The 100,000 Participant Laboratory -
A Crowd-Centered Approach to Design
and Evaluate the Usability of Mobile Apps

Ming-Hui Wen[✉]

Department of Commercial Design and Management,
National Taipei University of Business, Taipei, Taiwan
donwen@ntub.edu.tw

Abstract. Mobile apps play important roles in our daily lives. Software developers typically follow the user-centered design (UCD) approach to develop products that optimize usability. However, most usability methodologies require lab experiments or field studies. Usability experiments require higher costs and efforts to prepare related tasks, such as selecting the environmental setting, subject recruiting and conducting the study. Compared to the laboratory experiment of personal computers (PC), mobile usability laboratory settings present greater challenges related to controlling the research context variables.

The purpose of this study is to build an online usability lab by applying Internet and crowdsourcing technologies. The lab allows researchers to invite Internet users to participate in the usability experiment online. This study follows the usability engineering theoretical framework to define mobile usability contextual factors. The clouded-based online usability lab was introduced as a conceptual framework to support both Android and iOS mobile apps. Mobile app developers can upload their apps to the lab system to collect and monitor usability data and track the behavioral data of human-computer interaction. The ultimate goal of this lab was to help researchers conduct usability experiments in real world contexts and develop the dynamic mental model of product users.

Keywords: Mobile usability · Mental model · User experience · Crowdsourcing

1 Introduction

Usability engineering is an important research topic within the study of Human-Computer Interaction and based on the user-center design (UCD) approach. User-centered design seeks to ensure the designers' work meets the users' requirements and ability to conveniently use the design product. Usability engineering applies both qualitative and quantitative research methods to help researchers understand user-product interactions and experiences.

Regardless of the particular method, usability laboratories have played significant roles in the development of the conventional usability process. The process of scientific investigation has allowed researchers to collect user response data and interpret findings to understand user experience quality and product effectiveness (Norman 1988).

A usability experiment is composed of four contextual variables, including the user, task, environment and tool (Gould 1988). Usability research should consider these four factors simultaneously to ensure the experimental setting replicates the u. Traditional computer software requires users to sit in front of desks. Conversely, mobile apps are more complex and influenced by many factors, such as multi-tasking, cooperative work, social dynamics and, user interface. The quantity and complexity of contextual variables might affect mobile app use. Therefore, conventional lab-based usability research methods might not be able to provide experimental contexts that highly match real mobile app usage.

Some online services already exist for testing the mobile apps such as Testdroid, Testin, and MonkeyTalk. However, these services only test technical issues, such as operation system compatibility, functional stability, and system stress. These online testing services only ensure that mobile apps operate smoothly online. However, a smooth operating app does not necessarily meet user requirements.

Online survey systems, such as SurveyMonkey, Limesurvey, Google Forms, SurveyGizmo, KwikSurveys, SoSci Survey, Typeform, Client Heartbeat, Gravity Forms, WuFoo, Formsite, and Formassembly, have helped researchers gather data on users' opinions and preferences about mobile apps. However, these tools only allow researchers to develop survey items and analyze data with graphs and tables. These survey tools do not assess the usability experience when the users interacting with the apps.

This study uses the Internet and communication technology to develop an online usability laboratory. The crowdsourcing and social networking mechanism is used to recruit participants. Mobile app developers can upload their products to recruit appropriate candidates. The proposed online lab is expected to allow mobile app developers to collect real usability data (even dynamic data) from real users in a real context.

2   The Influence Variables of Usability Testing

Based on previous literature, the user, tasks, tool and environment are used as a framework for developing the online usability lab. This framework allows Internet users to participate in the usability experiments online. Figure 1 introduces the four contextual factors applied in this study.
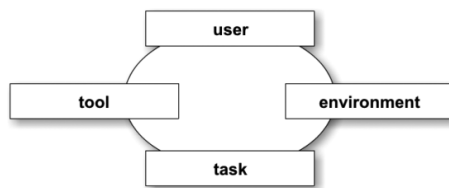
Fig. 1.   The four factors of a usability context

2.1   User

Demographics are the most frequently used factors to describe user profiles (Nielsen 1992; Allen et al. 1993; Levanthal et al. 1994). Demographic variable include nationality, race, age, date of birth, blood type, gender, education, income, occupation, family patterns, geographical location, urban living, family life cycle, marital status, and other variables. Demography can be used for basic human classification to help researchers define their user segments.

Lifestyle has also been used to classify user segments (Wansink and Park 2000). This study applied the AIO (Activity, Interest and Opinion) lifestyle rating scale to measure users' intrinsic and extrinsic lifestyles (William and Douglas 1971; Reynolds and Darden 1972). Personality is also commonly applied to psychologically classify users (Christine and Dewit 2001; Devaraj and Easley 2008). Researchers have most commonly used the Big-Five scale to assess personality, which includes five dimensions (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism; McCrae ans Costa 1992, McCrae and Costa 1999).

The online usability lab will provide APIs (Application Programming Interfaces) to allow researchers to add other rating scales or indicators classifying user groups or statuses. One example of which open-APIs applies is this addition of related survey items about TAM (Technology Acceptance Model) and Innovation adoption lifecycle.

2.2   Task

Tasks relate to the interactions between users and tools. This study applied activity theory, with three levels (i.e., activity, action, and operation) to describe user motivation and task goals (Leont 1978). Activity analysis allows for the assessment of users motivation and primary objectives related to a particular activity. Engeström (1999) proposed an activity triangle to explain other contextual variables related to user activity. This model includes tool, community, rules, and division of labor factors. These social and culture factors might also affect product usability and interaction design. Researchers need to understand users' motivations, objectives and how they conduct tasks with the tool to achieve their primary objective. Such research may allow mobile app designers to further understand users' mental models and design products accordingly (Norman 1983).

2.3   Tool

The online usability lab uses smartphones and tablets as vehicles to assess mobile app usage. The Android/iOS smartphones and tablets are the most common of such tools. Technical characteristics relating to the tool level include, but are not limited to, brand and model series, CPU number, memory size, operation system, software versions, Internet speed, telecom operators, and screen size. Although different smart devices have different capabilities regarding hardware and software levels, usability experiments should also consider the effects of the tools themselves to ensure experimental quality. Some online services provide a heatmap feature to track operational behaviors

step-by-step while the user interacts with the mobile apps (see Fig. 2). Such technologies are very useful for understanding users' activity, actions, and operational level behaviors. In order to promote future development of the online usability lab, the system provides scalable APIs to allow third party technologies. Examples of such technologies include heatmaps, screen video recorders, and mobile eye tracking. Users could operate the mobile app and perform the think-aloud methodology at the same time using the screen recorder. The system can collect this data and remotely send it back to the online usability lab through the Internet. Researchers can then collect usability experiment data from a real usage context. Such information should be very helpful for researchers when conducting large-scale experiments with worldwide samples.

Fig. 2. Head map feature for tracking user operations on a mobile app (Image Source: AppAnalytics.io).

2.4   Environment

Usability experiments should also consider elements of the physical environment that might affect the results of the experiment. Examples of environmental factors include moving state, mobile versus stationary, the noisy degree of illumination, temperature, humidity, vibration, and weather. Physical environmental factors are likely to affect usability during product interaction (Wickens 1992). The mobile app was usually used in the out-door environment and mobile context. The effect of the physical environment in mobile apps is much more complex than the simple PC software context.

To help recreate the physical environment condition of mobile app usage, sensors can be embedded in smart devices to identify pertinent environmental factors. For example, most smart devices are equipped with GPS (Global Positioning System) and Geographic Information Systems (GIS: Geographic Information System). These technologies can be used to indicate users' current locations and relay information back to the online usability lab. Therefore, GPS data can predict user mobility (e.g., standing, walking, running or driving. Computing device speed and direction data from the gravity sensor (G-sensor) and accelerometer of the mobile phone allows for the online lab system to predict possible user activities. The development of the Internet of Things (iOT) has led to the inclusion of temperature, humidity, gas, and pressure

sensors to monitor the environment and aid daily life. The online usability lab provides open data APIs to allow researchers to feed data from smart sensors to the system.

3   Motivating Users to Participate in the Online Usability Experiment

The main challenge in realizing the 100,000 participant lab is recruiting thousands of Internet users. The operational model is a platform, meaning we need to grow the number of usability experiment cases and participants. To grow the number of usability experiment cases, the system design SDKs (Software Development Kits) will be used to allow app developers to easily connect to the system and upload cases for testing though APIs. Crowdsourcing social networking technology is also used to recruit more Internet users. The gamification mechanism and associated tangible/intangible rewards system were also added to solicit participation.

Crowdsourcing is an Internet-based technique to facilitate large-scale tasks that are costly or time-consuming with traditional methods (Marzilli et al. 2009). Crowdsourcing has been applied to many Internet services, such as Amazon Mechanical Turk (MTurk), CloudFactory, Clickworker, CloudCrowd, and Fiverr. Research also has proven that crowdsourcing can reduce the cost and time associated with micro human resources tasks online (Kittur and Chi 2008).

Viitamaki (2008) proposed the Focus, Language, Incentives, Rules and Tool (FLIRT) model to guide the crowdsourcing service design. This author also identifies four kinds of peoples (creator, critic, connector, and crowd) who play different roles in the crowdsourcing system. The creator is responsible for generating original solutions for the crowdsourcing tasks. Critics and connectors emphasize their opinions and share information about crowdsourcing tasks to influence a large number of people. Finally, the crowd has a low-level of participation and only activates in some key events (Viitamaki 2008). The FLIRT model can be applied to human-to-human (the four kind of users) and human-to-computer (crowdsourcing tasks) interactions. This study first connects the usability lab to Facebook to assist the registration process. Facebook also allows participants to invite friends and family to participate. Tangible rewards (points) are given to users who invited members of their social network to participate, based on quantity.

Gamification is the process of applying in-game elements to attract and engage people in performing non-game tasks (Deterding and Dixon 2011). Game elements can provide positive experiences and involve people in the game playing process. Some people connect so strongly with games that they become addicted (Hsu and Wen 2009). Gamification has been applied in education (Raymer and Design 2011, Kapp 2012), marketing and sales (Huotari and Hamari 2011) and human resource management (Kumar 2013).

Three gamification features exist, including quests, experience points, challenge unlock and leaderboard access. The quest feature provides detailed information related to app type, function requiring testing, usability goal, testing procedure, expected number of participants, and reward points. Users who were meeting the inclusion criteria will receive the invitation to participate in the online experiment. The number

of reward points given by app providers may differ. Users who would like to participate in the quests for reward points should first complete at least 10 basis quests to collect experience points and unlock advanced quests. All testing of participants' achievement across different types of apps (e.g., educational apps, game apps, and tool apps) is calculated and displayed in the leaderboard.

4   The Development of the Usability Laboratory

This study was exploratory, and the author attempted to adapt the traditional laboratory usability experiment to an online environment. The main challenge in this effort was to reconstruct the usability context from offline to online. Therefore, this study applies four human-computer interaction context elements (user, task, tool, and environment) to replicate the usability environment. The use of mobile devices and smart sensors helps increase the efficiency of the usability data collection process.

The system lab supports mobile app developers with SDKs, allowing developers to easily connect their mobile apps to the online lab or upload the Android APK file to the system. The system provides a set of common usability methodologies, such as A/B testing, heuristic evaluation, think aloud, surveys, and open-question interviews. Researchers can easily form an online usability experiment by selecting a usability template from the menu. The computation module will then automatically calculate the related algorithm.

Data can be collected in two ways within this model, including human input and machine-collected data. The platform offers questionnaire design features to collect various forms of human input data. Machine-collected data include that derived from mobile device embedded features (e.g., GPS, GIS, G-sensor), and wireless smart sensors. All usability testing is sent back to the cloud server through APIs. The system can then provide qualitative or quantitative usability indicators as input for the data
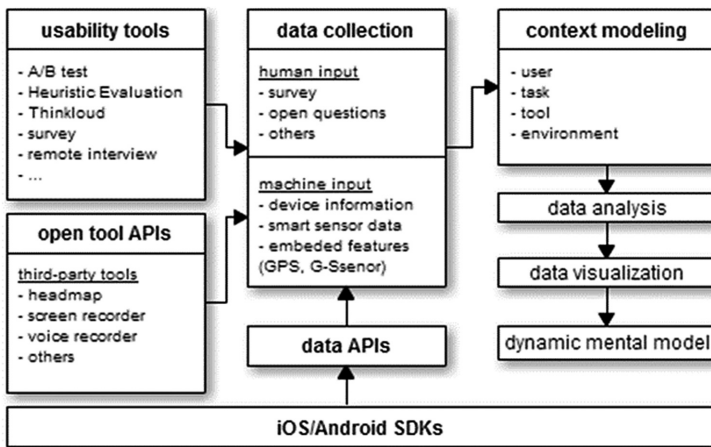

Fig. 3. Conceptual framework of the online usability texting laboratory

visualization model. Finally, the visualized data function was provided with a dynamic update function to allow researchers to easily interpret findings. By building an informative visualization graph with an appropriate algorithm to process the data, researchers can try to build a dynamic mental model of the mobile app indicating its quality. Figure 3 illustrates the lab's conceptual framework.

5   Conclusion

Quality outcomes related to an online, lab-based experiment are very close to those of physical labs. However, the former is timelier and cost-efficient. Additionally, Internet-based experiments allow much easier worldwide participant recruitment (Yen et al. 2013). The ultimate goal of this study is to development an online usability lab that allows hundreds of thousands of people to participate (iOS and Android). Compared to traditional lab experiments, it is easier to collect dynamic usability data and recruit participants in the online version.

The proposed online lab is expected to bring convenience to researchers conducting mobile app usability experiments in three ways. First, this lab prevents the need to replicate the experimental environment because of the mobile nature of the app. Second, this lab makes acquiring the target number of subjects or evaluating cross-cultural effects much easier through crowdsourcing. Finally, this lab allows users' behavioral data when interacting with the mobile app to be sent back to the cloud server. The experimenter can collect data within 24 h, obviating the need for human assistance to support the experiment. The proposed system is expected to help researchers conduct usability experiments in real world contexts and conceptualize the dynamic mental models of product users.

References

Allen, C.D., Ballman, D., Begg, V., Miller-Jacobs, H. H., Muller, M., Nielsen, J., Spool, J.: User involvement in the design process: why, when & how?. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, pp. 251–254. ACM (May 1993)

Roy, M.C., Dewit, O., Aubert, B.A.: The impact of interface usability on trust in web retailers. Internet Res. **11**(5), 388–398 (2001)

Costa, P.T., McCrae, R.R.: Four ways five factors are basic. Personality Individ. Differ. **13**(6), 653–665 (1992)

Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15. ACM (September 2011)

Devaraj, S., Easley, R.F., Crant, J.M.: Research note-how does personality matter? relating the five-factor model to technology acceptance and use. Inf. Syst. Res. **19**(1), 93–105 (2008)

Engeström, Y.: Activity theory and individual and social transformation. Perspect. Act. Theor., 19–38 (1999)

Gould, J.D.: How to design usable systems. In: Handbook of Human-Computer Interaction, pp. 757–789, Amsterdam (1988)

Hsu, S.H., Wen, M.H., Wu, M.C.: Exploring user experiences as predictors of MMORPG addiction. Comput. Educ. **53**(3), 990–995 (2009)

Huotari, K., Hamari, J.: Gamification" from the perspective of service marketing. In: Proceedings of CHI 2011 Workshop Gamification (May 2011)

Kapp, K.M. The gamification of learning and instruction: game-based methods and strategies for training and education. Wiley. com (2012)

Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with mechanical turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 453–456. ACM (April 2008)

Kumar, J.: Gamification at work: designing engaging business software. In: Marcus, A. (ed.) DUXU 2013, Part II. LNCS, vol. 8013, pp. 528–537. Springer, Heidelberg (2013)

Leontev, A.N.: Activity, consciousness, and personality. Moscow: Progress (1978)

McCrae, R.R., Costa Jr, P.T.: A five-factor theory of personality. Handb. pers. Theor. Res. **2**, 139–153 (1999)

Levanthal, L., Teasley, B., Stone, D., Lancaster, A.M., Marcus, A., Nardi, B., Nielsen, J., Kurosu, M., Heller, R.: Designing for diverse users: will just a better interface do?. In: Conference Companion on Human Factors in Computing Systems, pp. 191–192. ACM (April 1994)

Nielsen, J.: Finding usability problems through heuristic evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 373–380. ACM (June 1992)

Norman, D.A.: Some observations on mental models. In: Mental Models, pp. 7–14 (1983)

Norman, D.A.: The psychology of everyday things. Basic books, New York (1988)

Raymer, R., Design, E.L.: Gamification: using game mechanics to enhance eLearning. Elearn Mag. **2011**(9), 3 (2011)

Reynolds, F.D., Darden, W.R.: Intermarket patronage: a psychographic study of consumer outshoppers. J. Mark. Res. **36**(4), 50–54 (1972)

Viitamaki, S.: The FLIRT model of crowdsourcing: planning and executing collective customer collaboration. Marketing Master Thesis, Helsinki School of Economics, Spring (2008)

Wansink, B., Park, S.B.: Methods and measures that profile heavy users. J. Advertising Res. **40** (2000), 61–72 (2000)

Wickens, C.D.: Computational models of human performance. CSERIAC Gateway **3**, 8–12 (1992)

William, D.W., Douglas, J.T.: Activities, interests and opinions. J. Advertising Res. **11**(4), 27–35 (1971)

Marzilli, M., Yan, T., Holmes, R., Ganesan, D., Corner, M.: mCrowd: a platform for mobile crowdsourcing. In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, pp. 347–348. ACM (November 2009)

Yen, Y.C., Chu, C.Y., Yeh, S.L., Chu, H.H., Huang, P.: Lab Experiment vs. Crowdsourcing: A Comparative User Study on Skype Call Quality (2013)