Evaluating Mobile Applications: A Spreadsheet Case Study

Derek Flood, Rachel Harrison, Claudia Iacob*, David Duce
1Oxford Brookes University,
[derek.flood, rachel.harrison, iacob, daduce]@brookes.ac.uk

ABSTRACT

The power of mobile devices has increased dramatically in the last few years. These devices are becoming more sophisticated and allow users to accomplish a wide variety of tasks while on the move. The ease with which mobile apps can be created and distributed has resulted in a number of usability issues becoming more prevalent. This paper describes the range of usability issues encountered at all stages of the mobile app life cycle, from when users begin to search for an app to when they finally remove the app from their device. Using these results we developed a number of guidelines for both app developers and app platform developers that will improve the overall usability of mobile apps.

Keywords: Mobile technology, portable computing, application process model, mobile usability.

1 INTRODUCTION

Advances in technology have allowed a range of sophisticated devices to emerge which enable users to perform a variety of tasks in a mobile context. These tasks include both tasks which were previously only available on traditional desktop applications and tasks that are only enabled by the use of mobile devices. To facilitate these tasks a wide range of mobile applications, referred to here as apps, are available from easy to install locations, such as the App store provided by Apple or the Android Marketplace provided by Google. In order to improve the portability of both of these types of apps, a number of compromises are necessary. These compromises, such as smaller screen size, more limited processing power and the mobile context in which the device is used, have frequently had a negative effect on the usability of these apps. The compromises can most readily be seen during the use of mobile apps. There has been a large amount of research into the usability of specific mobile apps (Ahmadi, 2008; Geven, 2006; Schmied, 2009; Shrestha, 2007). However, little research has been conducted on the selection, installation and removal of mobile apps, all of which can be problematic. This paper presents a study in which we examined aspects of user interaction with mobile apps during various steps in the usage lifecycle of mobile applications: application identification, installation, usage, and removal. To illustrate the range of interactions that a user may have with a mobile app we also present a mobile app process model which shows the typical life cycle of a mobile app.

This paper is structured as follows. Section 2 describes related work. Section 3 addresses user interaction with mobile devices. Section 4 of the paper outlines the case study conducted to investigate the usability of mobile applications at various steps in their lifecycle. Based on this investigation, a set of guidelines for improving usability is provided in Section 5. A number of limitations to this study are presented in Section 6. The paper is then concluded in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 Usability Models and Standards

There are three major ISO usability standards, each one identifying different attributes for measuring usability. ISO/IEC 9241-11 (1998) defines efficiency, effectiveness, and satisfaction as the main
attributes of usability, while ISO/IEC 9126-1 (2001) proposes understandability, learnability, operability, attractiveness, and compliance with published style guides. ISO/IEC 9126-4 defines the concept of quality in use as the composition of three factors: effectiveness, productivity, and safety. The general limitation of these standards is that they are too abstract and give very little indication on how to interpret scores of specific usability metrics.

To answer these limitations, The Metrics for Usability Standards in Computing (MUSiC) model was developed to “provide valid and reliable means of specifying and measuring usability, while also giving diagnostic feedback which enables the design to be modified to improve usability” (Bevan, 94). In measuring usability, the MUSiC model assesses the user performance in terms of the context used for the evaluation, effectiveness, efficiency, productivity, learning, satisfaction, and cognitive workload. Performance-based assessments of usability, however, lack aspects such as user satisfaction.

Seffah & Donyae propose QUIM (Quality in Use Integrated Measurement), a hierarchical usability evaluation model which decomposes usability into factors, then into criteria, and finally into specific metrics (Seffah, 2006). There are 10 factors considered by the model: efficiency, effectiveness, productivity, satisfaction, learnability, safety, trustfulness, accessibility, universality, and usefulness. As criteria, the model includes time behaviour, resource utilization, attractiveness, likeability, flexibility, minimal action, minimal memory load, operability, user guidance, consistency, self-descriptiveness, feedback, accuracy, completeness, fault tolerance, resource safety, readability, controllability, navigability, simplicity, privacy, security, insurance, familiarity, load time, and appropriateness. To define the usability metrics associated to the model the authors identify a set of relationships between the factors and the criteria. For example, the efficiency factor is directly related to criteria such as resource utilization, operability, and feedback.

In (Hussain, 2008), the authors propose a method for developing usability metrics using the Goal Question Metrics (GQM) approach (Solingen, 1999). The GQM model is a hierarchical structure which starts with a goal which is further refined into questions. Metrics are then created for each question. For example, using this approach, Hussain et al. consider one goal as being “Efficiency” and derive a set of questions for it, e.g. “Once users have learned the design, how quickly can they perform tasks?”. Corresponding to this example question, the authors consider the following metrics: task completion time, duration used to finish given exercises, duration spent on each screen.

2.2 Usability in the Mobile World

2.2.1. Research methodologies

The rapid progression of technology has led to an increase in the number of mobile applications available. Although these applications offer a number of advantages in terms of portability and convenience they do so at the cost of usability. Whether performed in the laboratory or in the field, usability testing can be used to find the problems users face when interacting with a specific system. These problems have various levels of complexity, ranging from critical problems (which prevent the subjects from completing the task and recur over all test subjects) to serious problems (which increase severely the time to complete the task) and cosmetic problems (which increase the time to complete the task but only slightly) (Duh, 2006). There are many methods for assessing usability, including heuristic evaluations, cognitive walkthroughs, usability testing, and comparison against existing guidelines. Deciding which method to use is difficult and depends on the objectives of the overall evaluation process.

Support in making such a decision has been given by the work of (Duh, 2006) and (Kjeldskov, 2005). In (Kjeldskov, 2005), four different approaches for evaluating the usability of mobile guides are used: field evaluation, laboratory usability evaluation, heuristic walkthrough, and rapid reflection. The results show that none of the techniques managed to identify all the critical problems, and field
evaluation was the only technique to identify more than a half of the serious problems. In addition the type of problems identified by each technique differed. While field evaluation identified issues concerning the validity and precision of the data presented on the device, the heuristic walkthrough pointed to problems related to the overall use, flexibility and usefulness of the mobile guide. Rapid reflection brought to light issues related to the perceived relevance of the available information and linked it to the users’ emotional responses to the overall design of the mobile guide.

A slightly different approach is used in (Duh, 2006) for identifying the differences between usability testing conducted in laboratory and field environments in terms of the following criteria:

i) quality and quantity of the usability problems revealed,

ii) usability performance measure of each task,

iii) users’ satisfaction, and

iv) users’ behavioral patterns.

The final results showed that the number of usability problems identified in the field environment was twice the number of problems revealed by the laboratory testing. Moreover, the time needed to perform the task was much longer in the field environment. On the other hand, the behavior of the subjects was more positive in the laboratory environment which led the authors to conclude that this would affect the overall results of the usability testing, making the laboratory usability testing insufficient in evaluating the mobile application in its actual context of use.

2.2.2. Predicting usability

Zhang and Adipat (2005) have highlighted a number of issues that affect the usability of mobile apps, including Mobile Context, Connectivity, Small Screen Size, Display Resolution, Limited Processing Capability and Power and Data Entry Methods.

Efforts have been made to look into some of these issues, focusing mostly on mobile web browsing (Ahmadi, 2008; Geven, 2006; Schmield, 2009; Shrestha, 2007) and mobile guides (Kjeldskov, 2005). Acknowledging the importance of mobile web browsing, but also the limits in the usability of the applications addressing it, Ahmadi et al. introduce a method for automatically adapting a desktop presentation to a mobile presentation. Usability testing performed on the method applied revealed that “mobile users are likely to have an immediate, goal-directed intention of finding particular information rather than reading through a long document” (Ahmadi, 2008). As a consequence, efficient information searching and browsing is vital.

Building on this, Schmield et al. (Schmield, 2009) describe the results of a multidimensional study which investigates usage scenarios as well as the usability of mobile tailored websites, answering questions such as: What kind of websites are most often accessed using a mobile phone and who are the users?, What websites are currently available in a mobile tailored version? Should a mobile tailored version provide the same information as the full version or is a limited set of functionality an advantage?, Do the mobile optimized versions really have an advantage in comparison with the full version when viewed on a mobile phone?, Which type of mobile phone qualifies best for surfing the web (full version)? Some of the answers include:

- The typical mobile phone user is usually male, between 20 and 29.
- Out of the mobile tailored sites investigated, 55% of them fell in the category ‘Information services’, while 20% were social networking sites. The rest of them were search engines, online shops, or entertainment related sites.
- Generally, touch screens are preferred over pen-based or keyboard based devices. However, the type of device did not seem to influence the mean time required by the subject for completing the task.
A discussion on the optimal information hierarchy for mobile use and the usability effects for reduced screen size is proposed in (Geven, 2006). Slow reading speed, poorer comprehension and information retrieval performance are just a few of the usability effects reduced screen size brings (Duchnicky, 1983; Jones, 1999; Larson, 1998). Ways of dealing with such issues include adapting the way information is structured in hierarchies (Geven, 2006). A comparative study of four hierarchies on three different devices brought to light the following:

- Users clearly prefer narrower hierarchies on all three devices.
- The certainty of choice was equal for the four hierarchies tested.
- The quality of labels has a significant effect on the balance between depth and breadth.

Mobile devices generally do not contain a traditional keyboard as this would make the device too large to be portable. Some devices incorporate a physical keyboard which utilizes small keys while other devices use touch screen technology to present a keyboard to the user on screen. These keyboards require the physical keys to be smaller than traditional keyboards. Even with smaller keys it is not possible to include all required keys on a single screen. The iOS and Android platforms address this issue by providing users with three separate keyboards; one containing letters, one containing numbers and some special characters and a third containing additional special characters.

2.3 Spreadsheets and Usability

Research on spreadsheets has focused on various aspects, including expanding spreadsheet languages through direct manipulation and gestures (Burnett, 1998), communicating unit error messages in spreadsheets (Abraham, 2005), type inferences for spreadsheets (Abraham, 2006), graph-based visualizations (Kankuzi, 2008), testing (Rothermel, 2001), and web based spreadsheet-mediated collaboration (Ginige, 2010). However, little work has been done in identifying the usability issues revolving around spreadsheet applications (either mobile or desktop-based) and ways to go about them. In (Flood, 2011), an evaluation of mobile spreadsheets focuses on their effectiveness, efficiency, satisfaction, and simplicity. Results show that:

- The standard of mobile spreadsheet applications is low for the tasks the participants wish to accomplish.
- The subjects were slightly dissatisfied with the existing mobile spreadsheet applications.
- The limited screen size of mobile devices requires the subjects to perform considerably more navigation when looking at large spreadsheets.
- Subjects find it difficult to conceptualize the overall spreadsheet and to see how the section on-screen fits within the context of the spreadsheet causing cognitive overload.
- The ability to view spreadsheets on a mobile device convinced the subjects to put up with some of the limitations brought by the device.
- Half of the subjects admitted to having made errors while using a mobile spreadsheet application. On the other hand, half of the subjects considered the device to be sufficient for their needs.

Flood et al. (Flood, 2011), identified navigation as an issue that affects the performance of people debugging spreadsheets through voice recognition technology (an alternative input method for desktop computers) in comparison with the traditional input methods provided by the keyboard and mouse. By addressing this issue it was found that debugging performance could be improved. It was also found that participants audited more cells with the improved navigation system, and the ability to do this is an important aspect of the debugging process.

Chen et al. (Chen, 2010) conducted an evaluation of users entering text on a small size QWERTY keyboard. This evaluation required 15 participants to enter a passage of text using the small sized
keyboard. On average participants used 540 keystrokes to enter the passage of text. The most prevalent type of error made by these participants during the task was a key ambiguity error, which occurs when a user enters a character other than the target character. It was found that on average, participants made about 9 key errors on the first typing task. All participants made at least one error of this type during the study. Errors of this type, when made on a spreadsheet, may result in a misspelled word or in an incorrect reference in a cell formula, which could alter values in a spreadsheet substantially. It has been shown that even on desktop computers errors like this persist. Two independent studies (Panko, 1998; Powell, 2009), found that over 85% of the evaluated spreadsheets contained errors.

3 USABILITY OF MOBILE APPLICATIONS

3.1 What are Mobile Devices?

The meaning of the term ‘mobile device’ is now quite well understood. Traditionally the term referred to a smart phone but with the advent of larger devices the term is now used more widely. The differences between larger devices (such as the iPad) and small devices (such as smart phones) are considerable and reduce our ability to generalise research results.

![Figure 1 – The Portability Continuum](image)

To scope our work we have developed the Portability Continuum shown in Figure 1. We consider mobile devices to be a subset of portable devices: not all portable devices are mobile. For example, desktop computers can be moved between locations and are therefore portable however we do not consider these to be mobile devices here. Along the portability continuum devices are placed in increasing order of physical size from left to right. The devices in Figure 1 are used for illustrative purposes only. Other devices, such as portable MP3 players, medical devices, etc. could also be added to this continuum.

The meaning of ‘mobile device’ as used here only considers the physical size of the device. Other factors, such as the presence of an internet connection or the ability to connect to other devices or services, may also play an important role in the context of mobile apps but are not considered here. This paper focuses on smart phone devices: tablets and e-Readers are outside the scope of this paper. Although these are mobile devices according to our use of the term the differences in interaction styles and abilities raises different research questions to those presented here.

3.2 A Process Model for Mobile Apps

To better understand the way users interact with mobile apps a survey of smart phone users was conducted together with laboratory usability tests. We hope that understanding the nature of user interactions with mobile apps and the issues users face will help to improve the user experience of mobile apps. The full range of interactions considered during the survey and the laboratory tests is captured in the mobile app process model which begins at the time a user first encounters an app, usually when an app is downloaded as a result of a search, to when a user finally removes it from their device. It is important to consider all of these stages, as they will all impact upon a user’s overall
impression of the system and therefore its ultimate success or failure. Our mobile app process model has four stages as depicted in Figure 2.

Figure 2 – The Mobile App Process Model

The four stages encompassed in the app process model are:

- **App Identification**: At this stage the user has identified the need for a mobile app for a particular task. This task could be anything such as entertainment, ability to view documents or to log health information. There are many ways in which the user could identify an app including the use of mobile app stores such as the Android Marketplace.

- **App Installation/Upgrade**: Once the user has identified an appropriate app they are ready to install it. On traditional desktop systems this step offers the users many customisations; however, with mobile apps the user may not have this customisability because the operating system may make these decisions on behalf of the user. This stage also includes the installation of updates to the app as they are required.

- **App Usage**: Once the app has been installed it is ready to be used. This stage represents the most extensive range of interaction between the user and the app. Therefore it is at this stage that most usability problems will be noticed by the user.

- **App Removal**: Once the user has completed their task and they no longer need the app they may remove it from their device, usually to free up space for more apps. In some cases a long time will pass between the final usage of the app and its removal from the system.

The process model reflects the stages which the user is required to execute to complete their desired task. Before the user can accomplish this they must first identify and install a suitable app. Once the task has been completed, they may decide to remove the app from their mobile device. In some cases the user may decide after removing the app to reinstall the app on their device and thus will proceed to search for the app again.

3.3. **Evaluating User Interactions with Mobile Applications within the Process Model**

To evaluate user interactions with mobile applications in the framework of the process model we adopted the GQM approach (Basili, 1994) and formulated a series of questions related to each of the stages described by the model. In a first instance, a user searches for a mobile app once s/he has identified the exact purpose for which the application is needed. The search is performed over various sources and different factors may influence the user’s final choice. This process could be more or less tedious depending on the issues the user experiences. Also the ease of locating an application for a specific purpose may bias the user’s rating of the overall experience of using the application. Several
questions need to be answered when assessing interaction with a mobile app at this initial stage of the process, as shown in figure 3.

Once the user finds an application of interest, they progress to installing the application on the device. At times, searching for an application does not automatically lead to actually installing any application on the device and the reasons for this are many. The actual installation process of a mobile app can bring several constraints and difficulties; hence, a study of the usability of a mobile app should also consider such issues. Questions to ask for assessing the level of usability of a mobile app at this stage are depicted in Figure 4.

Having an application installed and ready to use, allows the user to take full advantage of the application’s features. Even though installed for a specific purpose, the app may not be needed after the purpose has been met, so the frequency with which an app is used may be directly related to several other factors worth exploring. Locating an app on the device once it has been installed is another issue of concern when analysing the usability of the application itself. Usability assessment factors such as learnability are of concern when evaluating the use phase of the model proposed. Some of the questions to ask when evaluating usability at the use stage of the model are presented in Figure 5.
Lastly, mobile apps are sometimes removed. The reasons for that are worth exploring when assessing the usability of the application since they may reveal usability issues which have convinced the user not to use the application anymore and remove it from the device. Hence, following the model proposed in this paper, evaluating the usability of a mobile app at this particular stage would require answering the question in Figure 6.

![Figure 6 – Research questions for the removal phase of mobile apps](image)

Most usability evaluation analyses focus on the actual use of a mobile app. However, a comprehensive evaluation of the usability of such an app requires evaluation at every step of its lifecycle. Different issues may occur at various stages and they have a direct influence on: a) the user experience and b) the other stages of the application’s lifecycle. Our process model represents the typical lifecycle of a mobile application. Evaluating each of the stages is important in order to identify the usability issues the user might face during each of them.

4 CASE STUDY

4.1 Objectives

The objective of the study is to evaluate user interaction with mobile applications (in particular mobile spreadsheet applications) in the framework of the process model described in Section 3. For this, the study focuses on the following questions:

| App Identification | (1) Where are apps installed from? |
| App Identification | (2) What factors affect the choice of an app? |
| App Identification | (3) What issues one might experience when searching for an app? |
| App Identification | (4) How easy it is to locate an app for a specific purpose? |
| App Installation/Upgrade | (5) How frequently would one install an app? |
| App Installation/Upgrade | (6) How easy is it to install an app? |
| App Installation/Upgrade | (7) What issues are experienced when installing an app? |
| App Usage | (8) How frequently are the mobile apps used? |
| App Usage | (9) Once installed, how easy is it to locate an app? |
| App Usage | (10) How long does it take to learn to use a mobile app? |
| App Usage | (11) What are the most common usability issues with the use of mobile apps, and in particular of mobile spreadsheet applications? |
| App Removal | (12) Why are mobile apps removed? |

Table 1 – Case study objectives

The overall context of the study is mobile apps in general. However, for a more precise understanding of the use stage described above, the paper considers one particular class of applications, mobile spreadsheet applications. The motivation for this comes from the results of a survey of experienced spreadsheet users (Flood, 2011) which show that almost 80% of participants required access to a spreadsheet while away from a traditional computing device, such as a laptop or desktop computer. Modern smart phones can provide the necessary functionality by allowing users to run a small scale version of a spreadsheet application while on the move. There are a number of applications available which can be used this way. However, these applications vary greatly in terms of usability and the range of features available.
4.2. Design

4.2.1. Research Method

Throughout the study, we employed two research methods for gathering data and analyzing the results:

- A survey was run using 51 participants. The survey was structured to answer question 1-10, and 12.
- Laboratory usability testing was performed with 12 participants. The participants were given a task and their interactions with the mobile application were recorded. The results of the laboratory usability testing provide answers to question 11.

4.2.2. Participants

The survey and the laboratory usability testing were conducted with different participants, the two groups being completely independent.

4.2.2.1. Survey

In total 51 participants took part in the survey. The gender analysis showed that 61% of the participants were male and 39% female and their ages ranged from 16 to 60 as shown in Figure 7.

![Figure 7 – Percentage of survey participants by age](image)

It can be seen that the 82% of the participants were aged between 21 and 40. None of the participants was over 60 years of age.

![Figure 8 – Breakdown of participants by platform](image)
There are a number of platforms that can be run on mobile devices. To ensure that the issues identified were not specific to a particular platform, the survey was open to users of all platforms. Figure 8 shows the breakdown of participants by platform. It can be seen that the two most represented platforms are the iOS platform and the Android platform, accounting for 84% of participants. The remaining participants used Blackberry OS, Symbian and the Windows OS. Three of the participants did not use Smartphone devices and make up the remaining 6% of the participants. These participants either did not indicate the model of their phone or their phone was not a traditional smart phone but did allow java apps to be executed. Participants may have had more than one device; this information was not collected.

![Figure 8](image)

**Figure 8** – Percentage of participants by platform

To determine the level of experience participants had with smart phones, they were asked “How long have you owned this/these device(s)?” Answers ranged from less than one month to approximately 3 years, with an average usage across all participants of nine months. Figure 9 shows the number of participants by duration of ownership. It can be seen that the majority of participants have owned a smart phone for less than one year.

4.2.2.2. Laboratory Usability Testing

Twelve participants were recruited to participate in the laboratory usability testing. Of these, 25% were female and 75% were male. Age wise (Figure 10), participants were from 21 to 60, the majority being less than 30.

![Figure 9](image)

**Figure 9** – Percentage of participants by duration of device ownership

To determine the level of experience participants had with smart phones, they were asked “How long have you owned this/these device(s)?” Answers ranged from less than one month to approximately 3 years, with an average usage across all participants of nine months. Figure 9 shows the number of participants by duration of ownership. It can be seen that the majority of participants have owned a smart phone for less than one year.

To determine the level of experience participants had with smart phones, they were asked “How long have you owned this/these device(s)?” Answers ranged from less than one month to approximately 3 years, with an average usage across all participants of nine months. Figure 9 shows the number of participants by duration of ownership. It can be seen that the majority of participants have owned a smart phone for less than one year.

4.2.2.2. Laboratory Usability Testing

Twelve participants were recruited to participate in the laboratory usability testing. Of these, 25% were female and 75% were male. Age wise (Figure 10), participants were from 21 to 60, the majority being less than 30.

![Figure 10](image)

**Figure 10** – Percentage of participants by age

Figure 11 shows the participant’s prior level of experience with desktop spreadsheet applications. It can be seen that 75% of the participants rated their level of experience with spreadsheets as being intermediate. Only 1 participant (8.33%) rated their level of experience as being novice. Only one of the participants had used a spreadsheet on a mobile device prior to the experiment. This participant found the previous experience to be frustrating and so only used the spreadsheet app twice on the mobile device before removing the app.
The participant’s previous level of experience with the iOS platform is displayed in Figure 12. It can be seen that approximately 60% of participants rated their experience with the iPhone as being either intermediate or expert. With the exception of one participant, the remaining participants all had experience of other smartphone devices.

The age distribution of the subjects means that the results obtained do not specifically address one age interval. Also, working with subjects with different levels of experience in the area of spreadsheet applications allowed us to identify a larger diversity of usability issues. Expert users were expected to report different issues than novices. However, the majority of the subjects were intermediate users.

### 4.2.3. Procedure

#### 4.2.3.1. Survey

The participants for the survey were recruited through a number of social media platforms, including Facebook and LinkedIn. A link to the survey was initially posted by the first author on his personal accounts with links to people from a number of countries, including England, Ireland, Spain and Austria, and so participants may have been recruited from each of these countries. This information however was not collected during the survey and is presented here only to indicate the range of the recruitment process. It was observed that participants from each of these countries placed comments stating that they had completed the survey. In addition to being placed on the first author’s account, a number of the participants reposted the link to the survey on their own accounts further propagating the survey. The survey was kept open for approximately 4 weeks during which time it was accessible to participants 24 hours a day. The results presented here are based on all participants who completed the survey and submitted their results. The survey allowed participants to attempt the survey but only recorded the results of those who completed and submitted the survey. It is possible that additional participants started the survey and subsequently abandoned it before the final submission.
4.2.3.2. Laboratory Usability Testing

The most common uses of spreadsheets on a mobile device have been found to be viewing and editing (Flood, 2011). Consequently the participants in the laboratory usability testing were asked to edit nine cells and locate a further seven pieces of information located on the spreadsheet. The spreadsheet used was adopted from Howe and Siskin (Howe, 2006) who used a similar spreadsheet during an investigation of spreadsheet debugging. The spreadsheet calculates the five year projections of Acme Ltd. It contains 3 worksheets; Payroll, Office Expenses and Projections. The breakdown of tasks was selected to reflect the frequency with which tasks would typically be done on a mobile device. For the location tasks, participants were given a brief description of the cell and asked to navigate to this cell and change the background color to one of their choice. Thus participants are forced to first locate and then enter the cell allowing the navigation time to the cell to be determined.

4.3. Data Collection

4.3.1. Survey

The survey was divided into 5 sections (Appendix A). The first 4 sections refer to the user’s experiences at each stage of the mobile app process model while the final section was used to collect demographic information about the participants. The sections are outlined below:

- **App Location:** This section examined the methods used by participants for locating new apps. It also examined the factors used by participants to decide between the many different apps available for a given purpose and the time participants spend searching for apps.
- **App Installation/Upgrade:** This section of the survey focused on the issues experienced by participants while installing apps. It examined the frequency with which users installed apps as well as the issues they have experienced during the installation process.
- **App Usage:** This section asked about using mobile apps including, frequency of use, learnability, and other usability issues experienced by the participants.
- **App Removal:** This section examined the strategies for removing apps from a device and the frequency with which apps were removed.
- **Demographic Information:** This section requested information about age, gender and experience of participants.

In each section participants were asked to respond based on their expectations but we found that participants responded according to their previous experiences of mobile apps and so this is what we report in this paper.

4.3.2. Laboratory Usability Testing

The laboratory usability testing captured the actions of participants while using the mobile spreadsheet application, providing data with respect to the users’ interactions. For a further analysis of the data the participants’ interactions with the spreadsheet application were recorded. These recordings were coded for analysis using the coding scheme described in Table 2.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>newCellSelect</td>
<td>subject selects the cell/character they are interested in</td>
</tr>
<tr>
<td>sameCellSelect</td>
<td>subject (re)selects the cell/character previously selected</td>
</tr>
<tr>
<td>wrongCellSelect</td>
<td>subject selects a cell/character they are not interested in selecting</td>
</tr>
<tr>
<td>multipleCellSelect</td>
<td>subject mistakenly selects multiple cells instead of one desired cell</td>
</tr>
<tr>
<td>selectOption</td>
<td>subject selects an option of the application</td>
</tr>
<tr>
<td>navigateOptions</td>
<td>subject browses through the options available within the application</td>
</tr>
<tr>
<td>scrollUpDown</td>
<td>subjects scrolls the screen vertically</td>
</tr>
<tr>
<td>scrollLeftRight</td>
<td>subject scrolls the screen horizontally</td>
</tr>
<tr>
<td>goBack</td>
<td>Backspace</td>
</tr>
<tr>
<td>save</td>
<td>subject saves the modifications</td>
</tr>
<tr>
<td>Action</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>cancel</td>
<td>subject cancels the modifications performed</td>
</tr>
<tr>
<td>giveUp</td>
<td>subject gives up the task</td>
</tr>
<tr>
<td>switchScreen</td>
<td>subjects selects a different screen of the application</td>
</tr>
<tr>
<td>typeLetter</td>
<td>subject types a letter</td>
</tr>
<tr>
<td>typeDigit</td>
<td>subject types a digit</td>
</tr>
<tr>
<td>typeUpperCase</td>
<td>subject types an upper case</td>
</tr>
<tr>
<td>retypeChar</td>
<td>subject retypes the same character again</td>
</tr>
<tr>
<td>lostKeyboard</td>
<td>subject tries to bring back the keyboard</td>
</tr>
<tr>
<td>hideKeyboard</td>
<td>subject tries to hide the keyboard</td>
</tr>
<tr>
<td>actionFeedback</td>
<td>feedback is provided to the subject on their action (more on this)</td>
</tr>
<tr>
<td>whatsThis</td>
<td>subject is uncertain about the results of his more recent action</td>
</tr>
<tr>
<td>close</td>
<td>subject closes the application</td>
</tr>
<tr>
<td>zoomIn</td>
<td>subject zooms in</td>
</tr>
<tr>
<td>zoomOut</td>
<td>subject zooms out</td>
</tr>
<tr>
<td>fishEye</td>
<td>subject makes use of the fishEye feature</td>
</tr>
</tbody>
</table>

Table 2 – Coding scheme used for the laboratory usability testing video analysis

The synthetized data collected from the laboratory usability testing supported our analysis of question 11, providing information on: a) the frequency of each action performed by the participants while using the app and the meaning of these frequencies and b) the recurring sequences of actions followed by the participants and the meaning of these.

4.4. Results

4.4.1 App Identification

Figure 13 shows the list of sources of apps location identified during the survey. It can be seen that the app stores are the main source of new apps. The high frequency of the App Store and the Android Market reflects the popularity of the iOS and Android platforms. Other sources include the Google search engine and installation from a desktop or laptop computer.

![Frequency of each source](image)

Figure 13 – Source of app location

A search of the app stores will usually return a high number of results, making it difficult for users to find the best app to install. In order to better understand how users decide between different apps, participants provided a wide variety of answers, using different terminology to represent the same concepts. In some cases the participants listed multiple factors which influenced their choice of app. A qualitative analysis of the answers was performed, whereby each response was classified based on its semantic content as shown in Figure 14.
Figure 14 shows the frequency with which each factor was identified as being important to the participants. It can be seen the two most important factors for the selection of an app are Function and Price. This highlights the need for app developers to carefully consider price and function. One participant remarked that apps “Must be free or come so highly recommended I’d find it essential”. A number of participants said that they would use the opinions of other users to help locate apps. In this context the “others” are not limited to acquaintances but also include app ratings, reviews and the numbers of downloads. Another important aspect to the selection of an app is User Interaction, which is used here to represent factors such as screen layout, usability, ease of use etc. It is unknown how users evaluate these factors before using the app. However, in most of the app stores users have access to screenshots which may help them to rate some of these factors.

Two lesser quoted factors are Curiosity and File Size. Two participants said that they would select an app just to better understand what the app was. Only one participant said that the size of the app would influence their choice. Despite the high number of factors affecting the users’ choice of app, it was found that participants spend on average only 10 minutes looking for an app. The longest time any single participant spent looking for an app was reported to be 1 hour.
Figure 15 shows a box plot of the search times quoted by participants. Participants who spent more than 20 minutes searching for an app are statistical outliers (indicated by a circle). We found that 75% of the participants spent less than 10 minutes looking for an app, with 50% spending 5 minutes or less.

The most prevalent issue participants experienced while searching for an app was the high volume of results obtained when searching the app stores. Approximately 20% of participants remarked that while searching for apps a large number of unrelated apps are returned. While evaluating apps for diabetes management, Garcia et al. (2011) performed a search in the App Store for “diabetes” and over 200 apps were returned including apps for cooking, information about diabetes, general health apps and apps for calculating Body Mass Index. The participants also identified the following issues while searching for an app:

- Determining the appropriateness of an app or difficulty finding reviews (14%);
- Unavailability of an app for a specific task (8%);
- Apps not available for specific devices or location (8%);
- Poor store interface (6%);
- Inappropriate naming of apps (e.g. Non-English titles) (6%).
- Excessive loading times (4%).

Despite the issues mentioned it was reported that participants still found it easy to locate apps, rating this with a median value of 2, on a five point Likert scale, where 1 is Very Easy and 5 is Very difficult.

4.4.2 App Installation

During the survey the frequency with which users install apps was examined. Figure 16 shows the percentage of participants that install apps with each frequency. It can be seen that most participants only install apps when needed. One third of the participants said that they install apps on a weekly basis. Some participants (6%) said they would never install an app on their device.

![Frequency of Installation](image)

Participants were asked to rate on a 5-point Likert scale how easy they found it to install an app. A median value of 1 was found, indicating that on average participants found it very easy to install apps. The modal value was also 1. When using the App stores, apps can be installed in one click. Unlike traditional desktop applications, where there may be a large number of configurable options, apps are installed automatically onto the handset, hiding a lot of the installation procedure from the user.

A small number of participants reported several issues encountered while installing an app, such as the handset would stop working during installation and the system would need to be restarted. One participant experienced problems when the app was installed in an unexpected location which
subsequently affected updates being installed for that app. However two thirds of the participants reported experiencing no issues while installing apps.

4.4.3 App Usage

4.4.3.1. Survey

The first aspect of app usage examined during the survey was the number of mobile apps installed on the participants’ devices. It was found that the participants had between 0 and 140 apps on their devices, with an average of 35 apps per participant. During the survey participants were asked how many of these apps they used on a Daily, Weekly and Monthly basis. As the actual number of apps on the participants’ devices varied widely, the percentage of the total installed apps used with each frequency was calculated. The results of this analysis are presented in Figure 17.

![Box plot showing app usage frequency](image)

**Figure 17 – Frequency of apps use by time**

It was found that the median number of installed apps used on a daily basis was approximately 25%. The median number of apps used on a weekly basis it was found to be slightly higher, at approximately 38% whereas the median number of apps used on a monthly basis was. These results show that only a small percentage of the installed apps are used on a daily basis.

The relatively large number of apps that can be installed on mobile devices means that finding the desired app can become more difficult as the number of apps installed increases. Participants were asked to rate, on the five point Likert scale the easiness to find an app on the device after being installed. The median value reported by the participants was 1, indicating that the participants found it very easy to locate an app, once it has been installed. The modal value was also 1.

The efficiency with which users can locate an app may be dependent upon the number of apps installed on the device which in turn may be dependent on the length of time that a user has owned the device. We analysed the data to determine whether the problem of locating apps becomes greater over time.
Figure 18 shows a scatter plot of the number of apps against the length of time the device has been owned. It can be seen that as time increases the number of apps also increases. A correlation coefficient of 0.386 was found using a one-tailed Spearman’s correlation test. This is statistically significant at the 1% level.

When users use an app for the first time they must spend time learning to use it. We found that on average the participants spent approximately 8 minutes learning to use a new app. Over 60% of the participants said that they spent 5 minutes or less learning to use a new app.

Approximately 20% of the participants reported no issues during the use of mobile apps. Of the remaining participants a number of issues were reported. The most prevalent issues and the percentage of participants who reported this issue were:

- **System Stability (32%)**: Participants reported the crashing and freezing of apps as being a common issue. In some cases the system exited the app without any warning or explanation.
- **Unreliable Connection (12%)**: A number of participants reported issues relating to accessing the internet. The most prevalent complaint was the access speed. A number of apps, such as the Facebook App, require an internet connection to synchronise information between the user’s device and their online account. Some users also reported that their connection was lost completely.
- **Continuity (10%)**: There is a lack of continuity between different apps and between the underlying platform and the individual apps. This lack of continuity can result in a number of problems, such as confusion between the back and exit buttons. One participant remarked “all blackberry messaging functions have an option to "mark as opened / unopened" which is not available in the menu in the messaging section of the facebook app”.
- **Other Issues** these included:
  - Unresponsiveness of both the app and the touch screen (8%)
  - Long loading times (6%)
  - Apps requiring updating (4%)
  - Inability to use multiple apps in parallel (4%)

The results give a general picture of the most prevalent issues experienced by mobile app users. It is not suggested that all apps give rise to these issues or that this is an exhaustive list of the issues experienced by mobile app users. However, these are the issues that survey respondents have experienced most often and should be considered by developers of mobile apps.

4.4.3.1. Laboratory usability testing: The Spreadsheet Case Study
During the laboratory usability testing we recorded and then coded the participants’ interactions with a specific mobile app (the spreadsheet app). The aim of this investigation was to capture behavioral patterns users develop around using mobile apps, and in particular mobile spreadsheet apps. Such results inform us of the usability issues that users may experience. Once the videos of the laboratory usability testing were coded, the distribution of these codes was computed and is shown in Figure 19.

The most often performed action while using the mobile spreadsheet app was scrolling from left to right. On analysis we found that zooming out was not performed often (less than 2% of the video fragments being assigned this code) which led to the need to navigate within a spreadsheet using scrolling. The subjects would scroll from left to right in a number of situations: to make sure that the cell selected was the right cell, to get information on the values stored by the cells not shown on the screen, to get an overall view of the data contained in the spreadsheet etc. There was very little feedback from the system on the subjects’ actions, so the subjects were at times uncertain of what part of the spreadsheet they are viewing and how this part was related to the rest of the spreadsheet.

From Figure 19 it is clear that a high percentage of video fragments were coded with “sameCellSelect” and “wrongCellSelect”. Often the subjects would try to select one cell, but: a) they accidentally selected another cell close to the one desired or b) they tried to select the same cell again and again due to lack of feedback. The percentage of fragments coded with “newCellSelect” (which means that the right cell was selected at a first attempt) is very similar to that for “sameCellSelect”.

“Whatsthis” was used when the subjects were puzzled by the response of the application to their actions. In addition to this, the system provided very little feedback and this did not help the user with questions such as: a) was the action successful? (e.g. Is the background color set?), b) what are the consequences of this action? (particularly if an action was made by mistake) c) where is the cursor located? There was very little use of features such as fisheye, zoom in and zoom out.
In addition to analyzing the frequency of each individual action, a forward step is to look at the common sequences of actions (Figure 20). Figure 20 shows, for example that in 21.4% of the cases, canceling an action was consecutively followed by selecting a new cell.

Some of the values in Figure 20 point to a number of recurring sequences of actions. Our analysis of these recurrences has brought to light potential usability issues with the application:

- “sameCellSelect” occurred in 66.9% of the cases, showing that selecting the same repeatedly was common. The subjects would select the same cell several times in a row because of a lack of feedback. The maximum repetition of this action was 9 times.
- “save” indicates that the subject successfully made the desired modifications. In more than half of the cases saving current modifications led to moving the cursor away from the selected cell.
- In 37.9% of the cases, scrolling from left to right led to more scrolling. This is partially explained by the fact that most of the time a scroll to the left would immediately be followed by a scroll to the right and vice versa. A similar case is the “scrollUpDown” code.
- Selecting an option led to saving the modifications in 73.3% of the cases. This result relates to the background color change task. The high percentage indicates that in the majority of cases the subjects had little trouble changing the background color of the cells. However, at times, selecting an option would lead to reselecting that particular option. This was the case when, after selecting a color on the palette, the application would provide no feedback leading to confusion on the completion of the action. To rectify this the subjects would at times try to select the color again.
- Switching to another screen led to navigating within the screen (scrolling from left to right and from up to down).
In the majority of cases, the subjects would hide the keyboard immediately after typing some characters because otherwise the keyboard would impede the navigation within the spreadsheet.

After selecting a wrong cell (different from the one desired), the subjects would select the one desired in more than a half of the cases. However, 17.9% percent of the cases point to the situation where the same cell (the wrong one) is being selected. This, together with the high percentage of “sameCellSelect” followed by “sameCellSelect” points to possible usability issues to be further explored and discussed.

75% of the cases when the subjects intended to close the application were followed by a cancelation of the action. This is particularly interesting since in most cases it was not clear to the subjects that the close button would close the whole application and not just the open spreadsheet. Only when asked whether they want to save the work (as result of closing), would it become clear that the close button closes the application entirely.

The actions preceding the “whatsthis” code included: a) zooming (the subjects were uncertain of what caused the zooming and how to go back to the previous state), and b) canceling an action (not getting any feedback and not observing the consequences they were expecting, the subjects became confused).

4.4.4 App Removal

Most (73%) of the participants said they removed the app because they no longer needed it while another 9% said they removed apps to free up space on their device. In addition to the answers provided by the survey, users entered a number of other reasons, including:

- The app is no good. (4%)
- Too many apps clutters up the menu (2%)
- Constant daily updates (2%)
- Better alternatives found (2%)
- The app changes part of its structure to one the user did not like (2%)

There are many strategies that can be used to remove apps. Some users will remove several apps at once whereas others remove apps one at a time, when they think they are no longer needed. Of the participants that had removed an app, 84% said that they removed apps individually when they felt they no longer needed them. The remaining participants said that they removed apps in groups. One participant remarked that “I clear out once every couple of weeks.”

![Frequency of App Removal](image)

Figure 21 – Frequency of app removal

Figure 21 shows the frequency with which users remove apps. It can be seen that participants mostly remove apps when needed. 27% of the participants said that they removed apps on a weekly basis while a further 23% removed them on a monthly basis. Only 4% of the participants removed apps on a daily basis.
4.5. Discussion

In this section we summarize the results for each phase of the life cycle.

4.5.1 App location

The results of the survey indicate that the installation of mobile apps is a relatively easy task. The most dominant issue identified by the participants was the excessive number of results that are returned during a search of the app store. The survey also found that participants predominantly use the app stores when acquiring new apps and so this is a problem that needs to be resolved. Participants also said that they consider function and price to be the most important factors when acquiring new apps. App developers should therefore consider these factors carefully.

4.5.2 App Installation

The results show that participants find installing mobile apps to be very easy. This is not surprising because the installation of a mobile app is an uncomplicated procedure compared to the installation of desktop applications. Mobile apps are not usually customised to the same extent as desktop applications and this simplifies their installation.

4.5.3 App Usage

Even though the results of the laboratory usability testing show that mobile spreadsheet app bring a lot of challenges during their use, the survey’s results show that mobile apps are still perceived to be easy to use. The wide range of apps that are available allows users to find alternatives to ones they find complicated or unpleasant to use. The number of apps that users have installed on their device varied widely among the participants; however on average participants had approximately 35 apps installed on their device. This relatively high number of apps can make it more difficult to locate a single app. The survey found that as time progresses the number of apps increases and this lengthens the search time to locate a single app. We also found that participants only used a small percentage of their apps on a daily basis. Unfamiliarity with the location of apps complicates the search problem. Providing users with a grouping option on the handset would help to solve this problem: we envisage that on installation the user could be given an option to group the apps in a meaningful way. The user then simply has to remember which group the app is in to reduce the search space.

4.5.4 App Removal

Removing apps from a mobile device was reported (through the survey) to be an easy task. The majority of participants removed apps once they were no longer needed, implying that participants may need to re-install an app if they unexpectedly need to repeat a given task. Although the majority of participants removed apps individually, the survey found that 16% of participants removed multiple apps simultaneously. For these participants it would be useful to provide a feature which allows them to select, from a list, the apps they want to remove so that this can be done with the click of a single button.

4.6. Threats to Validity

The validity of the laboratory usability testing we performed is limited by a number of factors [cite Wohlin]. Here we consider the threats to construct, internal and external validity.

4.6.1 Construct validity
Construct validity is concerned with whether the variables used in the usability testing accurately measure the concepts they purport to measure, i.e. has the data that we collected given us accurate estimates of the efficiency with which users can perform a number of typical tasks with a mobile spreadsheet? Since the timings that we measured were absolute and were collected automatically we are confident that the data collected does accurately reflect actual elapsed time. Evidence of the typicality of the tasks performed is provided by (Chan, 96).

4.6.2 Internal validity

Internal validity is concerned with whether the relationship between the data collected and the factors affecting user interaction with mobile apps is a causal relationship, and has not simply happened as a result of chance. The usability testing was performed in a laboratory setting and this could be a threat to validity because participants may have been distracted by the unfamiliar environment. However, the participants were given some time to familiarise themselves with the environment before starting the experiment and so we believe this threat has been minimised.

4.6.3 External validity

External validity is concerned with the generalizability of the results. The small number of participants in the usability testing is clearly a threat to external validity. However, the tasks chosen are typical of the use of mobile spreadsheets in business and so our work is a contribution to the understanding of issues which may arise.

5 USABILITY GUIDELINES IDENTIFIED

5.1 App location

Distribute apps through the appropriate app store. During the survey we found that participants mostly use the app stores as a source for new apps. Very few participants used other sources such as web searches or the iTunes desktop application. To be able to distribute apps through proprietary stores, developers must adhere to strict guidelines. This requires extra work on behalf of the developer.

Chose keywords to describe the app carefully. As with web searches, app developers should pay particular attention to the search terms used to identify their apps. We found that the majority of participants spend less than 10 minutes searching for an app. The high volume of apps available coupled with the general terms that are used by developers to describe their apps means that a many irrelevant apps are returned with every search. It is therefore important for app developers to carefully consider the keywords they use when entering their apps to the app stores.

Carefully consider the price and function of the app. There are a number of factors that can affect a user’s choice of app. We found that participants were mostly influenced by two things, the function of the app and its price. The high number of free apps available means that participants are reluctant to spend large amounts of money on an app, with some participants unwilling to pay anything at all for an app and some participants only willing to pay if the app was essential.

Improve search capabilities. This survey showed that users spend on average 5 minutes searching for an app. Research has also shown that app store searches typically return a high number of irrelevant apps and that less than 10% of the results returned were relevant to the search terms (Garcia et al., 2011; Flood et al, 2011b). To improve the probability of returning meaningful results a tiered keyword search could be used. This would prioritize the results by the level at which the search term is found. The app stores could allow developers to specify a few key words or phrases at each level and then present users with a list of all the apps matching the search terms, ordered by the level at which their search term was found.
5.2 App Usage

Ensure the app is easy to learn. During the survey it was discovered that users only spend on average 5 minutes learning to use a new app. If users cannot learn to use an app within this time, the high number of alternative apps enables them to select a different app with the same functionality. It is important therefore that developers try to minimize the learning time associated with their app.

Ensure that app use is memorable. The survey also found that participants used only a quarter of their apps on a daily basis. Almost half of the installed apps are used on a monthly basis, indicating that participants spend a large period of time between each use of the app. As a result of this we believe that app developers should design apps to help the user to remember how to use them.

Group related apps. We found that users had on average 35 apps installed on their device. Grouping apps helps users to access a desired app more efficiently.

Make navigation features obvious. The laboratory usability testing showed that features such as zooming in and out were rarely used this leading to a poorer navigation and overall awareness. Most of the participants were not even aware of the existence of these features. Such features should be clearly signaled so that the users become aware of the full potential of the application.

Provide feedback. Feedback supports users in understanding their actions and the impact of their actions in relation to the application. During the laboratory usability testing, five types of needed feedback became evident.

- **Validation feedback.** Once the user performs an action on the application, it is helpful for them to understand whether it was a correct action with respect to the interaction context they are in. For example, selecting a spreadsheet cell should be provided with feedback on whether the cell was the right cell or not.

- **Location feedback.** Due to the small size of mobile devices, it is easy for users to get disoriented with respect to the cursor’s location on the screen or to the position of the section they are viewing. For such disorientation to be avoided, it is important for the user to be provided with feedback on the location that they are currently viewing in the overall picture.

- **Neighborhood feedback.** Participants needed to navigate through the document for long periods of time to become aware of the data contained by the cells surrounding those visible on the screen. Feedback on such information would reduce the users’ navigation overload.

- **Selection feedback.** One of the main issues observed during the laboratory usability testing was that the participants were not always sure that the cell they were trying to select was indeed selected. In answer to that they kept selecting the same cell repeatedly. In part, this problem could be solved by providing feedback once a cell is selected by highlighting its margins or bringing the cell to the front. Similar techniques could be used for applications which use other types of selection.

- **Keyboard feedback.** The location of the keyboard was not always clear to the users. At times, they struggled to understand how to hide the keyboard or, on the contrary, how to make it accessible. The application should support the user by providing feedback on the current state of the keyboard at all times.

**Hints as to the consequences of each action.** Users might be unaware of the consequences of some actions and realize this only after the action has been performed. Even worse, in some cases users might not perceive the consequences of their action or fail to understand it. Hints on what each action led to would help users. As a concrete example, the color changing of a cell in the spreadsheet...
application could be modeled as a wizard where all the steps are specified a priori and the user is aware of what the next step is.

**Provide evident milestones.** At times during the laboratory usability testing, participants faced difficulties in trying to cancel an action by clicking mistakenly on the button closing the application. It was not always clear how to save, cancel or quit the application and this affected the overall usability. Such features should be made evident and distinct, so that at any time during their interaction with the application the user is aware of how such features work and how they are distinct.

5.3 **App Removal**

Allow apps to be removed in groups. The survey found that a small number of participants remove multiple apps at the same time. If users could select multiple apps for removal simultaneously this would improve the efficiency with which apps could be removed.

5. **CONCLUSIONS AND FUTURE WORK**

In this paper, we describe how we collected data to investigate the lifecycle of a mobile application from when it is released into the market until it is removed from a mobile device. We used both a survey and laboratory usability tests to get a better understanding of the users’ behavior during searching, installing, using, and removing mobile applications. Results show that searching processes can be tedious due to the large number of results they return, but installing apps is not problematic. We examined the use of a particular mobile app (a spreadsheet app) to discover that, although the survey did not point to any major usability issues reported by the participants, the laboratory usability tests brought to light several areas worth improving. Based on these results and observations, the paper proposes a set of guidelines to improve the usability of mobile apps.

**REFERENCES**


Appendix A – Survey used throughout the study

1. What type of platform are you using on your device?
2. How long have you owned this/these device(s)?
3. Where do you install Apps from?
4. What factors affect your choice of App?
5. What issues have you experienced while searching for an App?
6. How easy do you find it to locate an App for a desired purpose?
7. Which of the following best describes the frequency with which you would install an App?
8. How easy do you find it to install an App on your device?
9. What issues have you experienced while installing an App?
10. How frequently are you using mobile Apps?
11. Once installed, how easy do you find it to locate an App?
12. Approximately how long would you spend learning to use a new app?
13. What issues do you find most common while using Apps?
14. Why do you remove Apps?