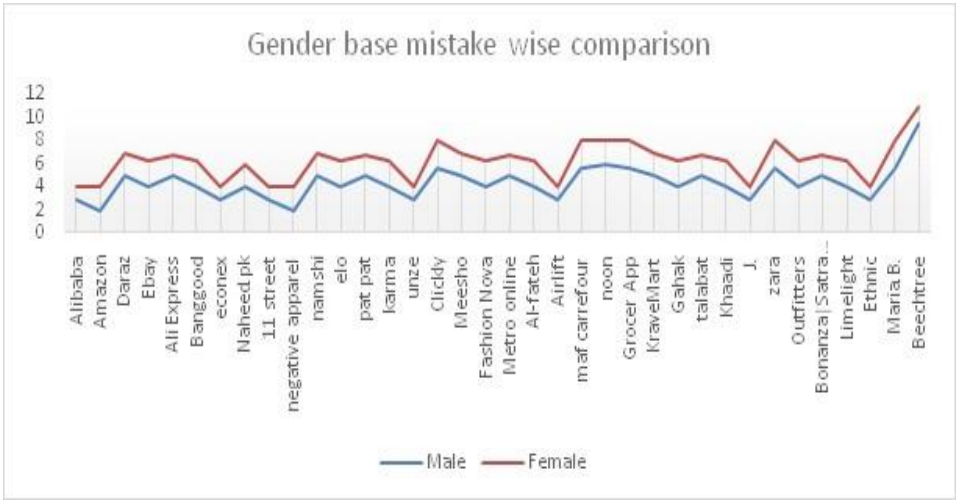
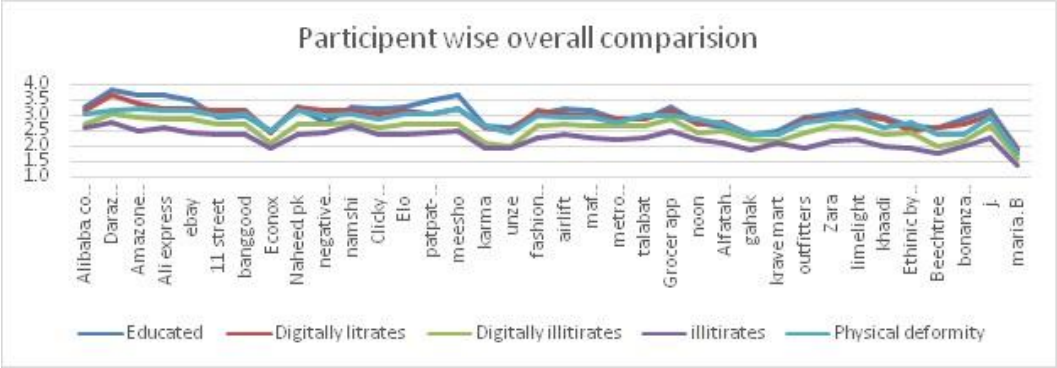


Figure 2: Overall comparison of evaluations by all participants' categories

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ



However, the apps that have an Urdu interface and are relatively more iconic (i.e. more pictorial) have been rated higher by these categories of participants. The participants with physical deformities take more time because they often select wrong icon due to cluttered interface. This shows that a poorly designed inter- face results in time wastage on behalf of clients which in-turn can result in frustration and conclusively leaving the usage of mobile app. A poor interface takes more time because participants commit more mistakes while completion of a task as shown in figure above.

We also highlight the evaluations among different age groups. For this purpose, we form 3 age categories of all participants i.e.

- Group 1 (age between 20 and 35)
- Group 2 (age between 36 and 50)
- Group 3 (age between 51 and 65)

Difficult to operate with mobile commerce apps for several usability problems while there was a mixed trend among younger category of participants. This shows that mobile apps needs to take special design measures in their user interface because existing ones are not able to meet old client's needs. The operability of mobile apps needs to be improved especially in such types of applications which targets diverse age categories and mobile commerce apps are one of them.

4.1.1. The Most Complicated Tasks

According to the graph the function complexity of F1 (registration function), F4(deals and discounts function), and F6 (payment option and address filling Takes more effort time, and mistakes because it had a lot of information to be filled and deals and discount function is not properly explained that is how to apply them on your final orders.

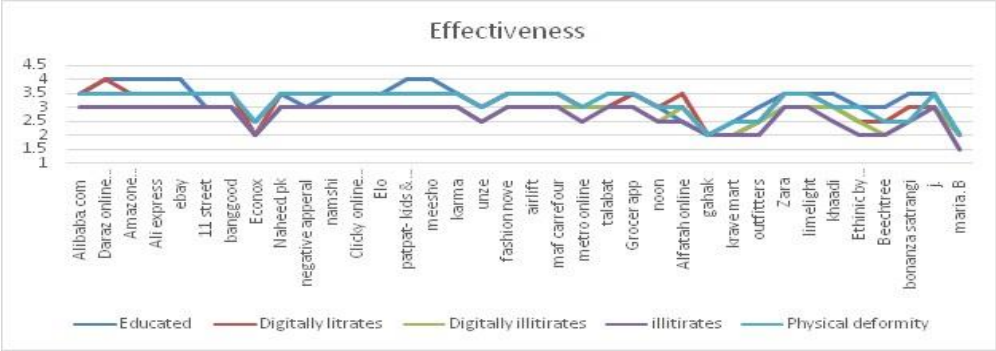
4.1.2. Applications with the best Evaluations

In this sub-section, we analyze evaluations for app and apps categories. For this purpose, we use bar graphs. We highlight the evaluations for the followings:

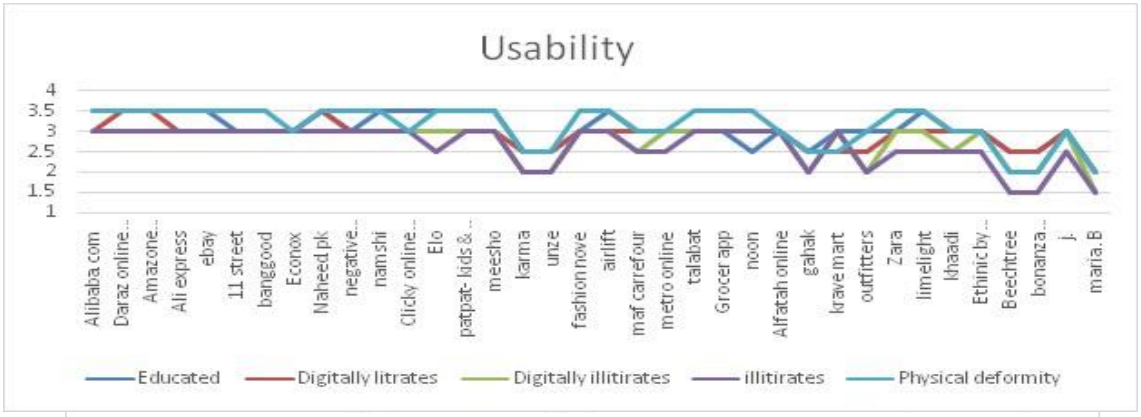
Figure 3: Participants Category-wise results – I

Usability Factor Vs Participant Category: Effectiveness

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

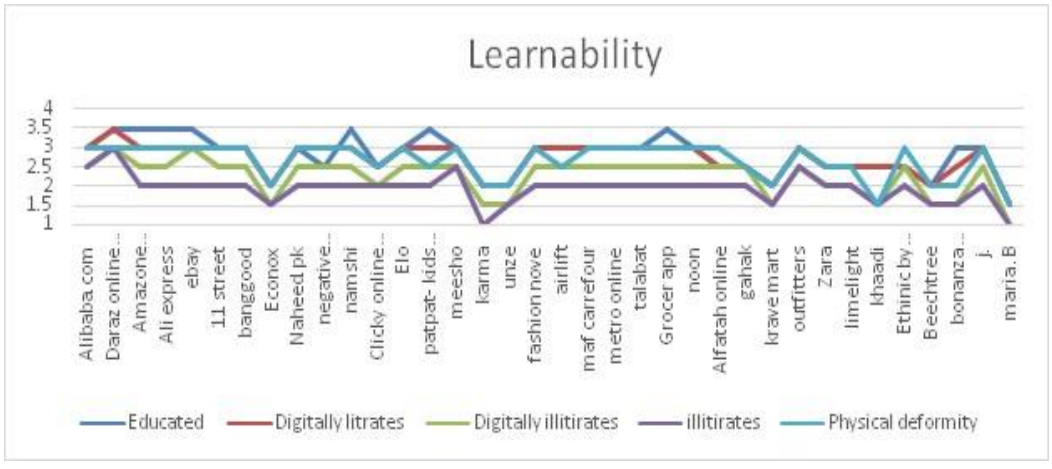


Usability Factor Vs Participant Category: Usability

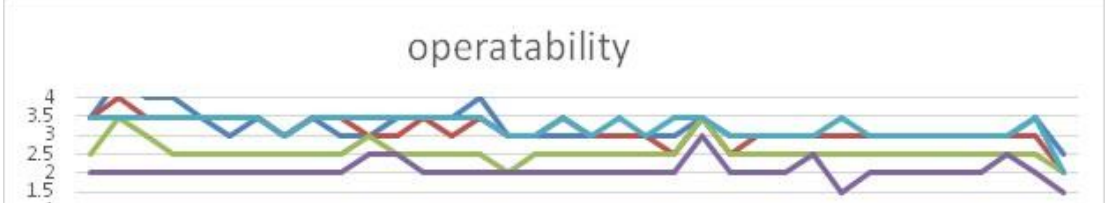


Usability Factor Vs Participant Category: Efficiency

Usability Factor Vs Participant Category: Learn ability

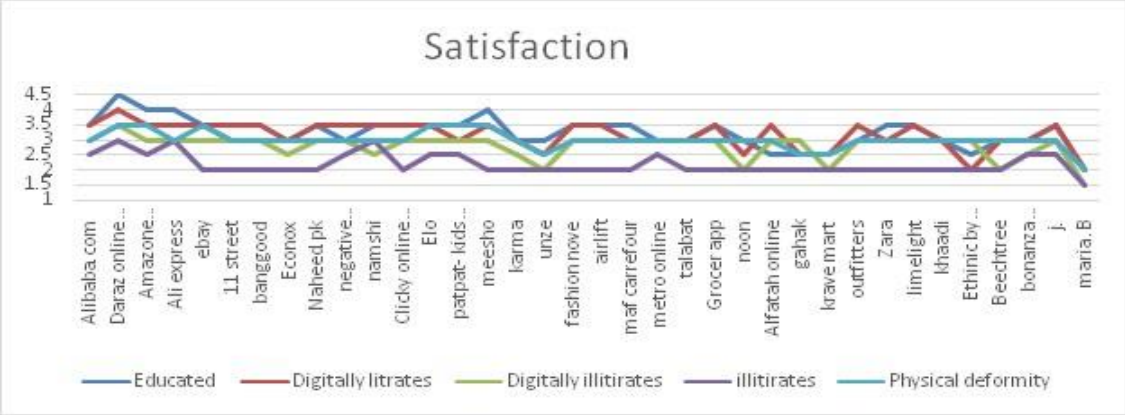


Usability Factor Vs Participant Category: Operability



Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

Usability Factor Vs Participant Category: Satisfaction



Usability Factor Vs Participant Category: Navigation

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

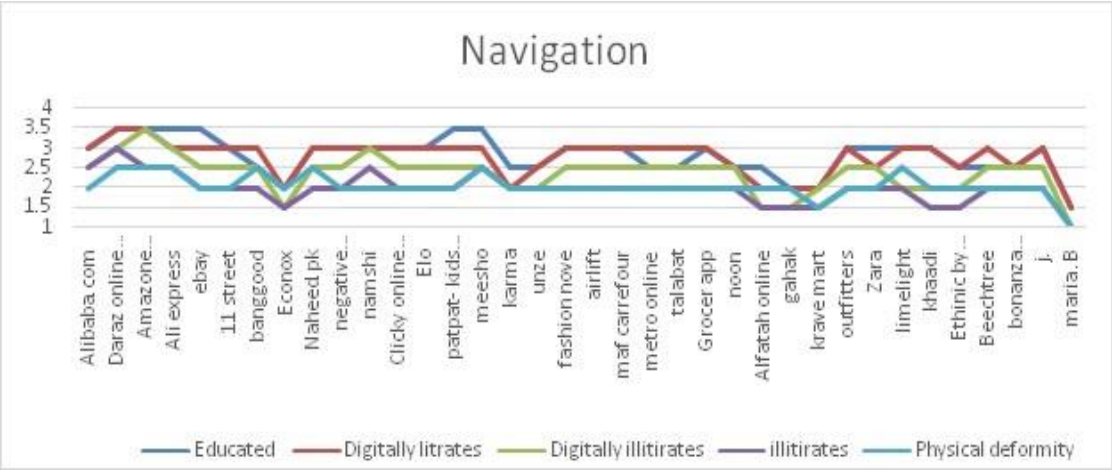
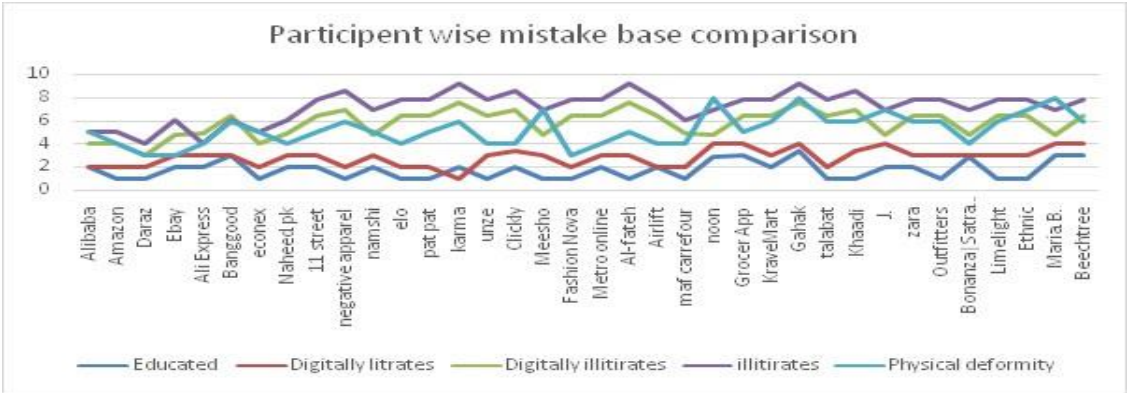


Figure 4: Participants Category-wise results - II

Comparison of committed mistakes by all participants' categories



Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

Comparison of task completion time among participant's categories

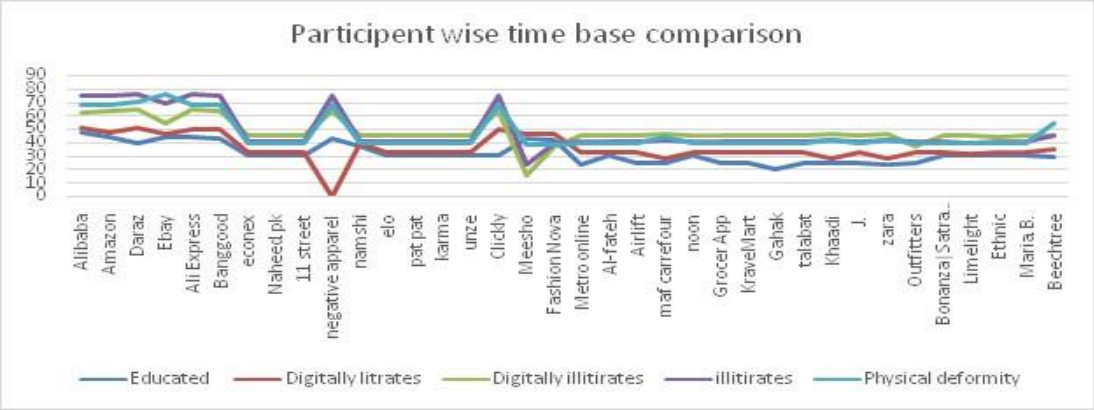
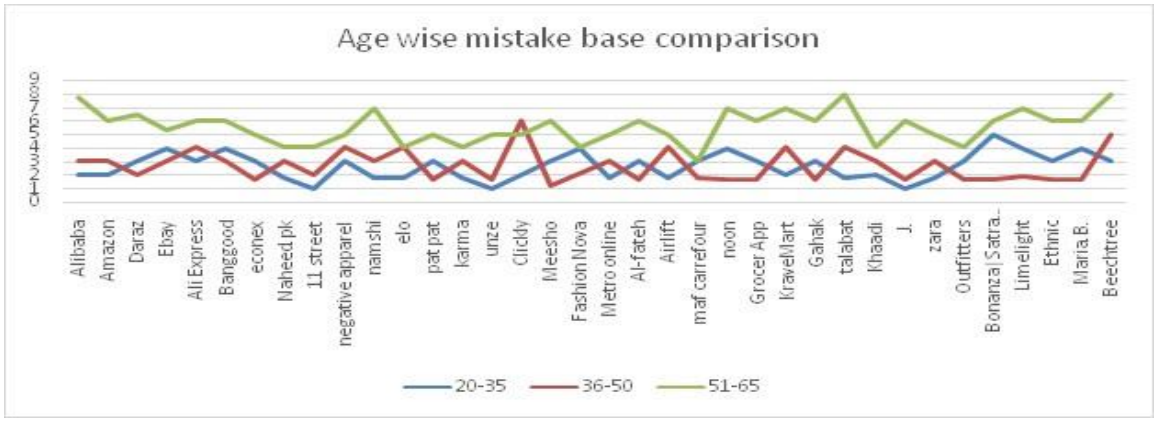
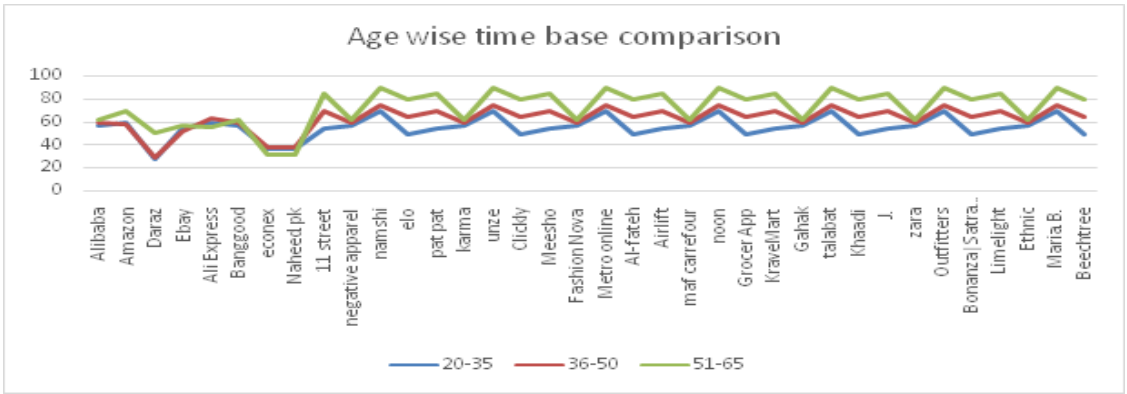


Figure 5:Age-wise comparison of participant's evaluations
Comparison of overall evaluations by participants age categories



Comparison of committed mistakes by participants age categories



Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

Figure 6: Complexity of different functions

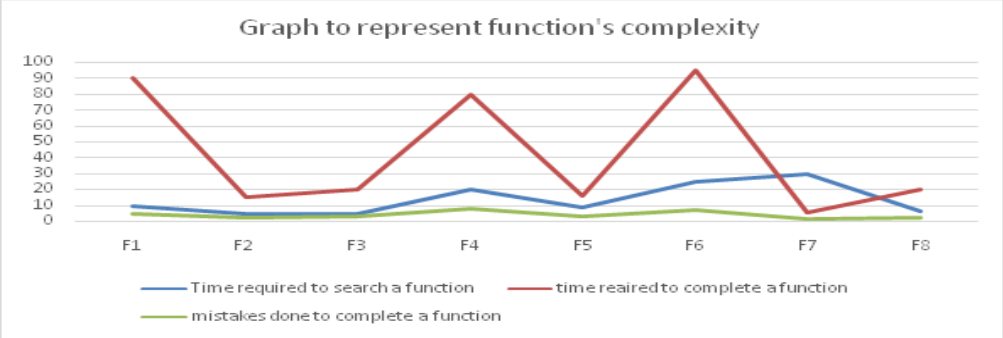
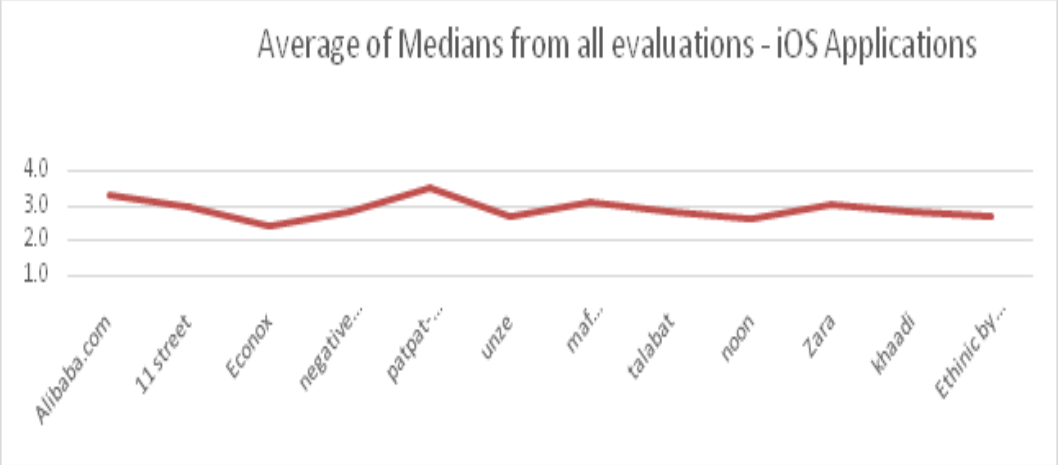
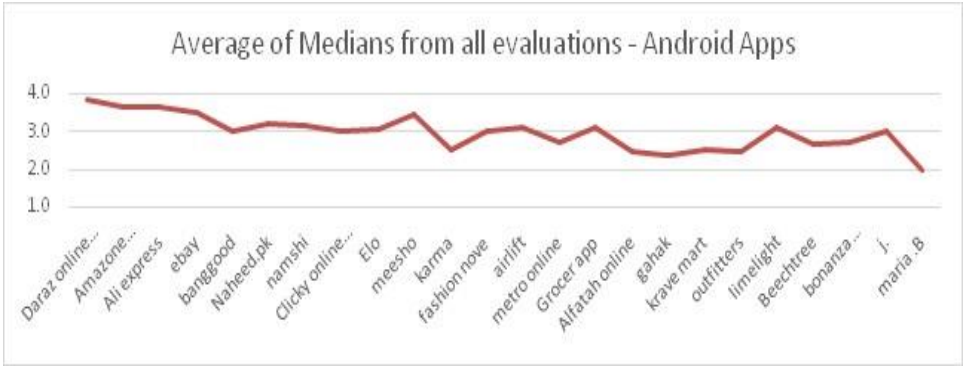


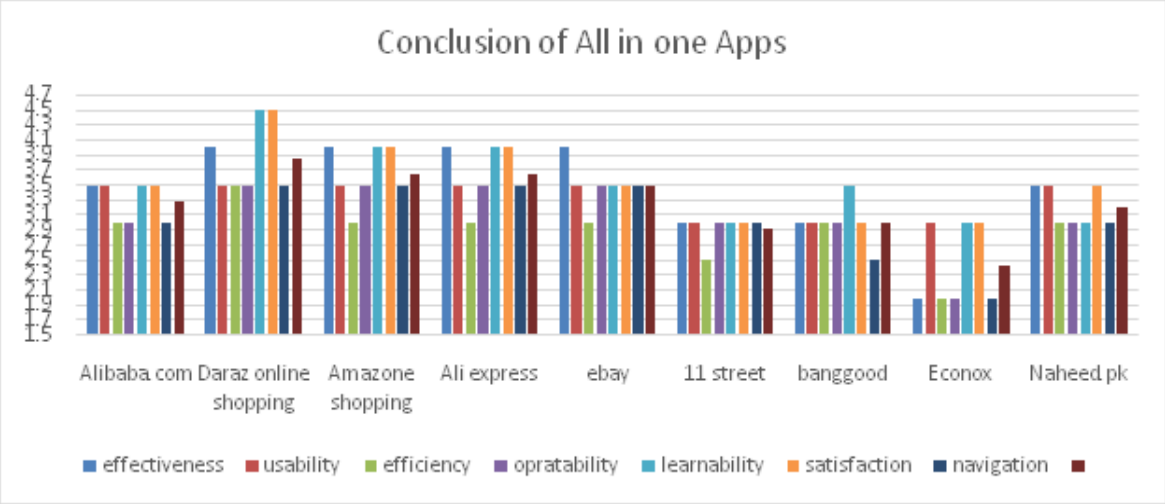
Figure 7: Platform-based evaluations
Average of medians from all evaluations – iOS apps



Average of medians from all evaluations – Android apps



Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ



Graph showing evaluations by all participants for different usability factors

- Overall evaluations by all participants for individual apps (Android and iOS apps),
- Overall evaluations by all participants for different app categories.

It is seen from the all-in-one app conclusion that in Pakistan Daraz online app gets the higher satisfaction and learn ability due to its iconic and Urdu language presence of interface other apps also have effective product searching etc but don't have urdu added in their languages till now

4.2. Identified Issues

The main objective of conducting usability testing was to identify the problems in user interface of mobile shopping apps available for Android and iOS platform. Wehave conducted a very extensive set of usability experiments and conclusively following problems have been identified in mobile commerce apps:

- Access to mental health problems symptoms is problematic.
- No visual display for loading process,
- Output is not easily understandable,
- Not easily understandable error messages in case of wrong input,
- No appropriate help provided,
- The help icon provided is not visible and understandable,
- Not enough information about functions being provided,
- No proper and multiple presentations provided for results,
- It is hard to use the app for a long time.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

5. CONCLUSION

In this paper, we performed a comprehensive usability evaluation of mobile commerce apps. In the literature, we cannot find any work which has performed usability testing on large scale to find usability problems in mobile commerce apps (to the best of our knowledge). A total of 1080 experiments have been performed for 36 apps using total 30 participants from various categories. We discuss the results in detail and as a conclusion; we identify some usability problems as reported.

References

1. Sharma, G., & Baoku, L. (2013). Customer satisfaction in Web 2.0 and information technology development. Information technology & people.
2. Arora, N. K. (2003). Interacting with cancer patients: the significance of physicians communication behavior. Social science &medicine, 57(5), 791-806.
3. Venkatesh, V., Ramesh, V., & Massey, A. P. (2003). Understanding usability in mobile commerce. Communications of theACM, 46(12),53-56.
4. Ahmad,N., Rextin,A., & Kulsoom, U.E.(2018).Perspectives on usability guidelines for smart phone applications: An empirical investigation and systematic literature review.InformationandSoftwareTechnology,94,130-149.
5. Silvis, I. M., Bothma, T. J., & de Beer, K. J. (2018). Evaluating the usability of the information architecture of academic library websites. Library hitech.
6. Masood, M., & Musman,A. (2015). The usability and its influence of an e-learning system on student participation. Procedia-Social and Behavioral Sciences, 197, 2325-2330.
7. Dalkıran,Ö,Aker,İ.,Öztemiz,S., Taşkın,Z., & Tunç,S.K.(2014). Usability testing of digital libraries: the experience of ePrints. Procedia-Socialand Behavioral Sciences, 147,535-543.
8. Albert, B.&Lewis,J.R.(2022).Welcome to the journal of user experience. Journal of Usability Studies, 17(3), 85-88.
9. Wasike, J.M., &Wabwile, P.N. (2022).Cloud Computing Usability Model for Quality University Education.5th KyU International Conference.
10. Majrashi,K.,Hamilton,M.,Uitdenbogerd,A.L.,&Al-Megren,S.(2020).Cross-platform usability model evaluation. Multimodal Technologies and Interaction, 4(4), 80.
11. Majrashi,K.,Hamilton,M.,Uitdenbogerd,A.L.,&Al-Megren,S.(2020).Cross-platform usability model evaluation. Multimodal Technologies and Interaction, 4(4),80.
12. Redish,J.,& Barnum,C.(2011).Overlap, influence, intertwining: The interplay of UX and technical communication. Journal of Usability Studies, 6(3), 90-101.
13. Missen, M. M. S., Javed, A., Asmat, H., Nosheen, M., Coustaty, M., Salamat, N., &Prasath,V.S.(2019).Systematic review and usability evaluation of writing mobile apps for children. New Review of Hypermedia and Multimedia, 25(3), 137-160.
14. Arnhold, M., Quade, M., & Kirch, W. (2014). Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. Journal of medical Internet research, 16(4), e2968.
15. Tahir, R., & Arif, F. (2015). A measurement model based on usability metrics for mobile learning user

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 55 Issue: 08:2022
DOI10.17605/OSF.IO/WHEGJ

interface for children. The International Journal of E-Learning and Educational Technologies in the Digital Media, 1(1),16-31.