Usability Evaluation of Mobile Health Application from AI Perspective in Rural Areas of Pakistan

https://doi.org/10.3991/ijim.v13i11.11513

Abdul Samad Dahri (✉), Ahmed Al-Athwari
Universiti Utara Malaysia, Kedah, Malaysia
Dahriabdulsamad@gmail.com

Azham Hussain
UUM School of Computing, Kedah, Malaysia

Abstract—Purpose: This study endorses AI enabled Mobile Health application and investigates usability evaluation of the Mobile Health application by patients’ task performance evaluation and satisfaction.

Materials and Methods: International Organization for Standards (ISO) 9241-11 standard metrics were used and 15 patients performed tasks on task success rate, errors, efficiency (time spent), satisfaction (SUS scale).

Results: Getting registered was a top easy task while finding a relevant doctor was the most difficult task for users. The satisfaction scores by SUS suggest good rather excellent application user experience. Male were successful task achievers, while educational level and mobile know-how influence the usability scores in terms of time consumed, task errors occurred, and task completed.

Conclusion: Methods used in this study suggest future research from different contexts. Using ISO 9241-11 usability standards, the SUS instrument for satisfaction, and measuring user characteristics influence performance and can provide considerable Mobile Health design.

Keywords—Healthcare, Mobile Health (mHealth), Usability, Artificial intelligence, Developing countries.

1 Introduction

1.1 Healthcare-Related Developing Countries Issues

Healthcare not only provides economic expansion potential but also the basic needs of the country (Mahmud & Parkhurst, 2007). Over a decade global healthcare grabbed attention assembling institutions and organizations to deliver better healthcare services that meet population needs in developing nations (Mills, 2014). The WHO recommends 5% of their GDP threshold to invest on healthcare to achieve set targets (William, 2003). However, this remains a dream for developing nations particularly, African and Asian countries which spend less than 5% of GDP on healthcare (Cassels & Janovsky, 1998). According to UN (2018) reported about 90% of the global...
population lives in rural areas, and this number will reach its peak by 2020. Where, life expectancy is worse, poverty, limited access to quality healthcare facilities, lack of trained healthcare workers, transport difficulty and so on, contributing to low quality of healthcare among rural population (Strasser, Kam, & Regalado, 2016). Besides that, skyrocketing costs, high price drugs, and hospital-acquired infection, and failure of care delivery leads to adverse healthcare events (Neill, 2013). Therefore, the immediate innovative intervention is needed in rural areas of developing countries.

For the past few years from a massive computer system to low-cost tablet, technology has decentralized computational capabilities and modification in healthcare system architect (Lanzola, Gatti, Falasconi, & Stefanelli, 1999). Moreover, in the 20th century, doctors operate with rare and expensive devices that allow ease with the accurate characterization of disease, such as heavy technology and specialist providers (Koop, Mosher, Kun, Geiling, Grigg, Long, & Rosen, 2008). These advances heavily initiated the inclusion of Artificial Intelligence (AI) agents to enhance the predictiveness in healthcare workflow (Bui, 2000).

In healthcare transaction flow between doctor and patient, where, the advantages of AI have broadly triumphed in the medical literature (Jha & Topol, 2016). An AI system uses sophisticated algorithm to ‘Learn’ and extracts useful information from a large patient population to assist in making real-time inferences for health outcome prediction (Neill, 2013). Moreover, the medical information volume is piling twice in every three years and the estimated reading work by a physician is 29 hours to remain up-to-date completely which is not possible (Curioni-Fontecedro, 2017). This enforces the use of AI techniques to analyse data into information to improve quality and lower the cost of patient care (Neill, 2013) to assist in the clinical practices.

There are more than 97000 AI enabled mobile healthcare applications (mHa) available on Google Play and Apple’s App which would be downloaded by 500 million people till 2015 (Jahn & Houck, 2013-2017) while, 50% of these apps were estimated to be downloaded on smartphones by 2017 (Siltala, 2013). This phenomenon has turned smartphones into medical kits for real-time health monitoring of patient’s activities, early predictability, disease screening, improved medication adherence (Alemdar & Ersoy, 2010) by medical professionals and reduce diagnostic errors that are inevitable in human clinical practices (Pearson, 2011).

Advantageously, over 85% of global population is under umbrella of commercial wireless signals (WHO, 2011) and 80% of them use the internet on smartphones (Dave, 2018), and developing countries like Pakistan has mobile user base above 90% with 2G internet coverage (PTA annual report 2014-2015). Keeping this impact, a satellite-based e-medicine initiative with USA training support to 45 doctor/nurses in Sindh with SUPARCO, the space agency of Pakistan, was launched, limited to Karachi and Shikarpur (Malik, 2007). Though all these initiatives ended in ashes yet, penetration of cheaper access to mobile technology is new breath to healthcare system in Pakistan through mHealth apps for cost-efficient, effective, and quality healthcare services. Therefore, by AI-powered mHealth apps holds the future for efficient quality healthcare services, particularly, in remote areas of developing countries like Pakistan.
The past research comprises heavily of a useful literature review or surveys rather than ground studies compared to its western counterparts. Therefore, there is a dire need to conduct field investigations on mHealth app usability (Ozdalga, Ozdalga, & Ahuja, 2012), as a higher percentage of studies lack empirical evidence of field validation for applications (Insfran & Fernandez, 2008). To serve the field-testing urge, mobile usability tests are best suited to best understand the usage of smartphone technology.

Accordingly, mobile application usability is considered a quality feature that indicates how this product enables users to learn and use without difficulty and actual performance of the application. Thus, usability evaluation has become vital for the smartphones as well to prevent the application being difficult to use (Coursaris & Kim, 2011) and indeed it determines the success of that application (Baharuddin, Singh, & Razali, 2013). Usability, in other words, is the facility with which users can use a technology artefact to achieve a specific goal (Coursaris & Kim, 2011) with the key feature are efficiency, effectiveness, and satisfaction, according to ISO 9241–11.

Importantly, despite the mHealth system popularity, 95% of applications are not tested (Furlow, 2012). Therefore, prior to the trial of mHealth technology, it is necessary for designers to consider the usability of these technologies (Brown III, Yen, Rojas, & Schnall, 2013). Surprisingly, most of the advancements are exclusive of the patient’s interface (Heathfield & Wyatt, 1993). Where, often the patient complaint is rectified rather than patient treatment (Coiera, 1997). Therefore, patients’ involvement is important to evaluate and attempt to improve healthcare quality (Kim, Trace, Meyers, & Evens, 1997) including greater patient satisfaction, increased adherence to treatment, and positive treatment outcomes (Tennstedt, 2000). As a result, further investigation is needed to ensure the appropriate designing of mobile healthcare technologies before considering them for health interventions (Wolf, Moreau, Akilov, Patton, English, & Ho et al., 2013).

1.2 Measuring Usability

ISO 9241-11 provides quantitative usability design (see fig. 1) that outlines metrics for the user with steps by user-centred-design (UCD) which provides user hands-on procedure. Typical task performance is most common usability test to measures adequate effectiveness: to what extent the user achieves desired goals, efficiency: the level of effort and resource usage put in by the user in relation to accuracy and completeness, and satisfaction: association or discontent experienced by the user while performing the task. at an acceptable usability level (ISO, 1998, Jaja, Pares-Avila, & Wolpin, 2010).
Effectiveness and efficiency are measured by task completion and counting numbers of errors in an attempt to interact with the application. While efficiency is measured by efforts and resource put by the user deemed to achieve the task. Whereas, satisfaction is measured objectively with available instruments. Such as the System Usability Scale (SUS). Developed and designed by Brooke, the SUS encompasses 10-items on 0-4 range Likert scale (Brooke, 1996), typically administered after user application interaction to record the experience. This study used SUS for validity, reliability, and sensitivity scores ranging from 0-100. The score of 50 or below is poor, above 70 is considered as good, and a score of 85 or above is excellent usability score (Bangor, Kortum, & Miller, 2009).

2 Materials and Methods

Inclusion criteria for users in this study were patients with different types of seasonal diseases such as cold, cough, chest congestion etc, adults above 18 years of age, some knowledge of computer and mobile phone use, with at least read, listen, and speak English, own smartphone, and have no interaction with mHealth application before this study. After performing tasks, participants were requested to respond few scheduled questions at nearby suitable place or room within premises of conveniently selected hospital of Sindh province, Pakistan.

2.1 System Description

The ‘Pharmapedia Pakistan’ by pharma developers is the top mobile healthcare app by Google play downloaded by users in Pakistan. However, this application features medicine details from chemical compound to alternate names and availability to the users. Due to the limited scope of ‘Pharmapedia Pakistan’, ‘Mytabeeb’ mHealth application is currently downloaded by 3000 doctors in 60 hospitals across Pakistan since January 2016 (Farah, 2017). Moreover, it features doctor and patient interaction...
(see figure 2&3 as a sample). This app makes a list of recommended good doctors to access any time. Yet, developers are working for ratings on doctor efficiency and database that manages the efficiency of this app as well which makes this app sound for usability evaluation in this study.
2.2 User Assessment Tasks, instruments, and Measures

The tasks are given to participants included:

- Getting registered with the system
- Identifying the target disease of patient in the application
- Specifying the illness/problem
- Searching for related doctor
- Setting reminders as per doctor recommendations
- Setting appointment reminders with doctor

The tasks were based on the real case scenario as to how patients would interact with the system in a real-life situation and were validated by healthcare professionals as well. Followed by instrument measuring their age, gender, educational background, and their experience.

The ISO measures were the guidelines, as effectiveness was measured by:

- User was able to complete the task without help
- Completed the task with some difficulty or help
- User failed to complete the task even after help

An error was coded that user fails to solve to complete the task. Efficiency was measured by averaging the time taken by the individual users to complete each task from start to till exiting the app by the user. Satisfaction was measured by using SUS and scores were calculated following Brooke’s guidelines. Such that, 1 point was deducted from item 1,3,5,7,9, and 5 points from 2,4,6,8, 10 items respectively. The user application interaction process began after finishing demographic characters. The collected data from 15 users then proceeded for data analysis and results were calculated using Microsoft Excel spreadsheet. Descriptive statistics such as means and standard deviations were calculated in SPSS version 24.

3 Results

The 15 patient users shared different demographic characteristics. 8 patients were female and 7 were male. Most (70%) of them were middle age adults (30-39 years), 30% were between 20-29 years of age, 70% had a college education, while, 90% were familiar with computer and mobile usage. 60% of patients were diagnosed with the 3-month-chronic disease while remaining were affected less than 3 months.

3.1 Effectiveness

Task 5 and 6 were difficult with 30% and 40% failure rates respectively. Task 1, 2, 3, 5, and 6 were completed easily without errors. Task 4 was the most difficult to complete and accumulated the largest errors. The kinds of errors that occurred were difficult to remember steps by patients, seemingly similar options, specifically, (4) selecting the related doctor.
3.2 Efficiency

As may be seen in Table 1, task 3 and 4 consumed the longest amount of time, as might be expected given the difficulties with task success and errors mentioned above. On the other hand, Tasks 2 and 5 took the shortest times. Tasks of specifying the illness/problem and searching for related doctor (3 and 4) had mean scores 2–3 times as long as those related to interpreting values. As from table 1 most of the time consumed on task (3) and (4), while, task (5) took the shortest time. The mean scores for (2) and (6) were 2-3 times as long as those related to getting registered.

Table 1. Time on task per task

<table>
<thead>
<tr>
<th>Time per task (min)</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Task 5</th>
<th>Task 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>1.09 (0.70)</td>
<td>1.99 (0.69)</td>
<td>4.45 (1.67)</td>
<td>3.68 (1.76)</td>
<td>1.36 (0.39)</td>
<td>2.21 (0.73)</td>
</tr>
<tr>
<td>Range</td>
<td>0.94–2.38</td>
<td>0.91–2.50</td>
<td>1.56–6.47</td>
<td>1.39–6.89</td>
<td>0.89–2.01</td>
<td>1.27–3.77</td>
</tr>
</tbody>
</table>

3.3 Satisfaction

The average SUS score for the group was 85.5 (SD 12) indicating good satisfaction across these mHealth application users as seen in figure 4. However, a wide variation in the score from low of 61.5 and a high score of 98.5 (at a 37-point range). While the highest ranged from 78.5 to 98.5 good or excellent (30% of the patient sample) to the lowest from 61.5 to 75 to minimally acceptable (30% of the patient sample).

3.4 Descriptive and Usability Metrics

User characteristics and objective data were assessed, additional insight is observed in table 2. Descriptive statistics indicate a different across gender, age, and patient experiences. Males in this sample had higher average success, lower error rates, and higher mean SUS scores than female. The younger patient also had higher average

Fig. 4. Overview of modified SUS
task completion rates on tasks, only one error average on tasks, lower task completion times, and higher mean satisfaction scores. Patients with a recent disease had a higher mean satisfaction score, task success rates, fewer errors, short average time on tasks, and higher SUS score. Education seems to have no influence on satisfaction scores, while mobile familiarity added fewer errors to complete tasks.

<table>
<thead>
<tr>
<th>User characteristics</th>
<th>Success rate, mean (SD)</th>
<th>Error rate, mean (SD)</th>
<th>Time on task, mean (SD)</th>
<th>Satisfaction, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>95.05 (10.50)</td>
<td>1.25 (2.63)</td>
<td>1.05 (0.30)</td>
<td>85.31 (2.50)</td>
</tr>
<tr>
<td>Female</td>
<td>85.58 (11.92)</td>
<td>3.01 (2.71)</td>
<td>2.01 (0.45)</td>
<td>75.09 (13.08)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–29</td>
<td>100 (0.00)</td>
<td>1.00 (1.00)</td>
<td>1.50 (0.25)</td>
<td>89.65 (8.44)</td>
</tr>
<tr>
<td>30–39</td>
<td>90.50 (12.55)</td>
<td>4.40 (2.50)</td>
<td>2.45 (0.75)</td>
<td>75.40 (12.50)</td>
</tr>
<tr>
<td>Disease (months)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1</td>
<td>100 (0.00)</td>
<td>1.05 (1.09)</td>
<td>1.70 (0.30)</td>
<td>85.10 (6.07)</td>
</tr>
<tr>
<td>3</td>
<td>85.42 (12.29)</td>
<td>4.75 (2.78)</td>
<td>2.90 (0.80)</td>
<td>78.33 (11.75)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>93.75 (8.84)</td>
<td>3.40 (0.69)</td>
<td>2.31 (0.20)</td>
<td>80.25 (11.37)</td>
</tr>
<tr>
<td>College/university</td>
<td>90.63 (12.94)</td>
<td>3.65 (3.77)</td>
<td>2.22 (0.87)</td>
<td>85.35 (12.53)</td>
</tr>
<tr>
<td>Comp/ mobile Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less experienced</td>
<td>85.50 (16.04)</td>
<td>6.35 (3.54)</td>
<td>2.76 (0.20)</td>
<td>65.86 (19.14)</td>
</tr>
<tr>
<td>More experienced</td>
<td>92.10 (11.50)</td>
<td>2.60 (2.01)</td>
<td>2.10 (0.52)</td>
<td>89.52 (10.48)</td>
</tr>
<tr>
<td>Sample (n 10)</td>
<td>91.05 (11.80)</td>
<td>3.45 (2.73)</td>
<td>2.55 (0.65)</td>
<td>80.50 (11.85)</td>
</tr>
</tbody>
</table>

4 Discussion

This study demonstrates Usability Evaluation Model for the application and how researchers may consider relevant characteristics during mHealth application interaction. These recommendations are in line with growing mHealth usage rates (El-Gayar, Timsin, & Nawar, 2013). This study depicts the deployment of ISO standards to evaluate application usability in terms of effectiveness, efficiency, and patient satisfaction. The inclusive set of factors allow in-depth understanding of application usability by the user, their task, and their performance interaction requirements for health application. Moreover, this study shows how patient’s characteristics may influence interaction performance and how developers might enhance eHealth and mHealth applications on practical grounds.

4.1 Interpretation of Task Performance Results, Satisfaction, and Demographic Trends

Findings of this study show that task 4 (selecting relevant doctor) and 6 (setting appointment reminders with doctor) had errors, difficult, and were time-consuming to
user practice. These outcomes may be due to the steps involved and limited options due to application comprehensiveness. As compared to the task (1 registration) that involves one easy step and most registration, in general, it has become a routine task over internet portals in contemporary life settings. The overall satisfaction results were good for application usability rather excellent. These results indicate areas that need to be improved by developers. For example, multiple steps, repeated information, limited or similar options, and exact selection of keywords was difficult as in task 2 (identifying the target disease of the patient in the application) and task 4 (searching for related doctor).

Demography of the sample reveals that male were more aware than female respondents and performed with slightly better SUS scores. Similarly, younger patients also performed faster with few errors. This directs that developers must consider the demographic characters for more user-friendly interaction that will benefit a wide range of users. As results depict that user at ease with application achieved tasks with fewer errors, faster, and successfully completed the task which also increases user satisfaction.

4.2 Contributions to the Literature

To our knowledge, it is the first usability study on the mHealth application in the healthcare system in Sindh province of Pakistan in terms of assessing usability effectiveness, efficiency, and satisfaction by using validated measure. Additionally, the study utilized ISO designed usability standards and SUS instrument to compare usability metric performance outcomes to pertinent patient user characteristics. This study addressed the recommended gap of usability studies on patient product interaction by scholars (e.g., Mulvaney, Ritterband, & Bosslet, 2011; El-Gayar, Timsin, & Nawar, 2013). Further, addressed the need to explore the demographic characteristic and technology influence on usability performance scores to design future interventions for targeted populations (e.g., Coursaris & Kim, 2011; Or, & Tao, 2012). Also, addressed a recommendation to measure effectiveness, efficiency, and satisfaction (Lyles, Sarkar, & Osborn, 2014). In the past authors focused on negative outcomes of usability findings, this study depicted positive mHealth usability application. This study followed valid ISO standard methodology to explain the core usability issues that need to be addressed and provides specific technique on which scholars can capitalize to evaluate and improve mHealth usability application in the healthcare system in the Asian context.

5 Conclusion

The study findings serve as an exemplar for Usability Evaluation Model with good perceived usability satisfaction, and nearly 1/3rd of users rated at poor usability of the application. The results show objective data for developers and directly needed corrections. The objectivity of the results shows variation in the users that indicate developer and user mismatch that needs to be addressed.
This study utilizes a systematic quantitative approach by considering the different needs of the user who interacted with the mHealth application. It also reveals the practicality of the performance of different kinds of assessment measures. The study used the ISO standard method to measure effectiveness, efficiency, and satisfaction with validated tools such as the SUS instrument. Together, these application usability measures provide better understanding and serve as an exemplar for the methodological approaches by designers and researchers.

Literature gaps were addressed in this study by examining different characters among patients with technological know-how which provides feasibility for the developer to improve the design of mHealth application for better user usability on practical grounds. Moreover, results also indicate that socio-geographical, and personal features may influence the user experience. Therefore, in a wider perspective and applicability these results would play influencing role on mHealth application usability and developmental experience in future.

6 References


7 Authors

Abdul Samad Dahri, is a Ph.D. in HRM and progressive publication are in different fields of research mainly related to human resource management. His current research revolves around harmonizing management with innovation technology in different sectors.

Ahmed Al-Athwari is a former web developer at YemenSoft Company, Sana’a, Yemen. Currently, he is a master student at school of computing, Universiti Utara Malaysia. He is doing research on Software Engineering and IT Management. ah.alathwari2012@hotmail.com
Azham Hussain is the Associate Professor of Software Engineering at School of Computing, Universiti Utara Malaysia, Kedah, Malaysia. He is the founder of Human-Centered Computing Research Group, which is affiliated with the Software Technology Research Platform Center at School of Computing, Universiti Utara Malaysia. Azham Hussain is a member of the US-based Institute of Electrical and Electronic Engineers (IEEE), and actively involved in both IEEE Communications and IEEE Computer societies. azham.h@uum.edu.my