Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications

Abstract

Usability testing of software applications developed for mobile devices is an emerging research area that faces a variety of challenges due to unique features of mobile devices, limited bandwidth, unreliability of wireless networks, as well as the changing context (environmental factors). Traditional guidelines and methods used in usability testing of desktop applications may not be directly applicable to a mobile environment. Therefore, it is essential to develop and adopt appropriate research methodologies that can evaluate the usability of mobile applications. The contribution of this paper is to propose a generic framework for conducting usability tests for mobile applications through discussing research questions, methodologies, and usability attributes. The paper provides an overview of existing usability studies and discusses major research questions that have been investigated. Then, it proposes a generic framework and provides detailed guidelines on how to conduct such usability studies.

Keywords: Usability testing, mobile applications, mobile devices, evaluation

1. Introduction

With the continuous advances in wireless technology and the widespread use of mobile devices such as cell phones, personal digital assistants (PDAs), palms, and pocket PCs, many innovative mobile applications are emerging, aiming to enhance wireless communication and provide users with ubiquitous access to information (Li & Liao, 2000). Many businesses have deployed mobile applications to gain competitive advantage. Such applications developed specifically for small mobile devices include daily news alert services, classified mobile advertising, restaurant and entertainment listings, wireless Web portals, and mobile commerce (m-commerce) applications (Varshney & Vetter, 2002).

The high demand and fast growth of mobile applications have attracted extensive research interests. Because developing mobile applications with an easy-to-use interface is critical for successful adoption and use of applications, one of the important research issues is regarding how to conduct an appropriate usability test using mobile devices in a wireless environment. Usability testing is an evaluation method used to measure how well users can use a specific software system. It provides a third-party assessment of the ease with which end users view content or execute an application on a mobile device. An effective usability test has to be able to elicit feedback from users about whether they use an application without (or almost without) difficulty and how they like using the application, as well as evaluate levels of task performance achieved by users (Wichansky, 2000).

There are various guidelines for usability testing of desktop applications. However, those established concepts, methodologies, and approaches commonly used in traditional human-computer interaction research are not always applicable to mobile applications (Jones et al., 1999) due to mobility and the distinct features of mobile devices and wireless networks. Ideally, usability testing of mobile applications should be carefully designed to cover all or most possible situations of a mobile environment (Kim et al., 2002). In reality, however, this poses many challenges. For example, it is difficult to foresee the exact situations of the application use - users may be standing, walking, or sitting in a dark or bright environment while using an application. As a result, a usability test may have to concentrate only on certain aspects of a mobile application and sacrifice others. Furthermore, traditional research methodologies used in usability testing, including controlled laboratory experiments and field studies, have various limitations in a mobile environment, such as ignoring the mobile context or lack of sufficient procedural control. Therefore, it is essential to develop guidelines for usability testing of mobile applications.

This paper is aimed to contribute to this important research area. Built upon the literature, it proposes a generic framework for usability testing of mobile applications, discusses several important issues in this field, and provides insights on how to conduct usability studies according to the nature of applications and usability attributes being evaluated. The rest of the paper is organized as follows. In Section 2, the fundamental concepts and challenges of usability testing of mobile applications are introduced. The existing work in this field will be briefly summarized in Sections 3. In Section 4, a generic framework for usability testing of mobile applications is proposed and a variety of issues regarding the research methodology are discussed. Finally, the paper concludes in Section 5.

2. Usability of Mobile Applications

2.1 Mobile applications

Mobile applications, referred to software systems operating on mobile devices, are evolving rapidly, making ubiquitous information access at any time and anywhere a true reality. For example, many mobile applications have brought Internet services to mobile devices (Kaasinen et al., 2000). In the business area, M-Commerce (Mobile e-Commerce) applications, such as mobile banking and advertising, extend electronic businesses to mobile devices. Customers can check their bank account balances and carry out business transactions through their cell phones (Varshney & Vetter, 2002; Zhang, 2003).

There have been some usability studies for mobile applications. Some focus on Wireless Application Protocol (WAP) evaluation (Chittaro & Cin, 2002; Kassinen et al., 2000). In the field of mobile education, usability studies are conducted when mobile devices are used for collaborative learning or information access (Danesh et al., 2001; Luchini et al., 2002). In the entertainment industry, mobile users can enjoy watching video or playing interactive games on their mobile devices. Those advanced features of mobile applications enable users to carry out a variety of activities through mobile devices. Because achieving a high level of user satisfaction is critical to the success of mobile applications, usability testing is a mandatory process to ensure that a mobile application is practical, effective, and easy to use, especially from a user's perspective.

2.2 Challenges in usability testing of mobile applications

The unique features of mobile devices and wireless networks pose a number of significant challenges for examining usability of mobile applications, including mobile context, multimodality, connectivity, small screen size, different display resolutions, limited processing capability and power, and restrictive data entry methods.

- **Mobile context**

It can be defined as “any information that characterizes a situation related to the interaction between users, applications, and the surrounding environment (Dey & Abowd, 2001).” It typically includes the location, identities of nearby people, objects, as well as environmental elements that may distract users’ attention. It is very difficult to select a methodology that can include all possibilities of mobile context in a single usability test (Longoria, 2001).

- **Connectivity**

The slow and unreliable wireless network connection with low bandwidth is a common hindrance for mobile applications (Longoria, 2001). This problem largely affects data downloading time and quality of streaming media (e.g., video and audio streams). Strength of signals and data transfer speed in a wireless network may vary at different time and locations, compounded by user mobility (Sears & Jacko, 2000). Therefore, how to deal with various network conditions must be taken into consideration in a usability study.

- **Small screen size**

Physical constraints of mobile devices, especially small screen size, can significantly affect the usability of mobile applications (Jones et al., 1999; L. Kim & Albers, 2001). Direct presentation of most WWW pages on small mobile devices can be aesthetically unpleasant, un-navigable, and in the worst case, completely illegible (Bickmore, 1997).

- **Different Display Resolutions**

The display capability of mobile devices supports much less display resolution (normally 640*480 pixels or below) in comparison with desktops. Low resolution can degrade the quality of multimedia information displayed on the screen of a mobile device. As a result, different levels of display resolution on different mobile devices may cause different usability test results (Jones et al., 1999).

- **Limited Processing Capability and Power**

Computational power and memory capacity of mobile devices lag far behind desktop computers. Some applications that require a large amount of memory for graphic support or fast processing speed, such as an application of 3D city maps for PDAs (Rakkolainen & Vainio, 2001), may not be practical for mobile devices. Because of limited processing capability of mobile devices, developers may have to disable some functions (e.g., high resolution images and dynamic frame movement).

- **Data Entry Methods**

Providing input to small devices is difficult and requires a certain level of proficiency (Longoria, 2001). Small buttons and labels limit users' effectiveness and efficiency in entering data, which may reduce the input speed and increase errors. Results of a usability study can be affected by the use of different data entry methods (e.g., soft versus physical keyboards) (MacKenzie & Zhang, 1999; Soukoreff & MacKenzie, 1995; Zhang, 1998). Different user status (e.g., sitting versus walking; holding a device in hand or putting it on a table) while using a mobile device can further exacerbate the data entry problem.

There are also some other challenges. Today, multimodal mobile applications are emerging. Multimodality combines voice and touch (via a keypad or stylus) as input with relevant spoken output (e.g., users are able to hear synthesized, prerecorded streaming or live instructions, sounds and music on their mobile devices) and onscreen visual displays in order to enhance the mobile user experience and expand network operator service offerings. Blending multiple access channels provides new avenues of interaction to users, but it poses dramatic challenges to usability testing as well.

The above problems caused by physical restrictions of mobile devices and wireless networks imply that while designing and conducting usability studies for mobile applications, these issues must be carefully examined in order to select an appropriate research methodology and minimize the potential effect of contextual factors on perceived usability when they are not the focus of studies.

3. Major Research Questions of Usability Studies of Mobile Applications

There have been some interesting studies on the usability testing of mobile applications. In general, the surveyed studies mainly attempted to address one or several research questions (RQ) as follows:

RQ #1: Can proposed presentation methods help users easily search for/browse/ understand specific information of their interest on mobile devices? This line of research focuses on exploring and evaluating different approaches to effective content presentation on the small screen of mobile devices (e.g., Buyukkokten, Garcia-Molina, Paepcke, & Winograd, 2002; Masoodian & Lane, 2003). Some studies, for example, have reported that showing a combination of summary and keywords of each document at first is more efficient for users to locate relevant information from a list of documents than showing entire documents directly (Buyukkokten, Garcia-Molina, & Paepcke, 2001).

RQ #2: *What are appropriate designs of menu and link structures that help users reach a destination page easily (navigation)?* This category of research concentrates on how to design menus and link structures to make them sufficiently simple and straightforward (e.g., Chittaro & Cin, 2002). Several general guidelines for menu and link design have been suggested. First, menu choices should be 1) clear with easily interpretable labels, 2) consistent throughout a navigation site, and 3) predictable so that users can foresee what will be results of actions based on their past interaction history. Second, designers should avoid displaying a long list of choices on the screen so that users' cognitive load can be minimized. Third, a structure menu should facilitate users to finish tasks with minimum interaction with a device (e.g., scrolling, data tries, and button clicks).

RQ #3: *Can users easily carry out specific activities (e.g., query searching, filling form, making notes) of an application on mobile devices?* Researchers aim to investigate how easily users can perform a variety of activities on mobile devices. It has been reported that users are unlikely to experience similar degrees of comfort while using applications on mobile devices as they do on desktops due to constraints of mobile devices (Bederson et al., 2002; Bederson et al., 2003; Buyukokten et al., 2002; Jones et al., 1999; Kaasinen et al., 2000). Different interface solutions have been proposed to enable users to carry out an application on mobile devices more effectively, such as fitting information on one screen in order to avoid scrolling, using hierarchical menu, and providing ways to go back to an earlier page/directory. Moreover, from a user's perspective, providing personalized features (e.g., utilize user preferences to adapt application behavior to help users fulfill tasks) and display control appears to be effective to improve usability.

RQ #4: *What kind of data entry methods can enable users to enter data easily and quickly?* This line of research investigates the effectiveness and efficiency of various data entry methods (e.g., external keyboard, stylus, and soft keyboard) that help users enter data into mobile devices (e.g., Mackenzie & Zhang, 1999; Lee & Zhai, 2004). Each data entry method has its own pros and cons. For example, although entering data into a mobile device via a soft keyboard is more accurate in comparison with other input methods, it is not convenient when users are walking around. Speech recognition techniques are very helpful for data entry, especially for users with physical disabilities, but they may produce high error rates. Therefore, selecting a data entry method really depends on the context in which mobile devices and applications are used. Recently, multimodal access that integrates multiple data entry methods has been developed for mobile devices.

RQ #5: *How well can mobile applications be used, considering mobile context, mobility, and slow network connection?* This line of research focuses on investigating the usability of mobile applications while being used in different contexts. In a controlled laboratory experiment, a number of mobile contextual issues are ignored. Therefore, a usability test in a real environment can help ensure that a mobile application can work properly and help users achieve their goals in real-world situations (e.g., Rakkoklainen & Vainio, 2001).

A summary of the existing usability research on mobile applications in terms of research questions addressed and usability attributes used in those studies is presented in Table 1.

<Insert Table 1 here>

4. A Generic Framework for Usability Testing of Mobile Applications

In this section, a generic framework (Figure 1) is proposed based on the literature to facilitate researchers on conducting usability studies for mobile applications. The framework involves some major issues that researchers need to take into consideration while designing a usability test for a mobile application. This Section also provides some suggestions and insights on how to select appropriate research methodologies and deal with unique issues of mobile applications and context.

4.1 Research methodologies for usability testing of mobile applications

Two major methodologies that have been applied to usability testing of mobile applications are laboratory experiments and field studies. In a laboratory experiment, human participants are required to accomplish specific tasks using a mobile application in a controlled laboratory setting, while a field study allows users to use mobile applications in the real environment. Both methodologies have pros and cons. Therefore, selection of an appropriate methodology for a usability study depends on its objectives and usability attributes.

Laboratory experiments

There are several advantages of performing usability testing of mobile applications through controlled laboratory experiments (e.g., Bautsch-Vtense et al., 2001; Buchanan et al., 2001; Buyukkokten et al., 2002). First, a tester has full control over an experiment. He/She can define particular tasks and procedures that match the goal of a usability study, and ensure that participants follow experimental instructions. For example, if the objective of a study is to investigate the effectiveness of a data entry method while a user is moving around, then a laboratory experiment is more appropriate than a field study, because testers can explicitly require and ensure participants to use a mobile device while moving. Second, it is easy to measure usability attributes and interpret results through controlling other irrelevant variables in a laboratory environment. As a result, the laboratory experiment approach is very helpful to usability studies that focus on comparing multiple interface designs or data input mechanisms for mobile devices. Third, it makes it possible to use video or audio recording to capture participants' reaction (including emotions) when using an application (Dumas & Redish, 1999).

A major limitation of the laboratory testing method is that it ignores mobile context and unreliable connection of wireless networks. A mobile application tested in a real environment may not work as well as it does in a controlled laboratory setting due to the changing and unpredictable network conditions and other environmental factors. In a lab, participants may not experience the potential adverse effects of those contextual factors.

Field studies

A major advantage of conducting usability tests through field studies is that it takes dynamic mobile context and unreliable wireless networks into consideration, which are difficult to simulate in laboratory experiments. The perceived usability of a mobile application is derived based on participants' experience in a real environment, which is potentially more reliable and realistic compared to laboratory experiments (Kjeldskov & Stage, 2003; Palen & Salzman, 2002a; Sharples, Corlett, & Westmancott, 2002).

However, performing field studies for mobile applications is far from trivial. A major challenge of this methodology lies in the lack of sufficient control over participants in a study. There are three fundamental difficulties reported in the literature (Beck, Christiansen & Kjeldskov, 2003). First, it can be complicated to establish realistic environments that capture the richness of the mobile context. Second, it is not easy to apply established evaluation techniques such as observation and verbal protocol when a test is conducted in a field. Third, because users will physically move around in a dynamically changing environment, it is challenging for data collection and condition control. Therefore, in a field study, testers must define the scope of mobile contexts (e.g., physical body movement such as walking, standing, or sitting, and environment such as home/office, quiet/noisy, bright/dark) and use effective methods to collect data in the field.

Selection of Research Methodology

The selection of an appropriate research methodology for the usability testing of a mobile application depends on specific research questions and objectives. We argue that laboratory testing is more suitable for standalone mobile applications – those without the need of dealing with network connectivity. While designing and conducting a laboratory experiment for a mobile application that involves data transfer through a wireless network, the testers should focus on evaluating components of mobile applications, such as interface layout, information presentation schemes, design of menu and link structures, and data entry methods, that are not significantly influenced by mobility, network connectivity, and other contextual factors.

Field studies, on the other hand, are more appropriate for usability testing when major concerns are application performance related issues that are highly dependent on the mobile context. For example, it has shown that mobile context has strong effect on the usability of Internet surfing via mobile devices (Kim, Kim, Lee, Chae, & Choi, 2002). In addition, field studies are appropriate for studying user behavior and attitude toward mobile applications (Palen & Salzman, 2002a). For example, if a usability study attempts to examine the user perceived usefulness and efficiency of a mobile Web portal application, then a field study should be deployed in order to enable participants to provide feedback based on their experience with the system in a real-world setting.

As discussed above, both laboratory and field studies have distinct pros and cons. They really complement to each other in usability testing of mobile applications. Ideally, for a comprehensive usability study that examines a variety of issues such as interface design, user perceived ease-of-use and attitude, as well as application performance related measures, a hybrid approach that combines both laboratory and field studies should be considered, where different methods can be used for investigating different research questions. A research project that developed a ‘location finder’ mobile application (Rakkolainen & Vainio, 2001; Vainio & Kotala, 2002) is an example. This research project was aimed to facilitate users with mobile devices to locate themselves and search for directions to a specific place in a city. Usability testing of this application was performed through both laboratory and field studies. First, a laboratory experiment was conducted to evaluate the user interface (a 3D graphical design). In the experiment, an emulator on laptop computers was used. The tasks given to participants were to 1) find the nearest bookshop, 2) describe the current location during a short drive, and 3) locate all theaters at the center of the city. To avoid potential impact of network connection (mobile context), a local database of the city map was stored on the laptop locally. In addition, a field study to evaluate the usability of the application in the real environment was conducted. Fifteen participants were asked to use that mobile application on PDAs to search for a specific place and actually walk there. Participants were asked to think aloud during the study and their voice was recorded. They were also interviewed after finishing tasks. The results showed that although the application was perceived useful, the participants faced problems of slow downloading due to the low bandwidth of wireless connection.

4.2 Tools used in usability testing of mobile applications

Real mobile devices are used in field studies. Usability tests of mobile applications in laboratories can be carried out on either emulators or actual mobile devices. Both approaches have their pros and cons (Longoria, 2001). Using an emulator on a desktop computer enables testers to thoroughly capture user behavior such as the number of button clicks via software tools (Buyukkokten et al., 2002; Chittaro & Cin, 2002; Jones et al., 1999). The captured data are generally informative and useful for analyzing user performance and finding faulty designs of applications that frustrate users. However, using emulators omits some important aspects of actual mobile devices and mobile context. For example, it alleviates the problems of long transmission latency caused by limited bandwidth in real wireless networks, inefficient input mechanisms, and the changing wireless environment, potentially leading to untruthful user perception and satisfaction. We argue that emulators are more suitable to be used for improving the interface design of applications such as the layout of menu structures during the development process.

Testing an application on real mobile devices allows testers to collect more realistic information than testing on emulators, because users can test the application in a real environment. In comparison with the emulator testing, however, this approach has difficulty in capturing sufficient details of user behavior while users use a mobile application.

Mobile devices themselves, due to their unique, heterogeneous characteristics and physical constraints, may play a much more influential role in usability testing of mobile applications than desktop computers do in usability testing of desktop applications. Therefore, real mobile devices should be used whenever possible. Emulators may only be suitable for examining the usability of some device-related issues (e.g., interface design), while real mobile devices are more appropriate for finding usability problems involving mobile context. For example, a usability study of evaluating menu and link structure design for three different mobile phones (i.e., Nokia 3210, Siemens C35i, and Motorola P7389) was conducted in a laboratory experiment (Zifle, 2002). Emulators of three mobile phones running on a desktop were used. Participants were asked to solve six predefined tasks using three mobile phones within a time period. Usability attributes such as effectiveness (measured by the percentage of tasks solved), efficiency (measured by time used to solve tasks and number of clicks used to reach a destination page), and learnability (measured by the improvement in task performance in the second trial) were used to evaluate the menu and link structure design for each mobile phone. Results indicate that the basic principle of menu and link structure design for mobile applications is to minimize the number of clicks by users to reach destination pages.

4.3 Usability attributes

Usability attributes are various features that are used to measure the quality of applications. Based on the standard ISO 9241, HCI handbooks, and existing usability studies on mobile applications, there are nine generic usability attributes (Danesh et al., 2001; Frokjaer, Hertzum, & Hornbaek, 2000; Nielsen, 1993; Öquist & Goldstein, 2002; Ziefle, 2002): **learnability** focuses on how easily users can finish a task the first time using an application and how quickly users can improve their performance levels (i.e., ease-of-use); **efficiency** is defined as how fast users can accomplish a task while using an application. The difference between efficiency and learnability is that before measuring efficiency, users should have already had some experience of using a mobile application; **memorability** refers to the level of ease with which users can recall how to use an application after discontinuing its use for some time. The main idea is to measure how well users can re-establish the skill of using an application; **errors** can be measured by counting the number of mistakes that users make while using a mobile application, the severity levels of mistakes, and how easily users can correct them; **user satisfaction** reflects the attitude of users toward using a mobile application; **effectiveness** is defined as completeness and accuracy with which users achieve certain goals. It can be measured by comparing user performance with required levels. The effectiveness attribute is also used to assess the improvement of a new version of a mobile application; **simplicity** is the degree of comfort with which users find a way to accomplish tasks. This attribute is frequently used to assess the quality of menu structures as well as navigation design of mobile applications; **comprehensibility,** sometimes interchangeably with the term readability, measures how easily users can understand content presented on mobile devices. Because current mobile applications primarily deal with textual information, the presentation of information has significant effect on users' understanding of content; and **learning performance** measures the learning effectiveness of users in mobile education (using mobile applications to facilitate learning or communication with other learners or instructors).

In addition to the above nine usability attributes, there are several other commonly used attributes such as user perceived usefulness and system adaptability (Goldman, Pea, Maldonado, Martin, & White, 2004; Sharples et al., 2002). Different usability attributes may be best evaluated by different methods and variables. Selecting appropriate usability attributes to evaluate a mobile application depends on the nature of the mobile application and the objective(s) of the usability study. A variety of measures (e.g., time, speed, and number of button clicks) have been used to evaluate different usability attributes of specific mobile applications, as shown in Table 2.

4.4 Data collection methods

In comparison with field studies, data collection in laboratory experiments is usually much easier. Fundamental data collection methods such as observation, interview, survey questionnaire, and verbal protocol have been employed in usability testing of mobile applications (Goldman et al., 2004; Kjeldskov & Stage, 2003; Rakkolainen & Vainio, 2001).

It is challenging to collect data in a precise and timely manner in a field usability study of a mobile application. Researchers have developed new techniques for data collection in field studies, such as voicemail diaries (Palen & Salzman, 2002b) and pocket and web diaries (Kim et al., 2002). In the voicemail diary method, participants are required to call a dedicated voicemail line to report problems or provide suggestions about the use of mobile applications. By using information obtained through voicemail diaries, testers can analyze reported usability problems and feelings of participants about mobile applications. For the pocket and web diary method, participants are required to write down detailed information about mobile applications in their mobile devices, then connect to a pre-determined Web server and upload their notes on a daily basis. Testers can obtain information from the server during the study. If they have questions, they can contact participants for more information. In addition to those two methods, other approaches such as regular meetings, email reports, daily online questionnaires, and audio/video recorders can also be adopted for collecting data. Another issue about data collection in field studies is to find effective ways of reminding participants to report results or provide feedback timely because quality of data will solely (or almost solely) rely on responses from participants.

5. Conclusion and Future Research

With the rapid advances of mobile technology and applications, effective usability testing becomes increasingly important for the design, development, and deployment of successful mobile applications. However, due to unique features of mobile devices, limited bandwidth, unreliability of wireless networks, as well as other changing mobile context (e.g., location), traditional guidelines and methods used in usability testing of desktop applications may not be directly applicable to mobile applications. Therefore, it is essential to develop and adopt appropriate research methodologies and tools to evaluate the usability of mobile applications. This paper highlights major research questions and issues in this field, and proposes a generic framework built upon the past literature to guide the selection of research methodologies, usability attributes, mobile tools, and data collection methods for usability testing of mobile applications.

The latest advance of mobile technology and increasing wireless network bandwidth makes it a reality for users to gain access to multimedia information (combines several communication media such as text, graphics, video, animation and sound) available on the Internet or other sources from mobile devices. However, the constraints of mobile devices and wireless communication pose a variety of challenges for mobile devices to handle mobile multimedia applications (Smith, Mohan, & Li, 1999). For example, low wireless network bandwidth may cause significant transmission delay, which can affect both user perception and performance while using mobile applications. Although many mobile multimedia applications have used audio or video compression techniques to compress multimedia content in order to reduce the file size and shorten the transmission delay (Brachtl, Slajs, & Slavik, 2001; Smith et al., 1999), such data compression can result in reduced quality of multimedia content presented on mobile devices.

So far, most usability studies of mobile applications deal with either traditional database data or textual documents. Few studies have focused on usability testing of multimedia applications. This raises an interesting and even more challenging research question: *how can usability of mobile multimedia applications be evaluated effectively?* Based on previous studies of traditional multimedia applications in wired environments (Garzotto et al., 1993; 1994; 1995; 1998, Peterson, 1998), we suggest that usability testing of mobile multimedia applications should be planned from a QoS (quality of service) perspective. Researchers can adopt evaluation principles for usability of multimedia applications such as discipline, interactivity, quality, usefulness, and aesthetics (Heller et al., 2001).

References:

Bautsch - Vtense, H. S., Marmet, G. J., & Jacko, J. A. (2001). Investigating PDA web browsing through eye movement analysis. In *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality* (pp. 6-10). Mahwah, NJ: Lawrence Erlbaum Associates.

Beck, E. T., Christiansen, M. K., & Kjeldskov, J. (2003). Experimental Evaluation of Techniques for Usability Testing of Mobile Systems in a Laboratory Setting. *Proceedings of OzCHI 2003*, Brisbane, Australia.

Bederson, B. B., Clamage, A., Czerwinski, M. P., & Robertson, G. G. (2003, April 5-10). A fisheye calendar interface for PDAs: providing overviews for small displays. *Proceedings of CHI'03* extended abstracts on Human Factors in computing systems, Ft. Lauderdale, FL.

Bederson, B. B., Czerwinski, M. P., & Robertson, G. G. (2002). A Fisheye Calendar Interface for PDAs: Providing Overviews for Small Displays. Technical report (No. #HCIL-2002-09): University of Maryland College Park.

Bickmore, T. B., & Schilit, B. N. (1997, April 7-11). Digestor: Device-independent Access to the World Wide Web. *Proceedings of the 6th World Wide Web Conference*, Santa Clara, CA.

Björk, S., Holmquist, L. E., Redström, J., Bretan, I., Danielsson, R., Karlgren, J., et al. (1999). WEST: a Web browser for small terminals. *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, Asheville, North Carolina.

Björk, S., Redström, J., Ljungstrand, P., & Holmquist, L. E. (2000). POWERVIEW: Using information links and information views to navigate and visualize information on small displays. *Proceedings of Handheld and Ubiquitous Computing 2000 (HUC 2K)*, Bristol, U.K.

Brachtl, M., Slajs, J., & Slavik, P. (2001). PDA based navigation system for a 3D environment. *Computers & Graphics, 25*, 627-634.

Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., & Pazzani, M. J. (2001, May 1-5). Improving Mobile Internet Usability. *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong.

Buyukkokten, O., Garcia-Molina, H., & Paepcke, A. (2001, May 1-5). Seeing the whole in parts: text summarization for web browsing on handheld devices. *Proceedings of the tenth international conference on World Wide Web*, Hong Kong.

Buyukkokten, O., Garcia-Molina, H., Paepcke, A., & Winograd, T. (2002). Efficient Web Browsing on Handheld Devices Using Page and Form Summarization. *ACM Transaction on Information Systems, 20*(1), 82-115.

Chittaro, L., & Cin, P. D. (2002). Evaluating Interface Design Choices on WAP Phones: Navigation and Selection. *Personal and Ubiquitous Computing, 6*, 237-244.

Christie, J., Klein, R. M., & Watters, C. (2004). A comparison of simple hierarchy and grid metaphors for option layouts on small-size screens. *International Journal of Human-Computer Studies, 60*(5-6), 564-584.

Danesh, A., Inkpen, K., Lau, F., Shu, K., & Booth, K. (2001). Geney TM: Designing a Collaborative Activity for the Palm Handheld Computer. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, WA.

Dey, A. K., Salber, D., & Abowd, G. D. (2001). A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction, 16,* 2-4.

Dumas, J. S., & Redish, J. C. (1999). A Practical Guide to Usability Testing. *Intellect Book*. Portland: Intellect.

Ebling, M. R., & John, B. E. (2000). On the Contributions of Different Empirical Data in Usability Testing. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques*, New York.

Frokjaer, E., Hertzum, M., & Hornbaek, K. (2000, April 1-6). Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems, Hague, Netherlands.

Garzotto, F., Mainetti, L., & Paolini, P. (1993). HDM - A Model Based Approach to Hypermedia Application Design. *ACM Transaction on Information Systems, 11*(1), 1-26.

Garzotto, F., Mainetti, L., & Paolini, P. (1994). Adding Multimedia Collections to the Dexter Model. Proceedings of ACM Conference on Hypermedia Technology (ECHT'94), Edinburgh, UK.

Garzotto, F., Mainetti, L., & Paolini, P. (1995). Hypermedia design, analysis, and evaluation issues. *Communications of the ACM, 38,* 74-86.

Garzotto, F., Mainetti, L., & Paolini, P. (1998). Model-based Heuristic Evaluation of Hypermedia Usability. Proceedings of the working Conference on Advanced Visual Interfaces, L'Aquila, Italy.

Goldman, S. V., Pea, R., Maldonado, H., Martin, L., & White, T. (2004, March 23-25). Functioning in the Wireless Classroom. *Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'04)*, Taoyuan, Taiwan.

Gulliver, S. R., Serif, T., & Ghinea, G. (2004). Pervasive and standalone computing: the perceptual effects of variable multimedia quality. *International Journal of Human-Computer Studies, 60*(5-6), 640-665.

Heller, R. S., Martin, C. D., Haneef, N., & Gievska-Krliu, S. (2001, May 1-5). Using a theoretical multimedia taxonomy framework. *ACM Journal of Educational Resources in Computing, 1*(1), 6.

Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., & Buchanan, G. (1999, May 11-14). Improving Web Interaction on small displays. *Proceeding of the Eighth International Conference on World Wide Web*, Toronto, Canada.

Kaasinen, E., Aaltonen, M., Kolari, J., Melakoski, S., & Laakko, T. (2000). Two approaches to bringing Internet services to WAP devices. *Computer Network, 33,* 231-246.

Killi, K. (2002, August 29-30). Evaluation WAP Usability: "What Usability? *Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'02)*. Växjö, Sweden.

Kim, H., Kim, J., Lee, Y., Chae, M., & Choi, Y. (2002, January 07 - 10). An Empirical Study of the Use Contexts and Usability Problems in Mobile Internet. *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35'02),* Big Island, Hawaii.

Kim, L., & Albers, M. J. (2001). Web design issues when searching for information in a small screen display. Proceedings of the 19th Annual International Conference on Computer Documentation, Sante Fe, New Mexico, USA.

Kjeldskov, J., & Stage, J. (2003). New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies, 60*, 599-620.

Lee, P. U.-J., & Zhai, S. (2004). Top-down learning strategies: can they facilitate stylus keyboard learning? *International Journal of Human-Computer Studies, 60*(5-6), 585-598.

Li, V. O. K., & Liao, W. (2000, May 22 - 24). Wireless multimedia networks. *Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)*, Makati City, Metro Manila, Philippines.

Longoria, R. (2001). Designing Mobile Applications: Challenges, Methodologies, and Lessons Learned. In *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality* (pp. 91-95). New Jersey: Lawrence Erlbaum Associates Inc.

Luchini, K., Oehler, P., Quintana, C., & Soloway, E. (2001, August 6-8). An Engineering Process for Constructing Scaffold Work Environments to Support Student Inquiry: A Case Study in History. *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, Madison, WI, USA.

Luchini, K., Quintana, C., Krajcik, J., Farah, C., Nandihalli, N., Reese, K., et al. (2002, April 20-25). Scaffolding in the small: designing educational supports for concept mapping on handheld computers. *Proceedings of the Computer-Human Interaction (CHI) Conference on Human Factors in Computing Systems*, Minneapolis, MN.

Luchini, K., Quintana, C., & Soloway, E. (2003, April 5-10). Pocket PiCoMap: A Case Study in Designing and Assessing a Handheld Concept Mapping Tool for Learners. *Proceedings of the Computer-Human Interaction (CHI) Conference on Human Factors in Computing Systems*, Ft. Lauderdale, FL, USA.

MacKenzie, I. S., & Zhang, S. X. (1999, May 15-20). The design and evaluation of a high performance soft keyboard. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit (CHI'99),* Pittsburgh, Pennsylvania, USA.

MacKenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text Entry Using Soft Keyboards. *Behaviour & Information Technology, 18*, 235-244.

Masoodian, M., & Lane, N. (2003, February). An empirical study of textual and graphical travel itinerary visualization using mobile phones. Proceedings of the Fourth Australian user interface conference on User interfaces 2003, Adelaide, Australia.

Nielsen, J. (1993). *Usability Engineering*. New York: Academic.

Öquist, G., & Goldstein, M. (2002). Towards an Improved Readability on Mobile Devices: Evaluating Adaptive Rapid Serial Visual Presentation. *Interacting with Computers, 15*(4), 539-558.

Palen, L., & Salzman, M. (2002a). Beyond the handset: designing for wireless communications usability. *ACM Transactions on Computer-Human Interaction (TOCHI), 9*(2), 125 - 151.

Palen, L., & Salzman, M. (2002b, November 16-20). From methods to design: Voice-mail diary studies for naturalistic data capture under mobile conditions. *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, New Orleans, Louisiana, USA.

Parush, A., & Yuviler-Gavish, N. (2004). Web navigation structures in cellular phones: the depth/breadth trade-off issue. *International Journal of Human-Computer Studies, 60*(5-6), 753-770.

Petersen, M. G. (1998). Towards usability evaluation of multimedia applications. *Crossroads, 4*(4), 3-7.

Rakkolainen, I., & Vainio, T. (2001). A 3D City Info for mobile users. *Computers & Graphics, 25*, 619-625.

Sears, A., & Jacko, J. A. (2000). Understanding the Relation Between Network Quality of Service and the Usability of Distributed Multimedia Documents. *Human-Computer Interaction, 15*, 43-68.

Sharples, M., Corlett, D., & Westmancott, O. (2002). The Design and Implementation of a Mobile Learning Resource. *Personal and Ubiquitous Computing, 6*(3), 220-234.

Smith, J. R., Mohan, R., & Li, C.-S. (1999, October 30 - November 5). Scalable multimedia delivery for pervasive computing. *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, Orlando, Florida, United States.

Soukoreff, R. W., & MacKenzie, I. S. (1995). Theoretical upper and lower bounds on typing speed using a stylus and soft keyboard. *Behavior & Information Technology, 14*, 370-379.

Suwita, A., & Böcker:, M. (1999). Evaluating the Usability of the Siemens C10 Mobile Phone Going Beyond Common Practice in Industry. *Personal and Ubiquitous Computing, 3*(4), 173-181.

Vainio, T., & Kotala, O. (2002, October 19-23). Developing 3D Information Systems for Mobile Users: Some Usability Issues. *Proceedings of the second Nordic Conference on Human-Computer Interaction*, Aarhus, Denmark.

Varshey, U., & Vetter, R. (2002). Mobile Commerce: Framework, Applications and Networking Support. *Mobile Networks and Applications, 7*, 185-198.

Wichansky, A. (2000). Usability test in 2000 and beyond. *Ergonomic, 43*(7), 998-1006.

Zhang, D. (2003). Delivery of personalized and adaptive content to mobile devices: a framework and enabling technology. *Communications of AIS, 12*, 183-202.

Zhang, S. X. (1998). *A high performance soft keyboard for mobile system*. University of Guelph, Guelph, Ontario, Canada.

Ziefle, M. (2002). The influence of user expertise and phone complexity on performance, ease of user and learnability of different mobile phones. *Behavior & Information Technology, 21*(5), 303-311.

Table 1: Summary of Previous Usability Studies of Mobile Applications

RQ	Usability attributes	Device used	Research methodology	Implications
1	**Efficiency** (Bautsch – Vtense et al.,2001; Beck et al.,2003;Staffan Björk et al.,2000; Christie et al.,2004; L. Kim & Albers,2001; Masoodian & Lane,2003; Öquist & Goldstein,2002) **Errors** (Christie et al.,2004; Gulliver et al., 2004; L. Kim & Albers,2001; Masoodian & Lane,2003) **User satisfaction** (Beck et al.,2003; Staffan Björk et al.,1999;2000; Christie et al.,2004; Gulliver et al.,2004; Kaasinen et al.,2000; L. Kim & Albers,2001; Masoodian & Lane,2003; Suwita & Böcker:,1999) **Effectiveness** (Staffan Björk et al.,2000; Buchanan et al.,2001; Orkut Buyukkokten et al.,2001;2002; Christie et al.,2004) **Comprehensibility** (L. Kim & Albers,2001; Öquist & Goldstein,2002)	**Emulator: cell phone** (Buchanan et al.,2001; Kaasinen et al.,2000; Masoodian & Lane,2003), PDA (Staffan Björk et al., 1999; Orkut Buyukkokten et al.,2001;2002;Christie et al., 2004; L. Kim & Albers,2001) **Actual device: PDA** (Beck et al.,2003; Staffan Björk et al.,2000;, Gulliver et al.,2004; Kaasinen et al.,2000; Öquist & Goldstein,2002) **Actual device: cell phone** (Beck et al.,2003 ;Suwita & Böcker:,1999)	**Laboratory experiment** (Bautsch – Vtense et al.,2001; Beck et al.,2003; Staffan Björk et al., 1999;2000; Buchanan et al.,2001; Orkut Buyukkokten et al., 2001;2002; Christie et al.,2004; Gulliver et al.,2004; Kaasinen et al.,2000; L. Kim & Albers,2001; Masoodian & Lane, 2003, Öquist & Goldstein,2002, Suwita & Böcker:, 1999)	Different presentation methods can affect the efficiency of information seeking and browsing on mobile devices, leading to different levels of usability.
2	**Efficiency** (Chittaro & Cin,2002; Killi,2002; Parush & Yuviler-Gavish,2004; Ziefle,2002) **Errors** (Chittaro & Cin,2002; Killi,2002; Parush & Yuviler-Gavish,2004; Ziefle,2002) **User satisfactions** (Chittaro & Cin,2002; Suwita & Böcker:,1999; Ziefle,2002) **Learnability** (Killi,2002), Effectiveness (Chittaro & Cin,2002; Parush & Yuviler-Gavish,2004; Suwita & Böcker:,1999; Ziefle, 2002) **Simplicity** (Ziefle,2002)	**Actual device: cell phone** (Chittaro & Cin,2002; Killi, 2002; Suwita & Böcker:, 1999; Ziefle,2002) **Emulator: cell phone** (Parush & Yuviler-Gavish, 2004)	**Laboratory experiment** (Chittaro & Cin,2002; Killi,2002; Parush & Yuviler-Gavish,2004; Suwita & Böcker:, 1999; Ziefle, 2002)	Menu choices should be clear, consistent, and predictable. Link structure should be designed to help users finish tasks with fewer number of clicks

3	**Efficiency** (Bederson et al.,2002;2003; Jones et al.,1999; Kjeldskov & Stage,2004) **Errors** (Jones et al.,1999) **User satisfactions** (Bederson et al.,2002;2003; Jones et al.,1999; Kaasinen et al.,2000; Kjeldskov & Stage,2004) **Effectiveness** (Bederson et al.,2002;2003; Kaasinen et al.,2000)	**Actual device: PDA** (Kjeldskov & Stage,2004) **Emulator: cell phone** (Kaasinen et al.,2000) **Emulator: PDA** (Bederson et al., 2002;2003; Jones et al., 1999;Kaasinen et al.,2000)	**Laboratory experiment** (Bederson et al., 2002;2003; Jones et al.,1999, Kaasinen et al.,2000, Kjeldskov & Stage,2004)	Providing users with higher levels of control and personalized features to improve usability in mobile devices.
4	**Efficiency** (Lee & Zhai,2004; MacKenzie et al.,1999;MacKenzie & Zhang,1999; Soukoreff & MacKenzie,1995; D. Zhang,2003) **Memorability** (Lee & Zhai,2004; MacKenzie et al.,1999;MacKenzie & Zhang,1999; D. Zhang,2003) **Errors** (Lee & Zhai,2004; MacKenzie et al., 1999; MacKenzie & Zhang,1999; D. Zhang, 2003) **Learnability** (Lee & Zhai,2004; MacKenzie et al.,1999; MacKenzie & Zhang,1999; D. Zhang,2003) **Satisfaction** (Lee & Zhai,2004)	**Emulator: PDA** (Lee & Zhai, 2004;MacKenzie et al., 1999; MacKenzie & Zhang, 1999; Soukoreff & MacKenzie, 1995; D. Zhang, 2003)	**Laboratory experiment** (Lee & Zhai,2004; MacKenzie et al., 1999; MacKenzie & Zhang,1999; Soukoreff & MacKenzie,1995; D. Zhang,2003)	Multimodal interaction seems to be an effective data entry method for mobile devices.
5	**Usability problems (based on users' feedback)** (H. Kim et al.,2002; Palen & Salzman,2002a;2002b; Rakkolainen & Vainio,2001)	**Actual device: PDA** (Rakkolainen & Vainio, 2001) **Actual device: cell phone** (H. Kim et al., 2002; Palen & Salzman, 2002a;2002b)	**Field studies** (H. Kim et al., 2002;Palen & Salzman,2002a; 2002b; Rakkolainen & Vainio,2001; Vainio & Kotala, 2002)	Mobile context has major effect on user performance and satisfaction, so the information transmitted to mobile devices should be adapted according to the context

Note: The numbers in the first column 'RQ' represent 'Research Questions' discussed in Section 3.

Table 2. Measuring Usability Attributes In Mobile Applications	
Usability Attributes	**Measuring Variables**
Learnability	Time used to accomplish tasks at the first use (Killi, 2002; Parush & Yuviler-Gavish, 2004; Ziefle, 2002); time spent on training users until reaching a level of satisfaction (Killi, 2002); amount of training (e.g., number of trials, corrections) (Killi, 2002; Ziefle, 2002); learning curve of several uses (e.g., speed, errors) (MacKenzie et al.,1999; MacKenzie & Zhang, 1999; Parush & Yuviler-Gavish, 2004; Ziefle, 2002)
Efficiency	Task completion time(Bautsch-Vtense et al., 2001; Chittaro & Cin, 2002; Christie et al., 2004; Ebling & John, 2000; Killi, 2002; L. Kim & Albers, 2001; Kjeldskov & Stage, 2004; Ziefle, 2002), duration used to finish given exercises(Buchanan et al., 2001; Masoodian & Lane, 2003), the duration spent on each screen (Ebling & John, 2000; Kjeldskov & Stage, 2004; Öquist & Goldstein, 2002; Ziefle, 2002)
Memorability	Time, number of button clicks, pages, and steps used to finish tasks after not using applications for a period of time (e.g., 3 days or weeks) (MacKenzie et al., 1999; MacKenzie & Zhang, 1999; Öquist & Goldstein, 2002; Ziefle, 2002)
Error	Number of errors (e.g., detour steps, deviating button clicks from the right path, wrong answers, percentage of completed task correctly) (Chittaro & Cin, 2002; Christie et al., 2004; Gulliver et al., 2004; Jones et al., 1999; Killi, 2002; L. Kim & Albers, 2001; MacKenzie et al., 1999; MacKenzie & Zhang, 1999; Masoodian & Lane, 2003; Öquist & Goldstein, 2002)
Satisfaction	Attitude of users toward applications after using them (e.g., level of difficulty, confidence, like/dislike, etc.) (Bederson et al., 2002; 2003; Chittaro & Cin, 2002; Christie et al., 2004; Gulliver et al., 2004; Jones et al., 1999; Kaasinen et al., 2000; L. Kim & Albers, 2001; Kjeldskov & Stage, 2004; Masoodian & Lane, 2003; Nielsen, 1993; Suwita & Böcker:, 1999; Ziefle, 2002)
Effectiveness	Comparison of user performance with a predefined level (e.g., finishing tasks in 9 minutes, using no more than 2 clicks) in terms of speed (Bederson et al., 2002; 2003; Buchanan et al., 2001; Chittaro & Cin, 2002; Christie et al., 2004; Jones et al., 1999; Suwita & Böcker:, 1999; Ziefle, 2002), errors (Christie et al., 2004; Suwita & Böcker:, 1999; Ziefle, 2002), number of steps (Orkut Buyukkokten et al., 2001; 2002; Ebling & John, 2000; Suwita & Böcker:, 1999; Ziefle, 2002), task solved in a time limit (Chittaro & Cin, 2002; Jones et al., 1999; L. Kim & Albers, 2001)

Simplicity (Complexity)	Amount of effort to find a solution: numbers of menu levels that users have to go through in order to solve a task (Chittaro & Cin, 2002; Ziefle, 2002), numbers of button clicks and selections to reach a destination page (Chittaro & Cin, 2002; Christie et al., 2004; Ziefle, 2002), time used to search a button to perform a specific function (Buchanan et al., 2001; Ziefle, 2002)
Comprehensibility (Readability)	Reading speed (word/minutes) (Öquist & Goldstein, 2002) and percentage of correct answers in a predefined test (Kim & Albers, 2001; Öquist & Goldstein, 2002)
Learning Performance	Evaluation of assignments in classrooms (e.g., exercises, notes, concept maps) (Danesh et al., 2001; Luchini et al., 2001; 2002; 2003)

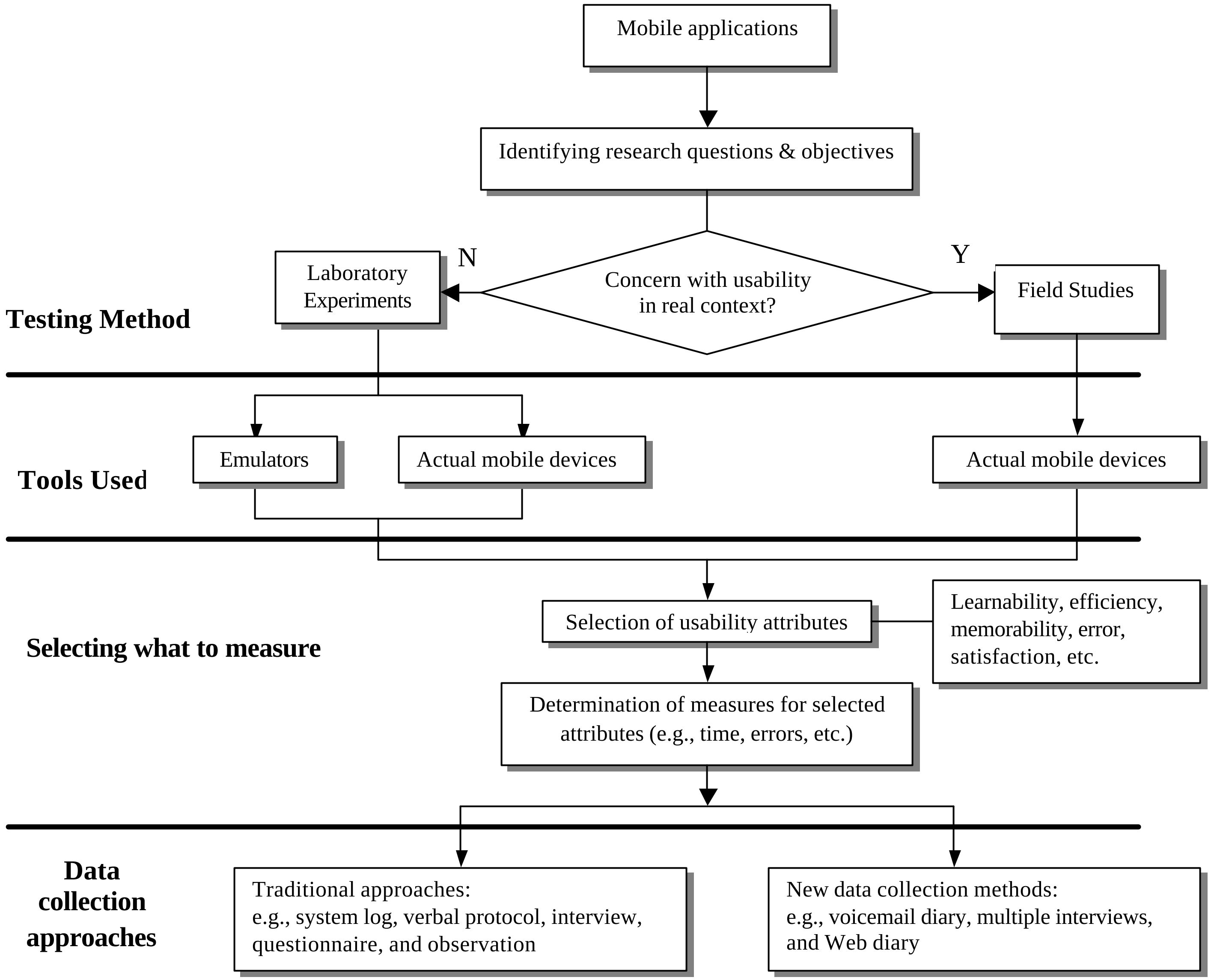

Figure 1. A Framework for the Design and Implementation of Usability Testing of Mobile Applications