Adaptation of Heuristic Evaluation for Mobile Applications and the Impact of Context of Use

Ger Joyce

Submitted to the University of Hertfordshire

in partial fulfilment of the requirements for the degree of Doctor of Philosophy

July 2020

School of Physics, Engineering and Computer Science

University of Hertfordshire

Dedicated to my wife, Denise.

ABSTRACT

The aim of this programme of research was to build upon prior work regarding the heuristic evaluation of mobile applications in two important areas. A mixed methods approach was taken to address the research questions using questionnaires, semi-structured interviews, document analysis and statistical analyses.

The first contribution of this programme of research was to adapt traditional usability heuristics, so they could be used to more effectively evaluate the usability of mobile applications. The resulting set of mobile application usability heuristics uncovered the highest number of usability issues in absolute terms, as well as the most critical issues within a well-known travel mobile application when compared to two other sets of usability heuristics. In addition, the set of mobile application usability heuristics defined within this programme of research was ranked as the most useful set by participants knowledgeable in the field of human-computer interaction when compared to two other sets of usability heuristics.

The second contribution of this programme of research was to extend the set of mobile application heuristics allowing human-computer interaction researchers and practitioners to more effectively consider the potential impact of context of use on the usability of mobile applications. This addressed a gap in the literature as few effective methods existed for this purpose. The protocol was deemed easy to use, easy to understand and easy to learn by participants knowledgeable in the field of human-computer interaction.

The beneficiaries of this programme of research are human-computer interaction researchers and practitioners, as well as mobile application users.

ACKNOWLEDGEMENTS

I would like to offer my thanks to those that helped me to complete this dissertation. First and foremost, my dissertation committee, Dr. Mariana Lilley, Dr. Trevor Barker and Prof. Amanda Jefferies, all of whom provided exceptional guidance over the years. In addition, I wish to thank my examination team, Dr. Nathan Baddoo from the University of Hertfordshire and Dr. Martin Maguire from Loughborough University, for their time during the examination process.

I would also like to thank Dr. Loi Sessions Goulet, Director, UX Research at TripAdvisor, and Dr. Mahtab Ghazizadeh, Head of Design Research at eBay, both of whom offered their support without which this dissertation would not have been possible.

I would also like to extend my gratitude to all of those that contributed to this dissertation, from those that critiqued submissions to conferences and journals and to those that participated in the various research studies conducted during the course of this programme of research.

Last, but not least, I wish to offer my heartfelt gratitude to my wife, Denise, to whom this dissertation is dedicated, for her support and patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. INTRODUCTION

Smartphones are relatively new, yet the number of smartphones sold globally have exceeded the number of personal computers and laptops sold in recent years (Rosoff, 2012; Taylor & Silver, 2019). Earlier handheld devices, such as feature phones, were less versatile (Lee, 2010), whereas smartphone users have substantially more functionality at their fingertips, including the ability to access mobile websites and install applications (Busk, 2011; Santhipriya et al., 2011). Even though mobile websites are less expensive to develop (McWherter & Gowell, 2012), they tend to have less visibility than mobile applications (Malan, 2011) as users need to open a mobile browser and type the web address of the website, which can be difficult on a smartphone keypad. Efforts to alleviate this are being developed, such as the introduction of quick response (QR) codes (Winter, 2011) (defined within Appendix A: Glossary). Despite these efforts, mobile applications continue to be more popular as they can work offline and have programmatic access to the camera, global positioning system (GPS), accelerometer, push notifications and other features of the mobile device that a mobile website cannot access (Hopkins & Turner, 2012).

Subsequently, billions of mobile applications have been downloaded (Murphy, 2012; Newton, 2012; Picoto et al., 2019), with 33.6 billion downloaded from Apple and Google app stores combined in the first quarter of 2020 (*Combined Global Apple App Store and Google Play App Downloads from 1st Quarter 2015 to 1st Quarter 2020*, 2020). Some of these mobile applications may be used occasionally, such as travel apps, while others are often used several times a week, such as calendars and email. Regardless of how many times they are used, mobile applications need to be usable (Amelung et al., 2010; Silva et al., 2014). Indeed, users have a public forum within which to report both easy to use and difficult to use mobile applications via reviews on app stores (Fu et al., 2013). Nevertheless, ensuring that a mobile application is usable is not without difficulty as smartphones are impacted by a plethora of constraints, which may also have an effect on the usability of mobile applications. For example, smartphone users have to contend with small screens, miniature keyboards, limited bandwidth, memory and processing power on devices that are often difficult to use in between ever-changing environments, such as strong sunlight or noisy indoor areas (Alkhafaji et al., 2019; Dunlop & Brewster, 2002; Joyce, 2014).

Despite the substantial differences with mobile computing, there are relatively few usability evaluation methods defined for mobile applications (Eshet & Bouwman, 2017) as most of the methods used were defined for desktop computing (Harrison et al., 2013; Salgado et al., 2016). To that end, exploring approaches that allow for an effective, holistic usability evaluation of mobile applications is an important area of research. Without effective ways to evaluate the usability of mobile applications, there is a higher likelihood that mobile applications will be difficult to use, resulting in frustrated users and lost revenue (Bloomberg, 2002; Nielsen, 2008; Sonnenberg, 2020). The concept of usability, therefore, forms the basis of this programme of research.

1.1 Defining the Concept of Usability

The first step in exploring the concept of usability is to look back several decades when software usability became an important commodity. This allows for a better understanding of the concept of usability and to ensure that it is defined in a satisfactory way. This, in turn, lends itself to a re-think of current approaches to evaluating software in that such approaches may not be adequate when evaluating the usability of mobile applications (Weichbroth, 2020; Zhang & Adipat, 2005). Additionally, both the opportunities and constraints of mobile devices, which are reflected in mobile applications, need to be considered when evaluating if a software artefact is usable or not.

Moving onto the definition of usability, human-computer interaction (HCI) is a science that encompasses elements of psychology, social sciences, behavioural sciences, and computer science that focuses on understanding how humans interact with computers. While the roots of this science have existed since World War II, HCI truly emerged as a scientific discipline during the 1970's and 1980's (Carroll, 2003; Dix et al., 2003). As far back as 1971, HCI researchers were concerned with the ease of use of software artefacts (Miller, 1971). During the late 1970's, the term 'usability' began to be used in a number of publications, following which attempts were made to define what it actually meant. One such definition originated from Boehm (1978), whereby the author defined software usability as the extent to which a product is convenient and practical to use. While this definition captured the essence of the term, many of those within the field of HCI were not content; evolving technology only made this matter more complex. For instance, as the World Wide Web became prevalent in the 1990's, the concept of usability was further extended by Nielsen (1993), who suggested that usability consisted of five attributes, namely learnability, effectiveness, tolerance for errors, satisfaction, and memorization. In 2002, Brinck et al. stated that they believed usability to mean a software artefact functions the way it should, tasks can be completed efficiently, and the software is easy to learn, tolerant of errors, and pleasing to use. In 2008, Rubin & Chisnell added to the discussion by suggesting that a product can only be deemed usable when a user can complete whatever they want their own way without difficulty or uncertainty. This definition was mirrored somewhat by Reiss (2012, p.xvii), who argued that "Usability deals with an individual's ability to accomplish specific tasks or achieve broader goals". Even the International Organization for Standardisation (ISO, n.d.), which creates global standards, became involved in an attempt to have one understanding of the term. Yet, ironically even the ISO has been inconsistent in defining what the term usability means (Abran et al., 2003). Over the years, this discussion has not abated (Coursaris & Kim, 2006), whereby Jeng (2005) and Nacheva (2020) asserted that the term usability has different meanings throughout the literature.

Most of the definitions of usability above do not include context of use even though this is a critical element of usability. In more recent years, Speicher (2015) argued that the consideration of usability must contain a product, specified users, specified goals, and a specified context of use. Without any one of these items, the author suggests that usability is not being evaluated. Indeed, the necessity to test digital products within the context that they are used is also specified within ISO 9241-210 (2010). Within that standard, it is stated "The extent to which products are usable and accessible depends on the context, i.e. the specified users, having specified goals, performing specified tasks in a specified environment". This definition is the lens through which this dissertation has been written. The next section considers how usability might be evaluated.

1.2 Evaluating Usability

Evaluating the usability of a software artefact is a critical phase in the software development lifecycle (SDLC) in that it allows for the assessment of a product by identifying usability issues. Such an evaluation offers two distinct benefits. Firstly, software developers know that they are creating quality work, which is one of the key motivators to staying with an organisation (Hall et al., 2008). Secondly, a usability evaluation helps to ensure that users can achieve their goals, which is the focus of this programme of research. To understand if a software artefact is usable, a usability evaluation should be carried out at least once prior to release (Abras et al., 2004;

Deraman & Salman, 2019). To accomplish this, there are multiple prolific usability evaluation approaches available (Vermeeren et al., 2010). Methods that have gained widespread use have various strengths and weaknesses that make them more suitable to different parts of the product development life cycle than others (Obrist & Roto, 2009). Additionally, a number of usability evaluation methods have been defined that involve representative users at their core, while other methods do not (Paz & Pow-Sang, 2016). The involvement, or lack thereof, of representative users does not make a usability evaluation method better or worse in itself as each approach has its time and place.

There are two main categories of usability evaluation types. When a usability evaluation method is used early in the design and development phase, it is known as a formative evaluation. Software artefacts and associated designs are less defined at this stage, so the evaluation tends to be carried out by usability experts as measuring constructs, such as user satisfaction, is not useful at this point. Instead, an evaluation should focus on ensuring that key usability guidelines are being met. Nielsen & Molich (1990) contend that formative usability evaluations are fast and inexpensive. Examples of formative usability evaluation methods are:

Table 1.1. Examples of formative usability evaluation methods

Method	**Strength(s)**	**Weakness(es)**
Cognitive walkthrough	Allows usability experts to evaluate defined user tasks (Blackmon, 2004)	Limited in that the scope is centred entirely on a single aspect, that being if the user interface is easy to learn (Wharton et al., 1994)
Heuristic evaluation	An easy method to employ at a low cost and can be carried out early in the development process (Nielsen & Molich, 1990)	Usability reports with evaluation results can have a lot of false positives (Hvannberg et al., 2007). Traditional heuristics lack context of use which is important to mobile computing (Varsaluoma, 2009)

Once a software artefact is more defined and is ready, or almost ready, for release to a target audience, measuring such constructs as time to complete tasks, satisfaction with the software and so on become more relevant. This is known as a summative evaluation. Usability evaluation methods used at this point should involve representative users (Norman & Draper, 1986; Privitera et al., 2019). It is also important at this stage to be clear on what usability elements to measure. Bevan et al. (2016, p. 276) argue that "almost any usability guideline (of which there are hundreds in the literature) could be treated as a measure". While this is true, summative evaluations should, at a minimum, measure the basic elements of usability as defined by ISO 9241-11 (1998), namely effectiveness, efficiency and satisfaction. (Maguire, 2001b) defines these elements as:

- *Effectiveness:* The success, or lack of, as users attempt to meet their goals
- *Efficiency:* The time it takes users to attempt their goals
- *Satisfaction:* The level of comfort and acceptability users have as they attempt their goals

There are a number of ways to measure each usability element. For instance, effectiveness might be measured by the number of tasks users have successfully completed; efficiency might be measured using time-on-task; and satisfaction might be measured using the level of satisfaction with features or the proportion of users complaining. A number of evaluation methods exist that can be used in this instance, such as those in Table 1.2:

Table 1.2. Examples of summative usability evaluation methods

Method	Strength(s)	Weakness(es)
Usability testing	Utilises real users in a laboratory and/or in the field as they would normally use applications (Rubin & Chisnell, 2008)	Cannot be employed until late in the development cycle, and it can be expensive and time-consuming (Abras et al., 2004)
Think aloud method	Evaluators can observe users and hear their thoughts as they use an application (Boren & Ramey, 2000)	Some participants in a think aloud usability evaluation might be embarrassed (Love, 2005). Evaluators focus on known problems (Nørgaard & Hornbæk, 2006)

Having contextualized the background to this work by defining the concept of usability and how it might be evaluated, it is important to consider this concept not only from the perspective of software in general, but from the viewpoint of mobile applications.

1.3 Evaluating the Usability of Mobile Applications

Evaluating the usability of mobile applications, in some respects, is no different to that of any software artefact. The basic concept is that a software user interface (UI) on a mobile device should be deemed usable prior to release to users (Cruz et al., 2019). As such, this reinforces the fact that the concept of usability must be understood before it is evaluated. Should HCI teams be unclear on this concept, there is no way to know if a mobile application is usable or not. Even when the concept of usability is clear, mobile computing brings its own unique challenges, which present significant challenges for usability experts (Bernhaupt et al., 2008; Rakotonirainy et al., 2000; Satyanarayanan et al., 2019). Yet, the majority of usability evaluation methods originated prior to the advent of smartphones. Even though evaluating in the laboratory using these methods is less expensive (Fiotakis et al., 2009), these methods were defined only to consider the usability of software artefacts from the perspective of a person that used a stationary device, such as a computer situated in the same well-lit environment each day (Kjeldskov & Stage, 2004). To that end, many of these methods do not and cannot account for the issues that smartphones users regularly contend with (Baber, 2009; Chittaro, 2011; Rahmati & Zhong, 2013). Consequently, it has been argued that traditional usability methods cannot be readily applied to mobile application evaluations (Beck et al., 2003).

Beginning in the early 2000's, HCI researchers began to call for new or adapted usability evaluation methods. Ketola & Röykkee (2001) recognized that the standard usability methods in use did not work well within the mobile domain. Around the same time, Po et al. (2004) called for future research into adapting traditional usability methods for the mobile domain. Simply adapting usability evaluation methods for the mobile domain, however, is not easy (Høegh et al., 2008). Following users of mobile devices, for example, around their home, work or in the street can be time-consuming, costly, uncomfortable and potentially even dangerous (Kaikkonen et al., 2005; Kjeldskov et al., 2004).

There are a number of ways to address this issue. First and foremost, it is important to know which areas of mobile application usability to focus on. There is much that might be evaluated if time and budgets were unlimited, however that is rarely the case. With this in mind, detailed mobile platform-specific guidelines were made available from Apple (*Apple Company Profile*, n.d.), Google (*Google Company Profile*, n.d.), Microsoft (*Microsoft Company Profile*, n.d.) and BlackBerry (*BlackBerry Company Profile*, n.d.) over the past decade, yet their focus tends to be on style and design, not on usability issues. Furthermore, some of these guidelines can be too specific, especially for enterprise-class native smartphone mobile applications built across multiple mobile operating systems (Joyce & Lilley, 2014). Consequently, HCI researchers and practitioners may benefit from a low-cost, effective, relatively fast usability evaluation method that considers issues unique to mobile applications. An approach to addressing this issue might be to create a new usability evaluation paradigm for native smartphone applications. However, completely new usability evaluation methods can take time to validate. An alternative approach would be to modify a well-known, tried-and-tested usability evaluation method for the mobile panorama (Joyce et al., 2014), which builds on over thirty years of existing research. Consequently, the goal of this programme of research is to modify an existing method for the usability evaluation of native smartphone applications. It is hoped that the modified method will form a framework to reliably evaluate the usability of mobile applications.

While there is no shortage of methods that might be modified, one well-known usability evaluation method that is low-cost, effective, and relatively fast is heuristic evaluation (Hollingsed & Novick, 2007). During a heuristic evaluation, between three and five usability experts review a UI using ten usability 'rules of thumb', following which each issue found is listed from a frequency and severity perspective (Nielsen & Landauer, 1993). All usability issues found are discussed by all evaluators prior to the creation of a single usability report. However, heuristic evaluation as originally defined does not consider issues unique to smartphone-deployed mobile applications, including the impact of context of use, which is critical within mobile computing (Baharuddin et al., 2013; Lubis et al., 2019). This needs to be considered within the aim of this programme of research.

1.4 AIM OF THIS PROGRAMME OF RESEARCH

The aim of this programme of research is to investigate how heuristic evaluation might be adapted to uncover the primary usability issues applicable to mobile applications. Further, the proposed method should be extended to consider the potential impact of context of use on the usability of mobile applications. The beneficiaries of this research effort are HCI educators, researchers, and practitioners, as well as individuals, teams and businesses that design and develop mobile applications, and ultimately mobile application users.

1.5 RESEARCH QUESTIONS

The research questions that will guide this programme of research are:

- RQ1. How might a set of mobile application usability heuristics be developed?
- RQ2. How might a set of mobile application usability heuristics be evaluated?
- RQ3. How might context of use be considered when evaluating mobile application usability?
- RQ4. How might a protocol that considers the impact of context of use on mobile application usability be evaluated?

1.6 STUDIES WITHIN THIS PROGRAMME OF RESEARCH

The studies within this programme of research were as follows:

Table 1.3. Studies related to the first research question

Research question	Method	Activity summary
RQ1. How might a set of mobile application usability heuristics be developed?		
SRQ1.1 How are mobile application usability heuristics currently being used?	Interviews	Thirteen participants were interviewed to learn more about the current usage of heuristic evaluation for mobile applications
SRQ1.2 What factors should be included in a set of mobile application usability heuristics?	Document Analysis	An exploration of the literature uncovered characteristics used to form an initial set of usability heuristics relevant to mobile applications

Table 1.4. Studies related to the second research question

Research question	Method	Activity summary
RQ2. How might a set of mobile application usability heuristics be evaluated?		
SRQ2.1 What is the attitude of those knowledgeable in HCI to the mobile application usability heuristics?	Questionnaire	Sixty participants offered feedback on an initial set of heuristics for mobile applications
SRQ2.2 How might the set of mobile application usability heuristics be compared to other heuristic sets?	Heuristic evaluation and evaluation of heuristics	Six participants took part in a heuristic evaluation and an evaluation of heuristics, which compared the mobile application usability heuristics defined within this programme of research to two other sets of heuristics

Table 1.5. Studies related to the third research question

Research question	Method	Activity summary
RQ3. How might context of use be considered when evaluating mobile application usability?		
SRQ3.1 How is context of use currently considered when evaluating mobile application usability?	Questionnaire	149 participants offered their thoughts on areas of importance in relation to understanding context of use and how they consider context of use when evaluating mobile application usability
SRQ3.2 What is a satisfactory protocol that allows for the consideration of context of use when evaluating mobile application usability?	Document Analysis	A protocol that allowed for an understanding of the impact of context of use on mobile application usability was defined based on other protocols within the literature that reduced complexity

Table 1.6. Studies related to the fourth research question

Research question	Method	Activity summary
RQ4. How might a protocol that considers the impact of context of use on mobile application usability be evaluated?		
SRQ4.1 How well does the protocol consider the actual contexts of use that mobile applications are used in?	Interviews	Thirty semi-structured interviews were conducted to understand if the proposed protocol was able to handle diverse user stories based on real-world contexts of use
SRQ4.2 What is the attitude of those knowledgeable in HCI to the contextual usability evaluation method?	Interviews and questionnaire	Twenty semi-structured interviews were conducted to gather the attitudes of those knowledgeable in HCI towards the proposed protocol
SRQ4.3 How might those knowledgeable in HCI determine the scope of a typical contextual usability evaluation?	Workshop	A workshop was conducted with eight participants to understand how those knowledgeable in HCI might determine the scope of a typical contextual usability evaluation

1.7 Contribution to Knowledge

There are two primary contributions from this programme of research. Firstly, a set of heuristics that can be used to evaluate the usability of mobile applications was defined. Secondly, the defined set of mobile application heuristics were extended to form the basis of a protocol that can be used by HCI researchers, educators and practitioners when considering the potential impact of context of use on the usability of mobile applications.

1.8 Scope of this Programme of Research

This dissertation is focused on smartphone-deployed mobile applications, not mobile websites. The focus is limited to task-based mobile applications, whereby users attempt a task to complete a goal. As such, other types of mobile applications, including games, are outside of the scope of this programme of research. Further, those that are involved within HCI nowadays are diverse, encompassing students, educators, researchers, and practitioners. Given the difficulty in involving all disparate groups, this programme of research focuses predominantly on HCI researchers and practitioners.

1.9 Outline of this Programme of Research

This programme of research is divided into eight chapters. This is the first chapter, which focuses on the background and motivation for investigating how heuristic evaluation might be adapted to uncover the primary usability issues applicable to mobile applications, while concurrently considering the impact of context of use on mobile application usability.

Chapter two considers the literature in order to critically review what has already been done by other researchers. Further, the purpose of chapter two is to demonstrate how this programme of research fits into the bigger picture within two main topics that this work ties together, namely heuristic evaluation and context of use, as well as how they fit together from the perspective of mobile application usability.

Chapter three provides a methodology to address the research questions. My philosophical perspective is also covered, as are the importance of research ethics. In addition, the overall conceptual framework that guides the research, as well as the limitations of this dissertation are discussed.

Chapters four and five explore the data gathered and analysed, as well as reporting on research findings and follow-up discussion for the first two research questions. These research questions address the topic of developing and evaluating mobile application usability heuristics.

Chapters six and seven shift the focus to the data gathered and analysed, research findings and follow-up discussion for the second two research questions, namely how context of use might be considered when evaluating the usability of mobile applications.

Finally, chapter eight summarises the programme of research as a whole and presents my final thoughts on the mobile application usability heuristics and context of use protocol defined within this work. Implications for theory and practice, further research opportunities, and the limitations of the work are also considered. The chapter concludes with a review of what went well and what might be improved when conducting future research studies.

Chapter 2. LITERATURE REVIEW

The previous chapter focused on the aim of this programme of research, which is to investigate how heuristic evaluation might be adapted and extended to uncover the primary usability issues applicable to mobile applications. As usability issues uncovered by a set of heuristics may differ depending on changing contexts of use within which mobile applications are used, the modified approach should also consider the potential impact of context of use. The literature review, therefore, chronologically explores the theoretical underpinnings of heuristic evaluation. This is followed by the evolution of heuristic evaluation for mobile applications from the definition of usability guidelines, which is in turn followed by the more recent consideration of context of use on mobile application usability. This programme of research builds upon prior work on this topic, not by an exhaustive review of all related articles, rather by critically evaluating the seminal articles, thus leading to gaps in knowledge.

2.1 Theoretical Underpinnings

A substantial amount of theoretical work has been conducted since the 1960's on better understanding what an evaluation is, who is involved, when and why an evaluation should be carried out, as well as the various types of evaluations. Since then, several definitions of the term 'evaluation' have been published. Lincoln & Guba (1980), for instance, maintained that an evaluation is a study that places value on an object. Trochim's (1998) definition goes further by suggesting that evaluations are used to gather evidence to help make decisions within environments that often lack appropriate resources and where people are often under time pressure. While this definition was not created from the perspective of researching, designing and developing mobile applications, it is surprising how much overlap there is within the fast-paced world of mobile applications. For instance, Cottingham & Snyder (2011, p. 17) stated "With accelerated [mobile application] product design, user researchers must modify traditional methodologies to provide results in a much shorter time frame".

Even with an understanding of what an evaluation is, it is still important to better understand the concept as the primary constructs in the definition from Lincoln & Guba (1980), namely 'object' and 'value' are quite broad. To differentiate any object in existence, the term 'evaluand' was used by Scriven (2003) to describe an object that can be subjected to an evaluation. Pertinent to this dissertation, an evaluand may also include a software artefact (Mathison, 2005). Regarding value, Stufflebeam & Coryn (2014) asserted that an evaluation assesses the worth or merit of an evaluand, which the authors stated may include the usability of an evaluand. Worth and merit were both deconstructed by Mertens & Wilson (2012, p. 6), whereby the authors claimed that "merit is the absolute or relative quality of something, either intrinsically or in regard to a particular criterion" and "worth is an outcome of an evaluation and refers to the evaluand's value in a particular context". In relation to the latter, the authors emphasized that evaluative data without contextual variables was not enough to assess the merit and worth of an evaluand. The inclusion of contextual variables in an evaluation has been considered within the CIPP (Context, Input, Process, Product) model from Stufflebeam (2000b). Using this model, an evaluator can consider the environment within which an evaluand is being assessed (context), which approach is best suited to the evaluation (input), the process that will be followed, including any limitations to the scope of the evaluation (process), and the output(s) of the evaluation and a determination of how well the evaluand meets specified criteria in terms of merit and worth, while taking the context, input, and process aspects of the evaluation into account (product).

With a better understanding of the constructs ‘evaluand’, ‘merit’ and ‘worth’, as well as elements that constitute an evaluation using the CIPP model, evaluation theory continues to be quite complex. Stufflebeam (2001), for instance, has suggested that there are twenty-two types of evaluations. To further facilitate the definition of a methodology that met the aim of this programme of research, as well as to assist in sense-making of the findings, it was necessary to establish which types of evaluation were best suited to this work. Of the twenty-two approaches, the most applicable was ‘Approach 17: Consumer-oriented studies’, a type of evaluation that Stufflebeam (2001, p. 59), believed would “help to produce and deliver products and services of excellent quality and of great use to consumers”. The author also claimed that a method to define an evaluand’s merit and worth using a consumer-oriented study was to use a checklist framed by an expert. The latter point was interesting in that Stufflebeam (2001) inferred that consumers themselves were not included within the approach, despite the name of the evaluation type. Further, the author maintained that a fellow theorist, namely Scriven (1996), insisted that the purpose of a consumer-oriented study was a final summative judgement of an evaluand’s merit and worth, not a formative judgement that led to continuous improvement as the evaluand was being developed. The lack of consumer involvement in a consumer-oriented study, as well as the insistence that a consumer-oriented evaluation must be summative in nature were interesting points of view. Each of these points were re-visited at a later stage in the programme of research when implications for theory were discussed.

Having deconstructed much of evaluation theory as it pertained to this programme of research, one specific aspect of the underlying theoretical foundation guided a practical aspect of this work. That is, assessing an evaluand’s merit and worth through the use of a checklist framed by an expert, which assists in the creation of improved products, such as mobile applications, that are useful and usable to consumers. When considering the assessment of an evaluand's merit and worth using a checklist, it was useful to bear in mind that there was nothing new about checklists—it has even been argued that the Ten Commandments was a checklist (Stufflebeam, 2000c). Today, checklists are common across a number of industries outside of HCI, such as the medical and aviation fields (Gawande, 2010). Done well, checklists are a useful and easy to use device for evaluating an artefact, be it a program or a product, that allow evaluators of all experience levels to remember all elements of an evaluation. Consequently, a checklist ensures that both novice and experienced evaluators review all key areas of the program or product, which increases the overall quality of the evaluation, in particular in terms of reliability, validity, and credibility. It is, therefore, not at all surprising that an evaluation theorist has created a checklist on creating a checklist. Stufflebeam (2000a) created such a checklist, which the author referred to as a checklist development checklist (CDC). The steps within the CDC ensure that there is a solid understanding regarding the reason(s) for defining a checklist. Training, experience, and relevant literature should be used to create a list of potential items for a checklist, which are then classified and sorted as needed. At this stage, potential users can critique the checklist, after which the checklist can be revised, if appropriate. The checklist can then be field-tested by potential users, after which it can be revised once more if needed, then shared widely. It is likely that further feedback will be received at this stage, thus the checklist might be reviewed occasionally.

Yet, there are different types of checklists. This is useful to bear in mind when defining the aim of a checklist per Stufflebeam's (2000a) CDC. Some checklists are simply created to remember various items. In terms of an evaluation, Scriven (2005) argued that a criteria of merit checklist, which the author referred to as a 'comlist', was the most important type. According to the author, comlists refer to criteria that are commensurable (items are all at the same level), clear, concise and confirmable (e.g. measurable), and that are as complete as possible without overlapping items. As the evolution of heuristic evaluation for mobile application was considered within the literature review, it became clear that guidance from evaluation theorists was not always followed.

As the theoretical lens of evaluation theory was being applied to this programme of research, care was taken to heed words of caution from an HCI practitioner regarding their perspective of theory. Rogers (2004, pp. 3-4), an author of multiple books and articles on HCI, warned that "you cannot simply lift theories out of an established field…then reapply them to explain other kinds of seemingly related phenomena in a different domain (i.e. interacting with computers)…the kinds of cognitive processes that are studied in basic research are quite different from what happens in the 'real' world of human-computer interactions". That is not to say that the application of theory does not apply to HCI. The author's point was that theory from a more established field, such as psychology, aspects of which may have been defined within controlled laboratory settings, should not be applied blindly to a field as new and less established as HCI, much of which is applied within real-world settings with substantially less control. Arguably, this includes evaluation theory, which has predominantly been applied to program and policy evaluation, not heuristic evaluation. Even while taking these words of caution into account, much was learned from the application of evaluation theory to this programme of research. The reverse was also true. Thus, within the final chapter of this programme of research the implications of using such theoretical underpinnings was discussed, including how theory helped and hindered when applied to the development of heuristics for the usability evaluation of mobile applications.

2.2 Evolution of Heuristic Evaluation for Mobile Applications

The use of heuristic evaluation to evaluate the usability of mobile applications is a relatively new approach with a long history. During the 1980's, long before the advent of heuristic evaluation and mobile devices, desktop computing became prevalent. For the first time, sophisticated computing devices came out of specialist labs and into the hands of non-technical users. To ensure that technical software developers were in a position to create software for non-technical users, usability principles and guidelines were developed (Gould & Lewis, 1985). As such, these sets of guidelines allowed software developers to shift their focus from lines of code to users of the software and the tasks those users needed to do. Many of these guidelines were long categorized lists in no particular order, which was less than ideal for usability evaluation. For instance, Smith & Mosier (1986) published a set of 944 usability guidelines. Such a list lacked the brevity important within Scriven's (2005) definition of comlists, did not consider the contextual elements required as critical as specified by Mertens & Wilson (2012), and were simply unworkable and too costly to fully implement (Quinn, 1996).

In a bid to cut costs and to ensure that sets of usability guidelines were easier to work with, researchers began to define shorter sets that addressed the most common usability issues. Known as inspection methods, examples included Gerhardt-Powals' (1996) cognitive engineering principles, Shneiderman's (1983) eight golden rules of interface design, and Nielsen & Molich's (1990) nine usability interface heuristics. The latter, referred to as heuristic evaluation, was later updated by Nielsen (1994) resulting in a set of ten heuristics. Heuristic evaluation is more flexible than other methods, such as cognitive walkthrough (Wharton et al., 1994), allowing the evaluation of any screen on a software application or other digital artefacts based on the task(s) a user might attempt (Jaspers, 2009). As such, the method is holistic in nature allowing the evaluation of a full software application across the entire product design and development life cycle. Two weaknesses of usability testing are mitigated with heuristic evaluation, namely the tendency to be narrow in focus (Dicks, 2002), and that usability testing is used relatively late in the SDLC (Jeffries et al., 1991). In addition to these strengths, research revealed that the method found more usability problems when compared to other methods (Jeffries & Desurvire, 1992), after which the method became popular (Alonso-Ríos et al., 2018).

Despite this, several HCI researchers have argued that heuristic evaluation results are subjective (Kirmani & Rajasekaran, 2007) and that the method may not be as effective as it claims (Cockton & Woolrych, 2002; Law & Hvannberg, 2004) as heuristic evaluation did not find the same issues as usability testing. These arguments against heuristic evaluation have been counter-argued, whereby other researchers have suggested that no usability evaluation method is perfect. Furthermore, alternative viewpoints argue that any differences in the perspective of usability issues found during an evaluation is actually a benefit, especially when multiple evaluation methods are used. Such approaches enable the discovery of more diverse usability issues (Wilson, 2013), leading to the continued popularity of the method (Fernandez et al., 2012; Jeddi et al., 2020; Paz & Pow-Sang, 2016; Rusu et al., 2019).

As mobile devices have become more ubiquitous, heuristic evaluation is one of many methods that have been used to evaluate the usability of mobile applications. In an interview with Jenny Preece (Rogers et al., 2011) Jakob Nielsen, one of the original creators of heuristic evaluation, stated “You can identify a lot of issues with a phone or other mobile user experience by using exactly the same heuristics as you would for any other platform”. Indeed, the majority of heuristic evaluations of mobile applications are conducted using Nielsen's (1994) heuristics (Salgado & Freire, 2014). Yet, there are extensive dissimilarities between the desktop-based websites of the early 1990's and the complex mobile applications of today.

Consequently, since 2003 a number of researchers have defined sets of heuristics that can be used to evaluate the usability of mobile UI's. A search of Google Scholar using two sets of terms, namely 'handheld' AND 'usability' AND 'guidelines', as well as 'mobile' AND 'usability' AND 'heuristics' for articles published between 1994 and 2020, uncovered forty-two articles containing sets of mobile application heuristics. The majority of these sets of heuristics have been based on Nielsen's (1994) set, which is a common approach when developing sets of usability heuristics within and outside the mobile domain (Pierre, 2015). A critique of selected sets of mobile application usability heuristics uncovered from the search of Google Scholar discovered if research on this topic was complete or if gaps in knowledge needed to be filled. The critique comprised of adherence to the principles of evaluation theory, the application of a reliable and valid methodological approach, as well as the evaluative scope of each set of heuristics. Sets of heuristics that were specific and not focused on general, task-based mobile applications were removed from the critique as they were outside the scope of this programme of research. These included Korhonen & Koivisto's (2006) set of playability heuristics for mobile games, Kuparinen et al.'s (2013) usability heuristics for mobile map applications, Kumar & Goundar's (2019) usability heuristics for mobile learning applications, Maguire's (2019) heuristic evaluation tool for voice user interfaces, and Muhanna et al.'s (2020) heuristics for the evaluation of Arabic interfaces. The critique of the remaining sets of mobile heuristics follows:

2.2.1 *Handheld Usability Guidelines - Weiss (2003)*

For the first set of heuristics, the handheld usability guidelines from Weiss (2003), the author did not present empirical evidence relating to how these particular heuristics were defined. As such Weiss' (2003) guidelines did not consider many aspects of evaluation theory. Consequently, the guidelines seemed to be based on the author's opinion and was not critiqued further.

2.2.2 *Mobile Application Heuristics - Bertini et al. (2006)*

A set of mobile application usability heuristics was defined by Bertini et al. (2006) by consolidating information from a number of sources, including the collection of mobile evaluation issues and the consideration of which of Nielsen's (1994) heuristics were applicable to mobile usability. Feedback was also sought by experts, and as such, Bertini et al.'s (2006) approach was consistent with aspects of Stufflebeam's (2000a) CDC. However, the authors were able to gather feedback from only eight HCI experts. Not only was this a relatively small number, the authors also did not specify what they considered to be an HCI expert. Regarding the scope of the set of heuristics, much emphasis was placed on elements less relevant to a mobile application, such as the ergonomics of a mobile device. For instance, 'Heuristic 1: Visibility of system status and losability/findability of the mobile device' and 'Heuristic 4: Good ergonomics and minimalist design' kept aspects of Nielsen's (1994) traditional desktop heuristics and combined them with the 'losability/findability of the mobile device' and 'good ergonomics' respectfully. Yet, it is unlikely that mobile application designers and developers have influence over the 'losability/findability of the mobile device' they are designing a mobile application for nor can they influence the 'good ergonomics' of any of the many mobile devices in use by millions of smartphone users worldwide. The perplexity of combined software and hardware issues called into question the validity of the entire set of heuristics from Bertini et al. (2006).

Continuing the critique, 'Heuristic 3: Consistency and mapping' from Bertini et al. (2006), described the importance of consistency in mapping user interactions. While important, this heuristic and others in the set did not consider experimentation and technological advances within the field of mobile computing. One example might have been to address consistent mapping in user interactions, while also introducing alternative ideas as mobile technology is modernised. 'Heuristic 6: Flexibility, efficiency of use and personalization' covered an important aspect in regard to the ability for a user to configure a mobile application to their needs. Yet, the heuristics also stated that the mobile application should be able to customize itself. While this was certainly an option, it was important that users believed that they are in control at all times, which Shneiderman (1983) referred to as a locus of control. 'Heuristic 8: Realistic error management' combines two of Nielsen's (1994) error-related heuristics. A combination of two related heuristics was welcome when a set of heuristics needs to be concise. However, the heuristic title was confusing in that the authors did not clearly define what 'Realistic error management' actually meant. While the heuristic description covered the basics of aiding the user should an error occur, what exactly does 'realistic' mean? This seemed to be quite subjective.

Furthermore, the heuristics from Bertini et al. (2006) did not focus on a number of important areas within mobile computing. For example, while the authors removed Nielsen's (1994) heuristic regarding 'Help and Documentation', deeming it less applicable to mobile applications, they did not consider alternatives. Such alternatives might have included onboarding screens displaying the main features and showing users how to interact with the application, thus allowing first-time users to get up-and-running quickly (Clark, 2010). Additionally, the authors made no mention of the utilization of increasingly complex sensors that are available in today's complex smartphones. Such utilization may provide users with more interesting and stimulating experiences, while reducing the burden on users. Finally, Bertini et al. (2006) placed limited focus on context of use.

2.2.3 *Usability Checklist for the Usability Evaluation of Mobile Phone User Interface - Ji et al. (2006)*

The approach from Ji et al. (2006) used a unique approach by comparing items from style guides to usability issues gathered from the literature—yet, both are quite different in nature. The authors conducted a comparison test using pairwise comparison, which resulted in a number of low scores. These low scores were discounted, thus making the entire approach questionable. Further, once a set of heuristics was defined, the authors validated their checklist by comparing the issues found using the heuristics against those found using usability testing. Despite some overlap, both usability testing and heuristic evaluation resulted in different types of usability problems (Jeffries and Desurvire, 1992; Wilson, 2013). Additionally, study participants had no HCI experience, which tends to compare less favourably to those with HCI experience in regard to the quantity and quality of issues uncovered (Nielsen, 1992). Finally, Ji et al. (2006) defined twenty-one usability principles, which did not consider Scriven's (2005) call for brevity to ensure that lists are concise and workable. Furthermore, Sharpe et al. (2007) claimed that the maximum number of heuristics should be ten. Despite the large number of heuristics, the issue of context of use was not mentioned within the article from Ji et al. (2006), thus recommendations from Stufflebeam's (2000b) CIPP model were not considered.

2.2.4 *Heuristics for the Assessment of Interfaces of Mobile Devices - Machado Neto & Pimentel (2013)*

In previous examples above, each approach taken to define mobile applications has been different. The approach taken by Machado Neto & Pimentel (2013) also differed in that the authors initially reviewed four mobile applications that were popular at that time. Following this, the authors compared issues found to Nielsen's (1994) heuristics to see which issues fit and which did not. Where they did not fit, new heuristics were defined. The authors then conducted a heuristic evaluation on a mobile application with five usability experts using Nielsen's (1994) heuristics and five usability experts using their new mobile heuristics. The resulting heuristics were either no different to Nielsen's (1994) set, such as 'Consistency and standards' or 'Help and documentation' or were reworded slightly. To improve their set of mobile application usability heuristics, the authors might have considered aspects of evaluation theory, including Stufflebeam's (2000a) CDC. Furthermore, the topic of context of use has been overlooked within the set of heuristics.

2.2.5 *Heuristic Evaluation on Mobile Interfaces: A New Checklist - Gómez et al. (2014)*

Despite the call for modified usability heuristics for the mobile domain, Gómez et al. (2014) did not challenge Nielsen's (1994) heuristics. Thus, the authors' first ten heuristics did not differ from Nielsen's (1994) set. The authors simply added three new heuristics to Nielsen's (1994) set, namely, 'Skills', 'Pleasurable and respectful interaction with the user', and 'Privacy'. Even though the majority of heuristics did not change, the authors added sub-heuristics to all thirteen heuristics. These were based on best practices and usability tests that differ from the types of issues commonly found during a heuristic evaluation (Wilson, 2013). This led to a large set of seventy-two sub-heuristics, which did not follow the theoretical principle of brevity (Scriven, 2005) and was simply going back in time to the long lists of usability guidelines of the 1990's. Additionally, many of the sub-heuristics were quite specific making the whole set less future-proofed. For instance, sub-heuristic 11 stated that "swipe ambiguity should be avoided: the same swipe gesture should not be used to mean different things on different areas of the same screen". Yet, Apple (*Apple Company Profile*, n.d.), one of the largest manufacturers of mobile devices, encourage mobile application designers and developers to use the same swipe gesture on different parts of the same iOS screen where it means different things—a swipe down from the middle of the home page results in a search box appearing, a swipe down from the top right of the home page results in the control centre appearing, while a swipe down from the middle top of the home page results in viewing the notification centre (*Use Gestures to Navigate Your IPhone X and Later*, n.d.).

Another best practice was mentioned in sub-heuristic 6, namely if an article spans several pages, use pagination at the bottom. Nevertheless, several popular mobile applications use infinite scrolling not pagination, such as Twitter, LinkedIn, Pinterest, Facebook, and Instagram. In such cases, Loranger (2014, p. 1) points out that "Endless scrolling saves people from having to attend to the mechanics of pagination in browsing tasks". Moving on, sub-heuristic number 40 stated that flash content needed to be removed. Such examples do not help to futureproof the heuristic set and ensure that it can be used for decades, as has Nielsen's (1994) set. Indeed, shortly after the publication of the heuristics from Gómez et al. (2014), Adobe (*Adobe Company Profile*, n.d.), the creators of Flash no longer supported the product (Barrett, 2017).

A welcome aspect of the work from Gómez et al. (2014) was that the authors sought to validate their set of heuristics by conducting a heuristic evaluation of a mobile application, not by comparing heuristic evaluation results to usability testing results as had other researchers. However, the authors did not recruit HCI researchers nor practitioners that were experienced in conducting such evaluations, which would have likely increased the quality and quantity of issues found (Nielsen, 1992). Rather the authors "used two software engineering students that had never been trained in usability" (Gómez et al., 2014, p. 14). Finally, the issue of context of use was not considered despite its criticality within the mobile domain. As such, aspects of evaluation theory, including Stufflebeam's (2000a) CDC and Stufflebeam's (2000b) CIPP model were not taken into account.

2.2.6 *Enhancing Usability Heuristics for Android Applications on Mobile Devices - Thitichaimongkhol & Senivongse (2016)*

The mobile application heuristics from Thitichaimongkhol & Senivongse (2016) had many of the same issues as those from Gómez et al. (2014)—Nielsen's (1994) heuristics were not challenged with only two new heuristics added, both of which were no different to those from Gómez et al. (2014). Additionally, the heuristic set from the authors was tested by programmers and system analysts, not those experienced in conducting heuristic evaluations. Furthermore, similar to Gómez et al. (2014), context of use was not mentioned and a large set of 146 sub-heuristics was added. To that end, recommendations from evaluation theory were not considered, including Stufflebeam's (2000a) CDC, Stufflebeam's (2000b) CIPP model, and Scriven's (2005) comlist criteria.

2.2.7 *SMASH: A set of SMArtphone's uSability Heuristics - Inostroza et al. (2016)*

According to Quiñones & Rusu (2017, p. 91), "there is no formal process to formulate, specify, validate and refine usability heuristics". That was clear from the various approaches taken to define a set of usability heuristics for mobile applications based on the aforementioned examples. The approach taken by Inostroza et al. (2016)

differed in that the authors used a six-step methodology defined by Rusu et al. (2011). A formal process to define usability heuristics was a welcome addition to the literature. Besides having a process to follow, Rusu et al. (2011) specifically stated that a set of heuristics should be validated using heuristic evaluation. While the authors stated that usability tests could also be used, this was simply a secondary, complementary step.

Despite using this methodology, issues remained within the work from Inostroza et al. (2016). During a previous experiment, the primary author failed to find statistical significance between the number of usability issues found using his mobile application heuristics against Nielsen's (1994) heuristics, potentially as the experiment was under-powered with two evaluators assigned to each set of heuristics (Inostroza et al., 2012). For the 2016 study, Inostroza et al. increased the number of evaluators to twenty-seven, following which the authors found a statistically significant difference between the number of issues found using their mobile application heuristics against Nielsen's (1994) heuristics. There were three potential problems with this approach. Firstly, none of the evaluators had experience conducting usability evaluations, which was likely to impact the quantity and quality of issued found (Nielsen, 1992). Secondly, increasing the sample size, in this case the number of evaluators, to find statistical significance is not recommended (Field, 2013). Thirdly, this approach may have limited value as Nielsen & Landauer (1993) recommended about five experienced HCI evaluators before the law of diminishing returns makes adding more evaluators not worth the time and cost. As a counter-argument to the last point, Inostroza et al. (2016) might have argued that evaluating a software artefact and evaluating usability heuristics were dissimilar requiring different numbers of evaluators.

Nevertheless, the authors' mobile application usability heuristics changed little when compared to Nielsen's (1994) set. In some cases, definitions differed slightly from Nielsen's (1994) heuristics even though the heuristic titles remained largely the same. The few changes in the new set compared to Nielsen's (1994) heuristics were also somewhat duplicative. For instance, 'TMD1— Visibility of system status' stated 'The device should keep the user informed about all the processes and state changes through feedback and in a reasonable time', whereas 'TMD8—Efficiency of use and performance' stated 'The device should be able to load and display the required information in a reasonable time'. Furthermore, while Bertini et al. (2006) combined two of Nielsen's (1994) heuristics regarding errors in an effort to be more concise, Inostroza et al. (2016) once again split these into two error-related heuristics, namely 'TMD5—Error prevention' and 'TMD10—Help users recognize, diagnose, and recover from errors'. Another potential issue with Inostroza et al.'s (2016) heuristics was the continued use of the term 'Help and documentation' in heuristic TMD11. While help and support may be part of more complex mobile applications, it is certainly rare to see documentation being used for mobile applications as this was even problematic when used for desktop applications (Novick & Ward, 2006). In a similar way to Bertini et al. (2006), Inostroza et al. (2016) focused on ergonomics with the heuristic 'TMD12— Physical interaction'. That is not to say that this topic is unimportant. It does, however, demonstrate that researchers defining a set of usability heuristics for mobile applications may place their emphasis on issues found during similar research efforts and not on other critical aspects of mobile computing. Finally, the critical issue of context of use was not mentioned within this set of heuristics, nor was the use of increasingly sophisticated technology within today's smartphones that can take the burden off the user.

2.2.8 *Heuristics for Evaluating Multi-touch Gestures in Mobile Applications - Humayoun et al. (2017)*

In 2017, a set of heuristics was collated from previous heuristic sets by Humayoun et al. (2017) for multi-touch gestures in mobile applications. The set did not consider recommendations regarding Scriven's (2005) comlist criteria, in particular the qualities of overlap, conciseness and confirmability. As such, overlap occurred in a number of cases—'Heuristic 8—Help and documentation', whereby the authors stated that documentation needed to be provided to explain which gestures were available overlapped with 'Heuristic 10—Learnability', where the authors stated that gestures should be easy to learn, which inferred that documentation was not needed. This pointed to the tension between learnability and the need for support. It could be argued that systems should be easy to learn, but also that not everything can be available on the screen at once. This is especially true with the comparison between novice and advanced users, and between casual and regular users. 'Heuristic 4—Consistency and standards', whereby the authors stated that gestures should be consistent with the mobile operating system (OS) overlapped with 'Heuristic 14—Do not lie to the user' in which the authors stated that gestures should do what they are supposed to do under normal circumstances. These elements of overlap within a set of heuristics that need to be broad, yet concise, may have been a result of the authors not considering recommendations from Stufflebeam's (2000a) CDC.

One of the additional outcomes from not taking theory into account is that the authors created a set of fifteen heuristics, which did not consider the recommendation from Sharpe et al. (2007) in that heuristics should be kept to ten at most. Regarding confirmability, the authors compared their heuristics in an evaluation against a set of eleven mobile application heuristics, after which they simply offered a count of the number of issues found between sets. It is reasonable to think that fifteen heuristics will find more issues than eleven heuristics, which indeed occurs. Yet, the authors did not consider the severity of the issues found between either set of heuristics, leaving open the question of effectiveness. In addition, the heuristic evaluation used to validate the heuristic set was conducted by five computer science postgraduate students, not those experienced in conducting heuristic evaluations. Finally, recommendations from Stufflebeam's (2000b) CIPP model were not considered. As such, the authors did not mention the environment within which an evaluand is being assessed, nor did any of the heuristics mention the topic of context of use.

2.2.9 *Usability heuristics for quality assessment of mobile applications on smartphones - Costa et al. (2019)*

Da Costa et al. (2019) defined a set of thirteen heuristics based on a review of the literature. There was only one major change compared to the heuristics from Gómez et al. (2014), which was the addition of a new heuristic: 'UH8 - Efficiency of use and performance'. This heuristic overlapped with 'UH1 - Visibility with system status', whereby UH8 states 'The device must be able to load and display information in a reasonable amount of time' and UH1 states 'The application should…state changes within a reasonable period of time'. Context of use was mentioned as a critical topic by the authors at the start of the publication, however it was not mentioned within the set of heuristics. Furthermore, the heuristic set was not validated. To that end, recommendations from evaluation theory were not considered, including Stufflebeam's (2000a) CDC, Stufflebeam's (2000b) CIPP model, and Scriven's (2005) comlist criteria.

2.3 A Consideration of Context of Use

A consideration of context of use when referring to the evaluation of usability of mobile applications is critical as the conditions within which mobile applications are used can change rapidly (Reichmuth & Möller, 2014). So critical, that Savio & Braiterman (2007, p. 284) claimed "For mobile computing, context is everything!", which, if not adequately addressed, may result in a less than optimal user experience (Tsiaousis, 2015). Within this section, the point of reviewing relevant literature is not to chronologically list all instances of how context of use has been considered within usability studies since the concept was conceived; rather it is to better understand the concept of context of use, to reflect upon several alternative approaches, and to critically review existing work that has coupled context of use with heuristic evaluation.

Defining the concept of context of use will help to ensure that heuristic evaluation can be adapted or extended accordingly. Yet, the term context of use can mean different things to different people (Trivedi, 2012). Dourish (2004, p. 22), for instance, identified context of use as the "features of the environment within which the activity takes place". Dey (2001, p. 5) went further by arguing that context of use was anything "relevant to the interaction between a user and an application, including the user and applications themselves". Context of use is also part of ISO standards, such as ISO 9241-210 (2019), which notes that context of use is not simply environmental, context of use also includes the user, the mobile device, and the task that the user is attempting to complete within a specified environment.

Even with an understanding of context of use, usability evaluations of mobile applications, especially within fast-paced Agile environments, are often eliminated. This is due to a number of reasons. For instance, HCI practitioners rarely conduct longitudinal studies (Humayoun et al., 2014; Raison & Schmidt, 2013; Wale-kolade & Nielsen, 2014), primarily as they are under pressure to release mobile applications as quickly as possible (Bergvall-Kåreborn, 2011; Cottingham & Snyder, 2011). Additionally, a loss of control of variables when evaluating the usability of software artefacts in the field can be a concern (Tan et al., 2016). Consequently, the impact that changing contexts of use over time may have on the usability of mobile applications is often not considered (Eshet, 2016; Hagen et al., 2005; Rogers et al., 2007).

A consideration of context of use consists of two parts. First of all, a better understanding of the contexts of use within which targeted user types of a mobile application needs to be carried out when planning the development of a mobile application. Secondly, it is necessary to understand the impact of context of use on the usability of a mobile application. While temporal in nature in order to capture the results of changing contexts of use, conversely such methods need to provide results relatively quickly to be suitable for today's fast-paced mobile application software development environments. In terms of the first point, that is understanding the contexts of use within which a mobile application will be used. Such approaches already exist. For instance, Maguire (2001a) defined the usability context analysis method to better understand the contexts of use within which a software artefact will be used. The approach was later modified slightly for mobile contexts by Jumisko-Pyykkö & Vainio (2010). The second point, that is understanding the impact of context of use on the usability of mobile applications, has been less of a focus for researchers. While some researchers have argued that such approaches need be investigated (Zhang & Adipat, 2005), others have countered that evaluating usability in the field can be difficult and expensive (Høegh et al., 2008). This culminated in a debate about the value of conducting usability evaluations in the laboratory versus in the field (Duh et al., 2006; Kaikkonen et al., 2005; Kjeldskov et al., 2004). This debate ended with the realisation that evaluating usability in the field uncovered issues that were not, and could not, have been found in the laboratory (Kjeldskov & Skov, 2014).

Thus, attention focused on experimenting with various field-based methods. A promising approach is to utilize the sensors in smartphones to better understand context, and to use that data to understand where the user is, what they are trying to achieve, and to better understand if the user has pain points (Abowd et al., 1999; Zhang et al., 2009). Nevertheless, this field of research is quite complex and while much has been learned, many obstacles remain unresolved leading to many mobile application developers under-utilizing sensors (Bobek et al., 2019; Nalepa et al., 2019; Yu et al., 2013). For example, only inferences can be made through context-aware data, whereby two situations can be very different and still be considered the same via a context-aware application (Gulliksson, 2012; McMillan et al., 2015). Thus, when Maguire (2001a, p. 17) defines context of use as a consideration of “whom the product was designed for, what it will be used for, and where it will be used”, context-aware applications can misinterpret a specific context at best or not consider contextual elements at worst. To that end, while the field of context-awareness holds much promise, more work is needed. Examples include the potential difficulty for context-aware applications to understand:

- Social situations
- Inattention or distractions
- Goals the user wishes to achieve
- How a mobile device is being held
- Encumbrance whereby a baby or groceries etc. are being held in one hand and a mobile device is being held in another hand
- Users’ feelings and emotions

- Potential user disabilities, such as motor disability or colour blindness
- Users' activities, such as standing on a street to cross the road versus standing on a street waiting for a bus
- Users' age and experience level and so forth.

Other researchers have been investigating alternative approaches. McGregor et al. (2014) recorded the mobile screens of fifteen mobile phone users over time. While this approach helped to understand general day-to-day usage of mobile devices, the authors pointed out that they had gathered a vast amount of video data, which was prohibitively slow and difficult to analyse. In some cases, the authors concluded, even with the use of GPS and video recording, there was still a level of uncertainty about the contexts of use. An analysis of insights that slows down a mobile application release might increase costs, thus decreasing any return on investment for software developers (Lane et al., 2015). Consequently, as more and more companies use fast-paced Agile software development methods, being relatively quick to discover insights can be deemed a requirement of any method that is used to evaluate the usability of mobile applications. On that note, several methods may work well within fast-paced Agile environments. These include field-based rapid ethnographic studies (Millen, 2000) and conducting field-based interviews (Ginsburg, 2010). However, these methods do not consider the temporal nature of changing contexts of use, which are prevalent when using mobile devices. One could argue that reviewing analytics data of users of mobile devices will give the insights needed to know how a mobile device is used with various contexts of use. Yet, Beasley (2013) believed this approach to be problematic in that tools, such as Google Analytics, were primarily designed for marketing teams, not usability experts, potentially leading to missed usability issues. Another approach that considered the temporal aspects of mobile computing were diary studies. However, Brandt & Weiss (2007) found that study participants were often unable to offer comprehensive diary entries when they were mobile themselves.

A number of researchers have considered how context of use might have adequate representation in regard to heuristic evaluation. This has been done in two different ways:

- Adding heuristics that consider context of use
- Conducting a heuristic evaluation using contextual scenarios in the laboratory or in the field

2.3.1 *Adding Heuristics that Consider Context of use*

The critical review of sets of mobile application heuristics in the previous section has demonstrated that context of use has not been given adequate representation within sets of heuristics defined within the past two decades. Two notable cases are the exception, as detailed below.

In an update to their 2006 article, Bertini et al. (2009) focused on context of use and revised their original heuristics. However, rather than actually considering the topic of context of use, only minor grammatical edits were made to the original heuristics, following which the authors stated that they have considered context of use. To that end, the critical review earlier in this chapter continued to be accurate as the authors simply highlight heuristics that focused on ergonomics of the mobile device, ensuring that the mobile device was robust to damage if accidentally dropped, and whereby a mobile device could be found if lost.

A set of fourteen mobile application heuristics was published from Bashir & Farooq (2019), which also used the methodology from Rusu et al. (2011). While the set was unremarkable overall as it was also derived from the literature, one heuristic stood out in that it was dedicated to context of use. That heuristic was referred to as

‘P8. Handling varied context of use in mobile environments’. While this was a welcome addition to a set of heuristics, the authors did not state how they expected evaluators to consider varied contexts of use, more so as usability evaluations are generally conducted within laboratories (Eshet, 2016). Furthermore, while the complexity of context of use was touched upon earlier in this section, the authors only referred to noise, bad light, motion, and social interaction, which may simplify the complexity of context of use too much for an adequate usability evaluation. Lastly, an evaluation was carried out by comparing the new set of heuristics with other sets using twenty-four evaluators with twelve assigned to two sets of heuristics each. Although the authors referred to the twenty-four participants as usability experts, in reality the majority of participants were software engineers, 'experts with no experience', and others that the authors simply assumed had knowledge about heuristic evaluation. The success of the heuristic dedicated to context of use was difficult to ascertain from the publication as the authors did not refer to specific issues discovered using this heuristic.

2.3.2 *Conducting a Heuristic Evaluation using Contextual Scenarios in the Laboratory or in the Field*

While heuristic evaluations are most often conducted in the lab, a study conducted by Po et al. (2004) considered context of use using three conditions: heuristic evaluation within a laboratory for two conditions, one of which simulated context of use, as well as a heuristic evaluation in the field. There were several threats to the validity of data from this study. For instance, during the heuristic evaluation, the authors used Nielsen's (1994) traditional heuristics, and did not attempt to define nor use a pre-defined set of usability heuristics for mobile applications. Furthermore, the internal validity of the study was threatened by selection bias, whereby the authors did not randomize participants across conditions. All conditions also had fewer than the recommended five evaluators (Nielsen, 1992). It seemed that none of the evaluators had practical HCI experience as several evaluators only had one semester in HCI and had little familiarity with heuristic evaluation. This could have been a threat to external validity, specifically population validity, in that the evaluators within this study cannot be generalized with HCI researchers and practitioners that have practical experience in commonly conducting heuristic evaluations.

While it can be difficult to keep all conditions the same in an experimental or in this case a quasi-experimental study, it should still be expected that all conditions would be as equal as possible. Yet, in this study the authors assigned user interface-related tasks to participants in just two of the three conditions. The authors stated that a by-product of this omission of task assignment in the first scenario was that participants focused on the product, more so than the operation of the product. As high construct validity is achieved only when intended constructs are measured accurately, it might be argued that the authors were measuring two operationalized variables defined from the construct 'usability issues'. Additionally, within the two conditions where evaluators completed six tasks, there was no indication that the tasks were counterbalanced. This could indicate an order bias.

Furthermore, a benefit of an in-situ evaluation is to increase ecological validity. Yet, participants in this study had to think aloud while they were being recorded. Consequently, the in-situ condition was not realistic. Participants were also aware which condition they were in within the study. This transparency could result in an internal threat to validity known as demand characteristics. This threat can occur when participants know they are in the experimental group, and they change their behaviour due to their expectations of the study. The authors of the paper might have minimized this threat by contriving a cover story, thus only offering the real reason for the study when participants were debriefed.

Reliability was also compromised as it would be impossible to replicate the study. The authors of the work were not fully transparent about the environments and activities of the participants for any condition. For instance, within the in-situ condition, the authors simply stated locations, such as cafeteria or bar, and minor details, such as varying light conditions. To replicate the study, other researchers would need to know about all conditions, including environmental, as well as participants' activities, such as walking and so forth. Any of these environmental and social factors may impact the findings yet were not published by the authors.

Within the in-situ condition, fewer evaluators uncovered substantially more issues than the lab-based heuristic evaluation. While the second condition, the laboratory-based condition with scenarios of use, uncovered more issues, this condition had a great number of evaluators, yet uncovered only 0.4% higher usability issues on average. Additionally, the in-situ study uncovered a higher number of critical usability issues, as well as issues where context of use had an impact. These issues were not found within the laboratory-based studies. Consequently, based on the threats to reliability and validity, it could be argued that the conclusion from Po et al. (2004), namely that there are no benefits to conducting heuristic evaluation within a specified context of use could be interpreted differently.

Several years after the work by Po et al. (2004), Varsaluoma (2009) focused on conducting a heuristic evaluation of mobile devices by adding contextual scenarios. This work had similar problems that may impact the validity of the findings and the reliability of the research effort. For instance, in order to validate the approach taken, the author compared findings from usability testing to heuristic evaluation. As previously stated, both approaches are different and will find different types of problems, even though some overlap may exist. Further, even though Bertini et al.'s (2006) mobile usability heuristics had already been defined at this time, Varsaluoma (2009) used Nielsen's (1994) set of traditional heuristics instead, which are not as effective when used to evaluate the usability of mobile applications.

Another area of concern was that the evaluators conducting the heuristic evaluations were requested to think of issues that might be found within real-world contexts of use without being able to conduct the heuristic evaluation within such conditions. Following this, the author stated that thirty-five problems uncovered by usability testing were not predicted during contextually imagined heuristic evaluations with only two of the issues being related to context of use.

A question remained—one that was not addressed in the work by Po et al. (2004) nor by Varsaluoma (2009), yet it was, if inconclusively, by Bashir & Farooq (2019)—it might be possible to achieve a better result by using a set of mobile application heuristics and adding a heuristic specific to context of use. This approach was evaluated when the first draft of mobile application heuristics as part of this programme of research was defined.

2.4 SUMMARY

Having critically reviewed the literature, the next step was to outline a methodology that would address the research questions, which would include:

- Increasing the reliability of proposed evaluation methods by following theoretical and practical models, including Stufflebeam's (2000a) CDC, Stufflebeam's (2000b) CIPP, Scriven's (2005) comlist criteria, and Rusu et al.'s (2011) six-step methodology to define a set of heuristics;
- Increasing the reliability of proposed evaluation methods by seeking feedback from experienced HCI researchers and/or practitioners;

- Increasing the validity of heuristic evaluations by ensuring that experienced HCI researchers and/or practitioners conduct studies;
- Validating a set of mobile application usability heuristics through a heuristic evaluation of a mobile application using the newly defined set of heuristics and comparing results to those gathered using traditional heuristics (at a minimum);
- Using the recommended number of evaluators (Nielsen, 1992) to ensure that heuristic evaluations do not simply seek to gain statistical significance;
- Considering the addition of a heuristic specific to context of use to the newly defined mobile application heuristics to understand if this approach achieved a better result than previous studies;
- Testing the point of view from Sharpe et al. (2007) that the number of usability heuristics should equal no more than ten.

Chapter 3. METHODOLOGY AND METHODS

Having positioned my work within the literature in the previous chapter, it was important to state my philosophical perspective and assumptions prior to addressing the research questions. By carefully considering your own worldview, Hesse-Biber (2010) argued that a researcher is less likely to follow the most frequently used paradigm within their respective field. Garner et al. (2009) and Collins (2010) concurred, stating that a researcher's philosophical perspective would inform which research strategy and methods would be used, as well as guiding practical considerations of the research approach, such as attention to ethics.

3.1 Philosophical Perspective

A philosophical perspective is made up of three primary areas for reflection, namely ontology, epistemology, and axiology. The primary ontologies are realism, critical realism, and constructivism, which focus on a researcher's viewpoint on the nature of social reality—in other words, their belief on what exists. Epistemologies are related to ontologies in that they focus on how one would know what their beliefs are, in other words, how might one gather the knowledge to know what exists. The primary epistemologies are positivism, post-positivism, and interpretivism, which are related to realism, critical realism, and constructivism respectively. Axiology, the final part of philosophical spectrum, focuses on the quality of research. A researcher's worldview can come from experience, be influenced by other researchers that they work closely with, or it might be defined based on the worldview prevalent in the literature for a specific field. Possessing a specific philosophical perspective may be more difficult within HCI as there are no clear paradigms nor theories within the field given its relative recency (Williamson & Brewster, 2012). This was certainly evident during the literature review.

Despite the lack of theoretical foundations within HCI, I am aware that my personal perspective does not relate to the viewpoint of realism/positivism, which deems the social world to be objective, one that can be measured scientifically. Rather, I lean toward a constructivist/interpretivist viewpoint, whereby I believe that social reality is subjective. Equally, my philosophical perspective is influenced by critical realism, not least because context of use is a central theme within that worldview. From an HCI perspective, Frauenberger (2016, p. 349) contended that "A critical realist perspective draws attention to the mechanisms through which a designed artefact interacts with other things in the real world". The influence of this viewpoint was evident as the final two research questions were addressed, specifically in regard to the concept of identifying generative mechanisms, which was in turn influenced by the work of Blom & Morén (2011).

Apart from the influence of constructivist and critical realist perspectives, this programme of research was not approached solely from either worldview. Rather, my epistemological approach centred on pragmatism. Pragmatists "set out to use whichever techniques will answer or address the research question" asserted Frost (2011, p. 5). While the following quote was not explicitly referring to pragmatism, I felt that it truly summed up my philosophical approach:

> *"Science is not seen as an activity of following methodological recipes that yield acceptable results. Science becomes the creative search to understand better, and it uses whatever approaches are responsive to the particular questions and subject matters addressed. Those methods are acceptable which produce results that convince the community that the new*

understanding is deeper, fuller, and more useful than the previous understanding" (Polkinghorne, 1983, p. 3).

This quote refers to a common practice within pragmatism, namely to use whatever research strategies and methods are required to address research questions, regardless of if those research strategies and methods are more commonly used within specific worldviews. As Creswell & Creswell (2017, p. 6) stated, this is more of a "problem-solving" and "real-world practice-oriented approach". Not all researchers are fully convinced that this is a viable approach, even those that have combined research strategies and methods. McEvoy & Richards (2006, p. 66), for instance, believed that researchers should exercise caution due to the "complex ontological and epistemological issues that are involved". Creswell & Plano Clark (2017) dismissed this by explaining that pragmatism could be used as an overall approach yet draw upon various objective and subjective ontological and epistemological underpinnings and assumptions, including constructivist and critical realist perspectives, as appropriate within a single research study. This point of view was particularly applicable vis-à-vis a programme of research that was sufficiently broad and complex enough to warrant the use of multiple methods.

3.2 RESEARCH STRATEGY

Pernecky & Jamal (2010) claimed that it is easier to define the research strategy once a researcher is clear in his or her worldview. The literature may strongly influence this choice, as it is quite common to see realists gravitating toward quantitative research strategies, such as experiments and surveys; constructivists, on the other hand, lean toward qualitative research strategies, such as case studies, ethnography, and phenomenology; critical realists, who look at social reality from a realist ontology and an interpretive epistemology, as well as pragmatists who integrate various methods to address each research question, tend to contemplate a mixed methods approach.

Thus, having identified as a pragmatist, the most appropriate research strategy would be a mixed methods approach. Yet, other researchers are more cautious when choosing a research strategy. Morgan (2014, p. 1045) maintained that "There may be an affinity between paradigms and methods, but there is no deterministic link that forces the use of a particular paradigm with a particular set of methods". Denscombe (2010) agreed, pointing out that a research approach needs to be feasible and should produce appropriate kinds of data required to address each research question. To that end, it was important to consider the main types of research strategies before deciding which approach was most appropriate for this programme of research.

3.2.1 *Primary Research Strategies*

The primary research strategies available are experiments, surveys, case studies, action research, grounded theory, archival research, mixed methods, ethnography, and phenomenology (Collins, 2010). While experiments and surveys tend to focus on numbers and statistical significance, other strategies, including grounded theory, ethnography, and phenomenology tend to be qualitative in nature. Qualitative research considers words more so than numbers, as this type of research is concerned with deeply understanding experiences (Jackson et al., 2007). Consequently, qualitative research tends to be inductive across disciplines (Schutt, 2014). As a deep understanding of experiences has been the focus, less so on generalizability of findings, qualitative research has been criticized for being unscientific in the past (Mays & Pope, 1995), however this perspective has changed in recent years (McAleese & Kilty, 2019). To better understand peoples' experiences, qualitative researchers seek patterns in order to define concepts, theories, explanations, and understanding. To uncover these patterns, analysis methods, such as grounded theory, discourse analysis, and narrative analysis, are often used. Research strategies, and their applicability to this programme of research, are summarised below.

3.2.1.1 Experiments

Experiments tend to be more aligned with a realist worldview as they empirically measure variables under controlled conditions (Denscombe, 2010). Realists with a positivist/post-positivist perspective conducting a similar programme of research might have used experiments with regression analysis to better understand the relation of contextual variables and their outcome on the impact on usability of mobile applications. However, this approach was not suitable for this programme of research as the focus was on understanding people's attitudes toward mobile application usability within real-world conditions, not to understand variables within a controlled laboratory environment.

3.2.1.2 Surveys

Surveys consist of a list of questions and can be used to collect quantitative and qualitative data on a small or large scale (Groves et al., 2011). They are useful in gathering wider points of view and attitudes than methods, such as case studies or interviews. These could be online, mailed, face-to-face, and phone surveys. Online surveys in the form of questionnaires were used twice during this programme of research.

3.2.1.3 Case Studies

Case studies focus on a very small number of phenomena (Yin, 2018). The data gathered, generally through observation over time, can offer a complex array of information around a particular account, be it a single event, situation, or an individual over a period of time. I decided against this approach as I wished to learn from a wider group of participants.

3.2.1.4 Action Research

Action research allows researchers to take subsequent action(s) to resolve organizational and/or academic problems following a research effort (Stringer, 2013). This was a viable option for this programme of research, however I wished to hear from participants from outside my organization and in some cases across the field of HCI, therefore this approach was discounted.

3.2.1.5 Grounded Theory

Grounded theory was defined by Strauss & Corbin (1997) to describe theories derived from empirical data using a systematic approach. This tends to be a time-consuming approach as defined theories are constantly checked against new data. Researchers following this approach start with an open mind, therefore they are not supposed to be influenced by other theories, although this tends to be difficult in practice. The purpose of this programme of research was not to define a theory. Therefore, the grounded theory approach was discounted.

3.2.1.6 Archival Research

The primary sources of data within an archival research strategy are archives, including photos, documents, manuscripts, videos, and audio files. These are generally from repositories, such as museums, libraries, and historical societies (Ventresca & Mohr, 2017). While this approach was discounted from an overall strategic perspective, the associated research method helped to address some of the research questions within this programme of research.

3.2.1.7 Mixed Methods

Mixed methods allow for the combination of qualitative and quantitative methods within one research project. The focus is less on which methods fit into a specific paradigm, shifting the focus instead to which methods are best applied to answering research questions (Creswell & Plano Clark, 2017). This approach was more aligned with this programme of research than other strategic approaches.

3.2.1.8 Ethnography

Ethnographers spend a lot of time with a particular group of people, using methods such as observation to learn about their everyday lives from an insider's perspective (Fetterman, 2019). While this approach may have helped with one research question within this programme of research, namely learning more about contexts of use, the approach would not have been beneficial from an overall research strategy as it would have been difficult to address all other research questions.

3.2.1.9 Phenomenology

Phenomenology focuses on the everyday lived experiences of individuals (Cerbone, 2014). While there may be overlap with ethnography and case studies, phenomenology differs in that the focus is usually the perspective of a single experience from multiple participants. As such, this strategy was initially considered, especially for the last two research questions within this programme of research, which focused on understanding the impact of context of use on the usability of mobile applications. However, the aim of this programme of research did not align fully with the purpose of phenomenology. Nonetheless, as phenomenology "emphasises the attempt to get to the truth of matters, to describe phenomena, in the broadest sense as whatever appears in the manner in which it appears, that is as it manifests itself to consciousness, to the experiencer" (Moran, 2000, p. 4), this programme of research will be of interest to those with a phenomenological worldview. As there are a variety of phenomenological interpretations (Giorgi et al., 2017), this work will be of particular interest to researchers that hold a hermeneutic phenomenological worldview, that is the interpretation of lived experiences (Hiller, 2016). Hermeneutic phenomenology as defined by Heidegger (1971) and Gadamer (1976), is a variant of the transcendental phenomenology, originally defined by Husserl (1931). As Morgan (2011, p. 17) pointed out, the primary thinkers within this philosophical field consider a person to be "the expert on his or her experience".

3.2.2 *A Focus on Mixed Methods*

A mixed methods approach was selected because it can use a variety of quantitative and qualitative methods as deemed appropriate to address each research question. Within mixed methods research strategies, a number of designs can be used, namely convergent design, explanatory design, exploratory design, embedded design, transformative design, and multiphase design (Creswell & Plano Clark, 2017). The multiphase design was best suited to this programme of research as the effort was conducted over several years, combining concurrent and sequential quantitative and qualitative data sets that addressed a set of research questions, which helped to define and evaluate a usability evaluation method. The primary types of methods commonly associated within mixed methods research were considered below:

3.2.2.1 Statistical Analysis

Statistical analysis is used within an experiments research strategy to make sense of a large amount of data. The method can also be used to make inferences from a sample to a population and to compare results across different groups (Williams, 2007). As the aims of this programme of research were being met, statistical analyses were used occasionally from a comparison perspective.

3.2.2.2 Observation

Observation is often used with strategies, such as ethnography. One of the advantages is that it allows a researcher to watch what participants do and not rely on what they say they do (Spradley, 2016). However, due to generally limited observation schedules, contextual information can be missed.

3.2.2.3 Focus Group

Focus groups allow a researcher to better understand the attitudes of several people at once toward a concept, product, or idea (Stewart et al., 2009). This method was an option for this programme of research, especially in terms of gathering attitudes toward outputs, such as a set of mobile application usability heuristics. Focus groups were not utilised due to the potential for group conformity (Rosenbaum et al., 2002). However, a style of focus group in the form of a workshop was used. In this instance, the potential risk for group conformity was reduced by splitting participants into three separate groups.

3.2.2.4 Questionnaire

Questionnaires are used within a survey research strategy and are generally useful in gathering attitudes from a large number of participants, potentially from a wide geographical area (Rahi et al., 2019). As such, questionnaires were utilised within this programme of research when gathering the attitudes of those knowledgeable in HCI.

3.2.2.5 Interviews

Interviews, which might be structured, semi-structured or unstructured, are used in a number of research strategies, such as phenomenology. This method allows a researcher to speak one-to-one with participants to gather their thoughts, perspectives, opinions, and ideas (King et al., 2018). Due to the ability to probe further on interesting responses from participants, semi-structured interviews were employed on a number of occasions within this programme of research.

3.2.2.6 Diary Study

Diary studies can be used to record events and decision-making processes from participants when the researcher is not present (Gunthert & Wenze, 2012). This approach was an option for this programme of research when learning about the contexts of use that participants found themselves in while using mobile applications. However, without access to dedicated diary study software, a large burden is placed on a researcher. Conversely, dedicated diary study software places the burden on the system, whereby a researcher can proactively keep participants engaged to reduce mortality rates.

3.2.2.7 Document Analysis

Document analysis is used within an archival research strategy. The primary goal is to analyse documentary evidence when addressing research questions (Bowen, 2009). While an archival research strategy was not used as a strategy for this programme of research, the document analysis method was utilised when analysing aspects of the literature as several research questions were addressed.

3.2.3 *Methods used within this Programme of Research*

The methods chosen to address research questions within this programme of research were questionnaires, semi-structured interviews, and document analysis. Statistical analyses were also used occasionally. The justification for each method was explained within each of the chapters that focus on each specific research question.

3.3 Research Ethics

Research ethics is a critical part of being a researcher (Mack, 2005). Resnik (2011, p. 1) described ethics as "norms for conduct that distinguish between acceptable and unacceptable behaviour". Fouka & Mantzorou (2011, p. 4) went further by stating ethical "code focuses on voluntary informed consent, liberty of withdrawal from research, protection from physical and mental harm, or suffering and death. It also emphasises the risk-benefit balance".

During this programme of research, ethical guidelines were always understood intimately as I attended ethics training conducted by the University of Hertfordshire. For each study that involved human subjects, ethics approval was sought and granted prior to the start of each study. All ethics approvals are available in Appendix B. Furthermore, informed consent forms were either read and signed by all participants in the case of interviews or acknowledged in the case of questionnaires. These forms allowed each participant to understand the research effort, what was expected of them, how their privacy was respected, while informing each participant that they were able to withdraw from a research study at any time without giving a reason.

3.4 Research Quality

The quality of any research effort, including research design, data collection and findings, is critical. Traditionally, research quality has been established using the concepts of reliability, validity and objectivity. Reliability focuses on consistency of procedures and instruments, including how capable both are of producing the same results in different contexts or with the same participants at different times (Mohajan, 2017). Validity is split into two concepts, internal and external. Internal validity focuses on the trustworthiness of instruments, data, and findings, whereas external validity considers the applicability of research findings to other contexts (Christensen et al., 2011). Objectivity is defined as the extent to which research findings truly represent participants' views and that the research is not contaminated by bias in any way (Letherby et al., 2012).

While some researchers, such as Franklin & Ballan (2001) continued to use the concepts of reliability, validity and objectivity within the context of qualitative research, others have argued that such concepts are only applicable to quantitative research (Anney, 2014; Cypress, 2017; Golafshani, 2003). For instance, Lacey & Luff (2001) maintained that the precise replication expected within quantitative research may not be appropriate within qualitative research. This argument was echoed by Leung (2015, p. 325), who suggested that "a margin of variability for results is tolerated" as qualitative research can often be context or participant dependent. Noble & Smith (2015, p. 1) agreed with these sentiments by stating "if qualitative methods are inherently different from quantitative methods in terms of philosophical positions and purpose, then alterative frameworks for establishing rigour are appropriate".

Within this mixed methods programme of research, quantitative methods, including questionnaires and statistical analyses were used occasionally, while qualitative research methods, such as interviews, were more prevalent. To that end, the viewpoints of two alternative qualitative research quality frameworks were contemplated. Rather than choose one approach over another, it seemed opportune to apply each at different levels, one at a dissertation level and one at an individual study level.

3.4.1 *Considering the Quality of Qualitative Research at a Programme of Research Level*

The quality of this programme of research met each of the eight criteria for excellent qualitative research set out by Tracy (2010) in the following ways:

- *Worthy topic:* The primary topics within this programme of research, namely defining a set of heuristics for the usability evaluation of mobile applications and creating a protocol to consider the potential impact of context of use on the usability of mobile applications, were relevant, timely, significant, and important.
- *Rich rigor:* The programme of research used well-established theoretical constructs based on the literature, quite some time had been spent in the field with a representative sample, namely HCI researchers and practitioners. Further, several data collection and analysis processes were considered and only the most appropriate were selected.
- *Sincerity:* Each study within the programme of research was transparent about the methods selected, whereby the decision process was fully apparent.
- *Credibility:* The programme of research considered the reflections of several hundred people knowledgeable in HCI, which increased the credibility of the work.
- *Resonance:* The programme of research influenced, affected, or moved readers through evocative representation and transferable findings to HCI researchers and practitioners and the usability evaluation of a variety of mobile applications.
- *Significant contribution:* The programme of research provided a significant contribution, both conceptually and theoretically, in that the contributions filled a gap in knowledge within the literature.
- *Ethical:* The programme of research was ethical in that a strong understanding of ethics was achieved through training prior to studies

being carried out. Further, procedural ethics was fully followed, and ethics approval was obtained prior to every study that involved humans.

- *Meaningful coherence:* The programme of research obtained meaningful coherence in that it achieved what it purported to be about, used methods and procedures that fit stated goals, meaningfully interconnected literature, research questions, findings, and interpretations with each other.

3.4.2 *Considering the Quality of Qualitative Research at a Study Level*

Lincoln & Guba (1985) viewed the trustworthiness of any research effort, be it quantitative or qualitative, under four constructs, namely truth value, applicability, consistency, and neutrality. Truth value refers to establishing confidence in the findings; applicability considers if the findings are relevant to any degree to other contexts; consistency denotes the degree to which findings are consistently repeated should the research be conducted within the same or similar contexts; and neutrality refers to the determination that findings are true representations of participants' perspectives. The approach to establishing trustworthiness for each construct is considered differently between quantitative or qualitative research (Table 3.1):

Table 3.1. Research trustworthiness

Construct	Quantitative	Qualitative
Truth value	Internal validity	Credibility
Applicability	External validity	Transferability
Consistency	Reliability	Dependability
Neutrality	Objectivity	Confirmability

Note. Adapted from Guba, E. G. (1981). Copyright 1981 by Guba, E. G

The qualitative terms in the table adapted from (Guba, 1981) refer to the following:

- *Credibility:* Findings are true to participants' perspectives;
- *Transferability:* Ability of findings to apply to other contexts;
- *Dependability:* Ensuring that other researchers can repeat the research;
- *Confirmability:* Confidence that bias was mitigated as much as possible.

The quality of studies within this programme of research were established using the appropriate qualitative or quantitative criteria.

3.5 SUMMARY

To summarise, philosophical perspectives were considered prior to the selection of pragmatism as the worldview through which this programme of research was written. A number of research strategies and methods were contemplated before a mixed methods approach was chosen. The next chapter focuses on addressing the first research question, namely how a set of mobile application heuristics might be developed.

Chapter 4. HOW MIGHT A SET OF MOBILE APPLICATION USABILITY HEURISTICS BE DEVELOPED?

The first research question focused on developing a set of mobile application heuristics. This research question was split into two sub-research questions that considered how mobile application usability heuristics were being used, and which factors needed to be included in a set of mobile application usability heuristics.

4.1 HOW ARE MOBILE APPLICATION USABILITY HEURISTICS CURRENTLY BEING USED?

The literature review demonstrated that the Nielsen's (1994) heuristics were not suitable for the usability evaluation of mobile applications (Baharuddin et al., 2013). To address this, part of this programme of research was to define a set of mobile application usability heuristics. While other sets of mobile applications usability heuristics were already proposed by other HCI researchers, the literature review also demonstrated that gaps in knowledge needed to be filled. As several sets of mobile application heuristics were already defined, the objective of asking this sub-research question was to understand if other sets of mobile application usability heuristics were used by HCI practitioners. Should issues have existed with the usage of existing mobile application heuristics, these could be addressed along with gaps in the literature when an improved set of mobile application usability heuristics was created.

4.1.1 *Method*

A common theme within the literature review was that researchers often consulted only the literature to understand which heuristics were the most commonly used for mobile applications. While it is important to review the literature to better understand the use or non-use of mobile application usability heuristics, it is also important to gather information from those knowledgeable in HCI. A questionnaire would have allowed for a wider array of answers over a wider geographical area, yet the ability to probe on interesting answers would have been difficult or impossible (Krosnick, 2018). To that end, due to the ability to probe, semi-structured interviews were used. This approach is also one of the most common ways of gathering qualitative data and has been for some time (Bernard, 2000). Indeed, Bingham & Moore (1959) referred to interviews as conversations with an end goal, whereby the researcher converses with participants and ends up with new information that they did not have prior to the conversation.

The next consideration centred around the number of interviews that needed to be conducted. The literature is inconclusive in this regard. Researchers have stated that the number of interviews required to understand a phenomenon is linked to sample heterogeneity and research objectives (Guest, 2006). Should the sample be homogenous, Kuzel (1992) recommended six to eight interviews. This is supported by Romney et al. (1986), who reported that samples as small as four individuals offered detailed information around a phenomenon once the individuals held enough knowledge and experience in the field of inquiry. As all participants within this study were expected to be knowledgeable in HCI, it was assumed that there would be a substantial amount of homogeneity.

The first round of interviews were conducted with attendees to a User Experience Professionals Association (UXPA) (*User Experience Professionals Association*, n.d.) conference in Boston on May 15th, 2015 (ethics protocol number: COM/PG/UH/00089). Purposive sampling (Taherdoost, 2016) was used, specifically stratified and systematic random sampling (Nassiuma, 2001) in that every tenth male and every tenth female was selected for an interview. In total, ten interviews were conducted at the conference. Between September 1st and 2nd, 2015, six more interviews were conducted (ethics protocol number: aCOM/PG/UH/00107). The latter set of interviews was conducted to gather further thoughts from participants.

4.1.1.1 Materials and Procedure

Participants were informed that the aim of the study was to interview those knowledgeable in HCI regarding their mobile application usability evaluation practices, as well as how long the interview would take. Participants were also informed that they could withdraw at any stage without giving a reason. Participants read both the informed consent form and a participant information sheet that gave further details about the study, including that any data gathered would be used in an anonymized form, that it could not be traced back to the participant, and that it would be kept secure. Once a participant signed the informed consent, they were asked if the interview could be recorded. All interviews were recorded with participants' permission using a high-quality audio recorder. No problems were noted with audio quality. The interviews were transcribed in Nvivo 10 (*Nvivo*, n.d.).

The interview guide for the first set of interviews was used to directly investigate usage of heuristic evaluation for mobile applications:

1) Do you use the heuristic evaluation method to evaluate the usability of mobile applications? Why/Why not?
2) Only if yes to Q1: Which set of heuristics do you use?
3) Only if yes to Q1: Is the heuristic evaluation done in-house or outside?
4) Only if yes to Q1: Are you able to or not able to find three or more evaluators for a heuristic evaluation?
5) Would a standard set of usability heuristics applicable to mobile applications be useful or not useful for you? Why/Why not?
6) What is your job role?
7) How many years' experience do you have in HCI/UX?
8) How many years' experience do you have in mobile HCI/UX?

The term UX was used in the interview guide; this refers to user experience. This term was used in conjunction with HCI as UX is more common outside of the academic arena. Following this set of interviews, it was decided to conduct a further set of interviews with those knowledgeable in HCI to gather more information. The following interview guide was used, which evolved based on the answers to the first round of interviews:

1) Which software development process do you use?
2) Can you tell me about your experience in evaluating the usability of applications, in particular mobile applications?
    a. Do you use the heuristic evaluation method to evaluate the usability of applications, in particular mobile applications? If so, which set of heuristics do you use?
    b. How long does it take to do the heuristic evaluation, including the full evaluation and analysis?
    c. Did you ever need to create new heuristics, and if so, how long did that take?
    d. How many people usually take part in the heuristic evaluation of an application?
3) How do you communicate usability issues with other stakeholders, such as product managers and developers?
4) How do you consider context of use when evaluating the usability of applications, in particular mobile applications?
5) What is your job role?
6) How many years' experience do you have in HCI/UX?
7) How many years' experience do you have in mobile HCI/UX?

4.1.1.2 Participants

Three of the original participants were interviewed once more. Also, as one participant in the original set of participants was a software engineer with no experience in HCI, that participant's answers were removed from the analysis. Thus, a total of thirteen participants took part in sixteen interviews (5 male, 8 female; Years' experience in HCI: Range=1-20; Mean= 8.66; SD=5.97; Years' experience in mobile HCI: Range=1-9; Mean=3.33; SD=2.42) (Table 4.1 / Table 4.2):

Table 4.1. Round one participants (May 15, 2015)

Participant	**Gender**	**Job role**	**Years' experience (HCI/Mobile HCI)**
P1	Female	UX Designer	15/5
P2	Male	Senior Designer	6/1
P3	Female	UX Designer	1/1
P4	Male	UX Strategist	2/1
P5	Female	UX Strategist	5/5
~~P6~~	~~Male~~	~~Software Engineer~~	~~0/0~~
P7	Female	UX Designer	7/1
P8	Male	UX Lead	20/9
P9	Female	Senior UX Researcher	15/5
P10	Male	Senior UX Designer	12/4

Table 4.2. Round two participants (September 1-2, 2015)

Participant	Gender	Job role	Years' experience (HCI/Mobile HCI)
P3 (Interviewed once more)	Female	UX Designer	1/1
P8 (Interviewed once more)	Male	UX Lead	20/9
P10 (Interviewed once more)	Male	Senior UX Designer	12/4
P11	Female	UX Designer	10/3
P12	Female	UX Designer	9/3
P13	Female	UX Designer	2/2

4.1.1.3 Analysis

The process that qualitative analysis methods follow tend to be similar, whereby all generate/re-generate codes, then a search for patterns is conducted (Auerbach & Silverstein, 2003). Saldaña (2015, p. 3) described codes as "a word or phrase symbolically assigns a summative, salient, essence-capturing and/or evocative attribute for a portion or language-based or visual data". While codes are useful in making sense of qualitative data, finding patterns between multiple codes across many pages during qualitative inquiry can be time-consuming and potentially erroneous (Male, 2016).

While helpful, even computer assisted qualitative data analysis software (CAQDAS) has limitations with this process (García-Horta & Guerra-Ramos, 2009). To assist further in the search for patterns, the analysis method chosen was framework analysis. While framework analysis is frequently used by qualitative researchers, particularly within healthcare (Georgsson & Staggers, 2016; Kwasnicka et al., 2014; Yon et al., 2015), it has not yet seen widespread use within mixed methods research

(Mason et al., 2018). Despite this, the case and thematic matrix approach used within framework analysis was an interesting way to understand insights as they emerged from the data. To give more context on the case and thematic matrix approach, the framework analysis method followed a standard thematic analysis (Blaxter, 2010), which consisted of familiarization with the data, data being transcribed, initial codes being generated, and themes being noted, reviewed and named. The main difference between framework analysis and other qualitative analysis methods is that the qualitative data is placed in a matrix, somewhat similar to a spreadsheet with cells of summarized data, which assists with the recognition of patterns and contradictory information (Gale et al., 2013).

During the implementation of the framework analysis approach, thematic categories (nodes) based on the interview questions were initially defined, followed by case nodes, which were based on individual entries from each interview. Following this, extracts from individual interviews were placed within each row of the matrix. This enabled the analysis of interview data on a case-by-case and thematic basis, during which codes or data extracts changed as patterns emerged (Joyce et al., 2017a; Ritchie et al., 2013).

4.1.1.4 Research Quality

As qualitative research methods were used, research quality focused on credibility, transferability, dependability, and confirmability. Credibility was established by prolonged interaction with thirteen participants knowledgeable in HCI. Transferability was established by purposive sampling of a representative set of those knowledgeable in HCI and sufficiently detailing the context of the study. From this, readers can establish the applicability of findings with other contexts. Transparency was linked to dependability, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. Confirmability was established by clearly and accurately presenting participants' perspectives, whereby transcriptions were revisited as many times as necessary to ensure that themes remained true to participants' accounts. Furthermore, the threat of artificiality, in particular experimenter expectancy (Maruyama & Ryan, 2014) was mitigated by remaining friendly, yet neutral at all times during the interviews.

4.1.2 *Results and Discussion*

The review of the literature demonstrated that the goal of other research efforts is to define a set of usability heuristics for mobile applications. Yet, understanding which sets of heuristics had been embraced by HCI practitioners when evaluating the usability of mobile applications and why this was the case is underrepresented in the literature. This study revealed that heuristic evaluation was not commonly used by HCI practitioners when evaluating the usability of mobile applications (Joyce et al., 2018). Two themes emerged from the data that allowed for a better understanding as to why this might have been the case:

1) Participants believed that good mobile design processes, accumulated knowledge, and usability testing negated the need for heuristic evaluation
2) Participants perceived heuristic evaluation as slow

4.1.2.1 Participants believed that good mobile design processes, accumulated knowledge, and usability testing negated the need for heuristic evaluation

Hussain & Mkpojiogu, (2016, p. 1) observed "Usability has become a central and superior pointer to mobile application design and quality". Yet, equally as important is that users' value mobile applications that are updated often more so than those that are updated less frequently (McIlroy et al., 2016). To address the latter issue, Agile software development processes are gaining popularity (Al-Zewairi et al., 2017). An important aspect of Agile software development approaches is speed, whereby teams make decisions and resolve challenges quickly (Micallef, 2015). To that end, Agile development teams also need insights from users quickly in order to address usability concerns. Should user insights not be presented on-time, a development team will often be put under pressure to release a mobile application or update a mobile application without the insights. This is despite a cautionary note from Sohaib & Khan (2011) that mobile applications will in the hands of users faster, yet the same users will also expect mobile applications to be usable.

In order to move quickly, yet continue to consider usability, participants often needed to justify taking shortcuts. For instance, participants' felt that good mobile design processes negated the need for heuristic evaluation:

- *"There's no call or drive to do a heuristic evaluation. Somebody is already doing the design, I guess some of those practices are being incorporated into it" and "I feel like some of those heuristic things are already maybe being talked about and dealt about there on the fly as they go through, cos they are trying to fail faster", P9*
- *"I feel like the practice of design has moved forward to the point where there's a shared understanding of what constitutes good design." P10*

On one level, participants were not wrong; smartphones have been in existence for over a decade and it was clear that mobile application design matured during that time. However, it would be erroneous to assume the number of issues found during heuristic evaluations would be zero, even for well-experienced product design teams that have multiple feedback loops built into their process. Rather, this mindset may make usability evaluators complacent, whereby they miss potentially critical issues.

Should an evaluation be conducted, many participants felt that they knew what the usability of a mobile application entailed, thus often relying on their own knowledge without the need to consult a set of heuristics:

- *"You're a user experience person, you know all your principles, what's wrong with the site, what can I do differently, what can I improve on, those are the things that half the time are very obvious to you", P1*
- *"You know we don't have anything written down that's like these are best practices, these are the principles", P2*
- *"It's some of my own discovery", P4*
- *"Just based on my understanding of usability", P5*
- *"I use my gained knowledge over the years", P12*

This mindset was cause for concern. Even though the application of accumulated usability knowledge was important, such an approach relies on a substantial amount of experience within HCI to be truly effective. Moreover, without a checklist, such as a set of usability heuristics, Scriven (2005) suggested that both novice and experienced evaluators would overlook key areas during a usability evaluation.

Other than reviewing the design of a software artefact using their own knowledge, participants often relied on tools such as UserTesting.com (*About UserTesting.Com*, n.d.). Such tools allow for quick and easy evaluations of UI's, leading some participants to believe that nothing else is needed—in effect, usability testing with representative users was considered the gold standard:

- *"I do remote usability testing with users on mobile devices", "Usability testing is going to give you more data" and "Nothing is better than user testing in front of the actual people it's intended for", P1*

- *“We use a product called UserTesting.com”, P2*
- *“More testing with the users”, P4*
- *“The way we validate designs, we use prototyping, then we show it to users” and “You know what I did on mobile applications were more like usability studies”, P8*
- *“We were more interested in looking at our audience to see what our audience would think about an application than an expert review”, P9*

The fallacy of this mindset was disquieting as the literature revealed that usability testing tends to be narrowly focused on a handful of representative users as they attempt to complete a small number of tasks over a short period of time. Thus, a substantial number of usability issues can be missed. In addition, usability testing can only be used late in the SDLC when an interactive prototype or working software has been developed, after which many of the design decisions have already been made.

4.1.2.2 Participants perceived heuristic evaluation as slow

Much of the literature refers to heuristic evaluation as a rapid usability evaluation method (Jantvongso & Nuansomsri, 2020; Lyzara et al., 2019). Surprisingly, in this study participants perceived heuristic evaluation as slow:

- *“By the time you provide the information from your heuristic evaluation, chances are that the screens have already changed” and “Going down the guidelines, coming up with the tasks and rules, then going down the guidelines, it's very time-consuming. I mean, you could put together a checklist, and that's fantastic, but the evaluation, it takes a long time”, P1*

- *[When analysing the heuristic evaluation data] "I think too many people may have too...ahhh...many.....conflicting opinions, so it may be more difficult", P3*
- *"If you have five [evaluators], you need to analyse and everyone comes up with different ones, then it increases your decision set of which ones are most important. You may not find a lot of commonality, maybe you will, but that's more analysis, rather than going straight to something that's evidence-based [such as usability testing]", P9*

When heuristic evaluation was originally defined, other approaches, such as usability testing and referencing guidelines, were neither cheap nor easy. As Nielsen stated (1994, p. 413) "...real users can be difficult and expensive to recruit in sufficient numbers to test all aspects of all the versions of an evolving design". Nowadays, this argument is less accurate, whereby users can be recruited relatively quickly and inexpensively. Conversely, Allen et al. (2006) demonstrated that four usability experts reviewing eighteen paper-based screen shots using heuristic evaluation created 108 comments. This is a substantial number of comments to review during an analysis. By the time the analysis is complete, a mobile application design may have been updated and the results of the heuristic evaluation may be outdated. Should there be a need to create a set of heuristics from scratch or add new heuristics to a base set, this could elongate the process even more, potentially by weeks. The usability process as a whole is already perceived to be a delay in building and releasing software to some project stakeholders (Raison & Schmidt, 2013).

When heuristic evaluation was conducted, generally with desktop applications, only one or two evaluators took part in the evaluation:

- *"I have done it on my own. No one is going to do the test again", P1*

- *[Heuristic evaluation will be conducted by] "One or two people", P3*
- *"Usually, it's a one-man show….evaluation", P8*
- *"We would have one person do it", P9*
- *"Two or three people out of a group of maybe fifty. It could be as little as one…it could be as low as one person", P10*

This could be due to saving time within fast-paced Agile environments and/or resource constraints, whereby there can often be a lack of personnel experienced in HCI methods (Borisova, 2019; MacDonald, 2019). Yet, the recommended number of evaluators is three to five (Nielsen & Landauer, 1993). This number of evaluators tends to find the maximum number of usability issues, while keeping costs low. Jeffries et al. (1991, p. 40) agreed with this sentiment when they stated that "A single heuristic evaluation is consistently the weakest way to evaluate an interface, in all the published studies".

Fundamentally, there were two issues to consider, namely that heuristic evaluation was perceived as slow and that fewer than recommended evaluators were utilised. Creating new heuristics from scratch can take quite some time, time that is better spent researching when product teams need insights quickly. One approach HCI teams can take is to review various sets of heuristics that have already been defined before creating a repository from which they can draw upon quickly. This is no different to pattern libraries that design teams often use to save time. Further, the perception that all application screens need to be reviewed against all heuristics is inaccurate. While this might have been encouraged in the 1990's, it is not required today. In many cases, product teams, especially those using an Agile software development process, will have a set of core user stories, use cases or user tasks from which a number of related heuristics and screens can be selected. Such an approach reduces the workload for an evaluator(s) and increases the speed of an evaluation.

Regarding the use of one or two evaluators, not three to five as recommended, this could be something that may not change. In such a case, the limitations of the approach should be made clear with product stakeholders in that there is a strong likelihood that fewer usability issues may be found. However, even if less effective with one or two evaluators, this approach continues to be a pragmatic way to allow heuristic evaluation to bring insights quickly to product teams, thus complementing other methods, such as usability testing.

4.1.3 *Summary*

This section sought to understand if other sets of mobile application usability heuristics were being used by HCI practitioners. Having interviewed thirteen participants, the primary finding was that the HCI practitioners that took part in the study considered their own HCI knowledge, good mobile design practices and usability testing as being good enough. As such, participants believed that there was no need to conduct a heuristic evaluation. Further, heuristic evaluation was perceived as too slow to offer quick insights. These issues were addressed accordingly. The next section focuses on understanding which factors might be included within a set of mobile application usability heuristics.

4.2 What Factors Should Be Included in a Set of Mobile Application Usability Heuristics?

In the previous section, an investigation of the use of mobile application usability heuristics was conducted. The next step was to define the factors that needed to be included within a set of mobile application usability heuristics. Most sets of heuristics tend to be defined and evaluated without a formal methodology (Quiñones et al., 2016). Indeed, Hermawati & Lawson (2016) suggested that there is no real consensus when defining heuristics for specific domains, including mobile applications. Jaferian et al. (2011) maintained that two informal approaches could be taken when defining heuristics—a bottom-up approach can be used to create heuristics from real-world data, and a top-down approach could use researchers' expert knowledge. The authors suggested that combining both approaches might mitigate the weaknesses of each approach, such as researcher bias potentially impacting a top-down approach. The original sets of usability heuristics were created using a top-down approach as they were based solely on researchers' expert knowledge. For example, the set of heuristics from Nielsen & Molich (1990, p. 250) were defined "during several years of experience with teaching and consulting about usability engineering".

Given the informal nature of defining heuristics to date, attempts have been made to formalize a process in recent years. Quiñones & Rusu (2017) suggested that the first step was to identify the features of the application being evaluated prior to defining a set of heuristics. At face value, this approach makes sense. It is likely that such a set of heuristics would work well in terms of number of issues identified, something that no other set of heuristics could replicate. However, the main issue with this approach is that it is specific to a particular application. While the approach might work well within a specific organization, the resulting heuristic set has limited utility outside of an organization. Further, if this approach were applied to mobile applications, the heuristic set may need to be re-defined for every update as certain features may come and go based on user research feedback.

An approach that can be taken to define a holistic set of usability heuristics, which can be more globally applied and have a longer life span, yet also be specific to an application, has been defined by Rusu et al. (2011). In their work, the authors suggested a linear approach comprised of six steps consisting of an exploratory stage, a descriptive stage, a correlational stage, an explicative stage, a validation stage, and a refinement stage.

4.2.1 *Method*

Within this chapter, the first four steps of the approach defined by Rusu et al. (2011) were conducted. Subsequent chapters continued with the remaining two steps.

4.2.1.1 Steps Taken to Define a Set of Mobile Application Heuristics

Step 1 (exploratory stage): During this step, a researcher is encouraged to explore the literature for topics that will assist in the creation of a set of usability heuristics. Potential topics and ideas were found by searching Google Scholar for combinations of keywords: “mobile”, “smartphone”, “usability”, “limitations”, “characteristics” with the period of publication between 1985 and 2020. Any potential topics and ideas were noted, yet there was no order of importance nor criteria other than listing interesting topics and ideas. Additionally, the list of references from the publications found from the Google search were reviewed and any references with relevant titles were read. In total, 105 papers were read, although topics and ideas that formed the basis of a set of mobile application usability heuristics were not gathered from all. Those that did contribute were listed within the references.

Step 2 (descriptive stage): The second step, which Rusu et al. (2011) refer to as the descriptive stage, is where characteristics are summarised from the first step. The summarised list of mobile characteristics is available in Appendix C, with the first six mobile characteristics shown below:

- Users’ awareness of the current mobile application status
- Ensuring that mobile platform consistency, conventions, and standards meet the users’ real-world mental model

- Graceful management of mobile application input errors
- Introducing users to a new mobile application
- Having a user interface that has a clean and simple presentation, while being focused and aesthetically pleasing
- A mobile application that is easy to understand, easy to learn, and easy to use

Step 3 (correlational stage): This step, which Rusu et al. (2011) referred to as a correlational stage, considers which characteristics should be included within a set of heuristics based on traditional heuristics. A number of heuristic sets existed that might have been used as a baseline. However, as Nielsen's (1994) heuristics tended to be the set most applied to mobile applications within the literature (Salgado et al., 2016; Salgado & Freire, 2014), this led to an investigation of Nielsen's (1994) heuristics against the collated mobile characteristics. This approach considered how each of Nielsen's (1994) heuristics either stays the same or where heuristics needed to change when applied to the usability of mobile applications (Joyce et al., 2015). Table 4.3 demonstrates that Nielsen's (1994) heuristics are missing many of mobile characteristics defined within Step 2 (descriptive stage).

Table 4.3. Correlation between mobile characteristics and Nielsen's (1994) heuristics

Characteristics	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**H8**	**H9**	**H10**
Awareness of status	X									
Platform consistency				X						
Manage input errors					X				X	
Introduction to app										X
Simple & aesthetic UI								X		
Ease of use		X								
Clear affordances										
Readability		X								
Ecosystem										
Customization							X			
Consider context of use										
Multimodal data entry										
Utilize camera/sensors										
Memorable app icon										
Screen size/resolution										
Network/Offline										
Storage & power										
UI elements										
Tap targets										
Adequate whitespace								X		
Application crashes										
Trustworthy										
Value proposition										
Essential elements								X		
Application gating										
Relevant adverts										
Battery drain										
Offers privacy										
Push notifications										
UI Localization										
User memory load						X				
Device ergonomics										

Beginning with the first of Nielsen's (1994) heuristics, '1. Visibility of system status'—even though the limitations of mobile devices include less processing power and potentially poor signal connection to cell base stations, both of which may cause delays in a request being processed, a mobile application should at the very least let the user know the request is being processed. This might be through a message in the user's language, a twirling icon, a status bar or anything that signifies that something is happening and the application has not stopped working (Neil, 2014). The reason being that while desktop-based application users can be impatient, which is why Tognazzini (2003) included 'Latency Reduction' in his principles of interaction design—yet, mobile application users tend to be even more impatient (Nilsson, 2009). To summarize, mobile application users should always be kept informed of what is going on through appropriate status notifications, not only within a reasonable amount of time, which is it itself subjective, but right away and for as long as it takes for an operation to finish or timeout. It was suggested that this heuristic should be titled 'Provide immediate notification of application status'.

Moving onto the second heuristic from Nielsen (1994): '2. Match between system and the real world'—mobile application designers should use scenarios (Elkoutbi et al., 1999; Nardi, 1992) and key characteristics of users (Jacko & Sears, 2003) when initially starting work on a mobile application. This allows the identification of the types of users that will use the proposed mobile application and the type of tasks they will attempt. From this, a style guide with the words, phrases and concepts familiar to that type of user would be selected for use on the interface (Galitz, 2007). The style guide will also help to ensure that those words, phrases and concepts are used consistently and in a natural and logical order throughout the application (Stone et al., 2005). Platform and industry standards should also be used as guidelines to apply consistent mapping to user interactions, including touchscreen gesture motions, to ensure interactions occur as users expect. Thus, another of Nielsen's (1994) heuristics '4. Consistency and standards' was removed and weaved into a new more concise heuristic. It is possible to ensure that 'different words, situations, or actions mean the same thing', while also ensuring that they match those expected by different types of users. This heuristic was named 'Use a theme and consistent terms, as well as conventions and standards familiar to the user'. While this heuristic seemed like a barrier to the introduction to new ideas, it was possible to apply consistent mapping to user interactions while introducing new ideas. For instance, a short frequently asked questions (FAQ) page or welcome mat (an interactive style of user help functionality) might be displayed when the mobile application launches for the first-time, which can introduce new ways to interact with the mobile application (Clark, 2010).

The next heuristic from Nielsen (1994) focused on '3. User control and freedom', which is related to navigation and user errors. To that end, this heuristic was partially merged with a single heuristic focused on errors. User errors to one side for now, the ability to navigate is extremely important. As users should be able to see right away how they can interact with a mobile application in order to navigate their way to task completion, this aspect of the heuristic was retained, whereby the heuristic was renamed 'Design a clear navigable path to task completion'.

Returning to the topic of errors, this was the focus of Nielsen's (1994) next heuristic: '5. Error prevention'. Yet, Nielsen (1994) has two error-related heuristics, which were combined to keep the set of mobile application heuristics concise. The other error-related heuristic from Nielsen (1994) is '9. Help users recognize, diagnose, and recover from errors'. Combing both error-related heuristics was an approach also employed by Bertini et al. (2006). Both heuristics were combined into one heuristic 'Prevent errors where possible; Assist users should an error occur'. This approach was influenced by Hoekman (2010), who called for mobile application designers to practice the concept of 'poka-yoke'. In effect, this is predicting the errors that users might make and to ensure they never become problems (Myszewski, 2012). If a situation exists where poka-yoke cannot be applied, whereby potential errors cannot be prevented, such as a navigation error, an incorrectly typed word and so on, users should be allowed to undo their mistake, go back to the previous screen or to exit the application easily if they wish to do so. Should the user try to attempt to proceed when it is not possible to do so, the application should display an error message that is understandable, letting the user know what the problem is, how to solve it and move on.

Moving onto the next heuristic from Nielsen (1994), which is '6: Recognition rather than recall', while an important heuristic, arguably mobile interfaces are quite small, and only the elements and information needed right away should be displayed at any one time. This was argued in-depth in the investigation on Nielsen's (1994) heuristic '8. Aesthetic and minimalist design'. It could also be claimed that having such a small amount of information available on a mobile application interface, using consistent terms, standards and gestures would ensure that users recognize rather than have to recall information or ways to interact with a mobile application. As all aspects of this heuristic were considered, this heuristic was removed.

The next heuristic from Nielsen (1994) is '7. Flexibility and efficiency of use'—some mobile application designers have ensured their mobile applications are far more usable and useful by allowing users to customize the applications to their needs. For instance, news readers that allow users to define the types of news stories they wish to see first. The ability to customize an application was noted by both Bertini et al. (2006) and Inostroza et al. (2012) in their respective heuristics 'Heuristic 6 - Flexibility, efficiency of use and personalization' and '(TMD7) Customization and shortcuts'. Key concepts from both heuristics were adapted in the new heuristic 'Allow configuration options and shortcuts'.

Moving onto the next heuristic from Nielsen (1994) '8. Aesthetic and minimalist design'—while this traditional heuristic seemed quite applicable to the mobile domain, and indeed it was reflected in the user interface principle 'Include in the displays only that information needed by the user at a given time' by Gerhardt-Powals (1996), in its current form the heuristic did not capture precisely the needs of a mobile application. Ideally, a mobile application interface needed to be:

- *Simple* – Hoekman (2010) supported this by applying the Japanese system of the 5S's, described in detail by Ortiz (2012), to mobile interfaces to decide what is absolutely necessary and what could be discounted, arguing that every element chosen caused cognitive load on users;
- *Focused on one task* – Though simple, the interface needed to have all the elements required to complete a singular task, be that to display the information a user needs, to show options or settings, to allow the user to interact with the application and so on. This allows single tasks, even if part of a larger application, to be completed quickly and easily, even by people on the move;
- *Visually pleasing* – While Clark (2010) stated that the beauty of an application is primarily in its functionality, Gong & Tarasewich (2004) suggested that a mobile application will be more prominent if eye-catching. In practical terms, both were correct. A mobile application needs to be functional and aesthetic. Marinacci (2012) even suggested that if an application was visually pleasing, a user would be more inclined to wait longer for information to download, more so than a mobile application that does the same thing but did not look as good;
- *Learnable and Intuitive* – Given the lack of in-depth help and documentation available for mobile devices, as well as the interruptions users face as they interact with mobile applications, an intuitive, easy to learn interface is vital (Lee et al., 2004). In fact, it has been suggested that even though the learnability of a software application has always been

important, it is even more important in the mobile context (Longoria, 2004);

- *Glanceable by allowing main information to be taken in quickly* – Chittaro (2011) was an advocate of mobile application displays that can be viewed at a glance for people on the go. While Bertini et al. (2006) also recognized the importance of glanceable mobile interfaces, the authors decided to create one heuristic to cover different problems in their heuristic 'Ease of input, screen readability and glancability'. Similarly, another heuristic was created by Bertini et al. (2006) named 'Good ergonomics and minimalist design', whereby other aspects of the interface, as well as the design of the actual device itself, were discussed. Some of these issues are beyond the scope of a mobile application evaluation and the influence of application developers. Thus, it was proposed that one heuristic needed to be created that considered the items discussed in this section, while simultaneously omitting items beyond the scope of a mobile application evaluation. Likewise, 'Ease of input' was also omitted as mobile application users either view the information on an interface or they interact with it. Therefore, user input needed to be handled in a separate heuristic. To capture all of this information in an easy to apply heuristic, this heuristic was titled 'Employ a simplistic, focused, glanceable, visually pleasing, intuitive interface'.

Finally, the last heuristic from Nielsen (1994) considered how to assist users, '10. Help and documentation'—while it seemed unlikely that a usable mobile application would require help and documentation, from their findings Bertini et al. (2006, p. 124) suggested that "people using mobile applications still expect such applications to provide help. Though they preferred help to be ‘interactive', non-distractive, not be a separate task”. On a related note, Clark (2010) observed that the use of an interactive ‘welcome mat’ for first timers showing an overlay pointing to the main features of the interface and how to interact with the application made the mobile more appealing. As a welcome mat is interactive, non-distractive and not a separate task, this complemented the findings from Bertini et al. (2006). Once first-time users interact with the application and discover its main features, they can dismiss the welcome mat and later become familiar with the more intricate settings, should they wish to do so.

Despite this, mobile application tutorials are the subject of a strong debate. On one hand, some HCI researchers argue that mobile applications need to be intuitive, and the existence of a mobile application tutorial infers that the mobile application is not as usable and intuitive as it should be (Echessa, 2014). On the other hand, other HCI researchers suggest that learnability has always been important in any context, and that there are times when mobile application tutorials can be useful (Bedford, 2014; Joyce et al., 2016b; Satia, 2014). This is an important debate given the potential for frustrated users should mobile applications be uninstalled if they cannot be learned quickly without mobile application tutorials being available. Prior to mobile application tutorials, software and web application users had printed documentation and online help at their disposal. Yet, the availability of printed documentation and online help proved to be largely ineffective (Grayling, 1998; Novick & Ward, 2006). Based on Clark’s (2010) recommendation, a similar heuristic was retained: ‘Use a welcome mat for first-time users’.

4.2.1.2 Closing the Gaps

Having investigated Nielsen's (1994) traditional heuristics for their applicability to the evaluation of mobile applications, it was noted that important information related to mobile applications and their environment was still absent, which includes:

- *Context of use* – The importance of glanceability was already discussed, whereby a user may glance at their mobile devices while walking, cycling or driving etc. Other types of contextual factors include users having to contend with poor lighting conditions and high ambient noise (Duh et al., 2006). Yet, it is impracticable to cover all scenarios and environments where a mobile application is likely to be used.
- *Content input* – Mobile devices are difficult to use from a content input perspective, and even slower while on the move (Arif et al., 2011). Mobile application designers need to ensure users can input the content they need accurately via large keyboard buttons and multimodal types of input entry (Tan & Lindberg, 2010).
- *Use of sensors* – Mobile devices contain a multitude of complex sensors, such as GPS, accelerometer, gyroscope etc. (Han et al., 2012; Waqar et al., 2011). These can be used within different scenarios, such as changing the mobile interface when a user is driving, to inform users when they are close to one of their friends, as well as for motion gestures as suggested by Negulescu et al. (2012). It is important that designers utilize these sensors as much as possible to provide users with a more stimulating experience.
- *Identifiable, aesthetic icon* – Clark (2010) suggested that an icon for a mobile application should be aesthetic and identifiable as it is the only item a user sees when searching the device interface for the application they

wish to launch. Indeed, a report from comScore asserted that 21% of millennials removed a mobile application as they did not like the logo (*The 2017 U.S. Mobile App Report*, 2017).

4.2.1.3 Research Quality

As qualitative research methods were used, research quality focused on credibility, transferability, dependability, and confirmability. Credibility was established by utilising peer-reviewed sources to determine the factors that should be included within a set of mobile application usability heuristics. Transferability was established by sufficiently detailing the context of the study. From this, readers can establish the applicability of findings with other contexts. Transparency was linked to dependability, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. Finally, confirmability was established by ensuring that there was an audit trail, enabling an auditor to trace conclusions to their respective sources.

4.2.2 *Results and Discussion*

Following a correlational review of Nielsen's (1994) heuristics, an initial set of heuristics for mobile applications was defined, which were named SMART heuristics. This was not an acronym; this was to differentiate the heuristics from existing sets. As previously mentioned, Sharpe et al. (2007) suggested that a set of heuristics should be limited to ten due to memorability. As the suggestion from Sharpe et al. (2007) was not accompanied by empirical evidence, the assumption was tested by adding one more heuristic to the initial set of mobile application heuristics.

With the correlational stage complete, the next phase of Rusu et al.'s (2011) approach was step 4 (explicative stage), which considered and summarised the most important characteristics based on the correlational stage. These characteristics formed the basis of the mobile application usability heuristics (Table 4.4):

Table 4.4. Important characteristics used to form the basis of the mobile application usability heuristics

Important characteristics	Initial set of mobile application heuristics
Offer immediate notification of application status	SMART1: Provide immediate notification of application status
Use a consistent theme and platform conventions and standards	SMART2: Use a theme and consistent terms, as well as conventions and standards familiar to the user
Error-proof as much as possible	SMART3: Prevent errors where possible; Assist users should an error occur
Be helpful for first-time users	SMART4: Use a welcome mat for first-time users
Use only essential elements, be viewable at a glance, be aesthetically pleasing, as well as learnable and intuitive	SMART5: Employ a simplistic, focused, glanceable, visually pleasing, intuitive interface
Offer clear affordances to navigate to task completion	SMART6: Design a clear navigable path to task completion
Allow configuration options and shortcuts to the most important information and frequent tasks	SMART7: Allow configuration options and shortcuts
Consider and cater for diverse contexts of use that mobile devices are used in	SMART8: Cater for diverse mobile environments
Support easy-to-use, multimodal data input entry	SMART9: Facilitate effortlessness input
Make appropriate use of the camera and sensors	SMART10: Make good use of sensors
Use an easily identifiable icon	SMART11: Create an aesthetic and identifiable icon

4.2.2.1 Initial Set of Mobile Application Heuristics

To summarize, the description stage considered many mobile application usability characteristics. The correlational stage positioned these characteristics against existing traditional heuristics. However, there was still a lot of content to mentally process. The explicative stage, therefore, took all of this information and enabled the selection of the most important characteristics for mobile application usability heuristics. All of this information was juxtapositioned against reliable and valid traditional heuristics, which increased the credibility of the new set of heuristics. However, there may have been an element of researcher bias even after the correlational stage as a researcher may inadvertently highlight areas that they are most familiar with. Even conducting a heuristic evaluation against the new heuristics other sets of heuristics may not help mitigate this implied bias. Consequently, steps were taken to mitigate this potential bias by gathering feedback on the initial set of mobile application usability heuristics. Based on the first four steps of the six-step methodology proposed by Rusu et al. (2011), the newly defined mobile application usability heuristics (Joyce & Lilley, 2014), were:

- *SMART1: Provide immediate notification of application status* – Ensure the mobile application user is informed of the application status immediately and as long as is necessary.
- *SMART2: Use a theme and consistent terms, as well as conventions and standards familiar to the user* – Use a theme for the mobile application to ensure different screens look alike. Also create a style guide from which words, phrases and concepts familiar to the user will be applied consistently throughout the interface, using a natural and logical order.

Use platform conventions and standards that users have come to expect in a mobile application such as the same effects when gestures are used.

- *SMART3: Prevent errors where possible; Assist users should an error occur* – Ensure the mobile application is error-proofed as much as is possible. Should an error occur, let the user know what the error is in a way they will understand, and offer advice in how they might fix the error or otherwise proceed.
- *SMART4: Use a welcome mat for first-time users* – A welcome mat displaying the main features and how to interact with the application allows first-time users to get up-and-running quickly, after which they can explore the mobile application at their leisure.
- *SMART5: Employ a simplistic, focused, glanceable, visually pleasing, intuitive interface* – Main interfaces should be easy to learn whereby next steps are obvious, focused on one task, be simple to the point of only having the absolute necessary elements to complete that task which will allow access to vital information while users are interrupted frequently and are themselves mobile, yet the interface should still be attractive and memorable.
- *SMART6: Design a clear navigable path to task completion* – Users should be able to see right away how they can interact with the application and navigate their way to task completion.
- *SMART7: Allow configuration options and shortcuts* – The mobile application should allow configuration options and shortcuts to the most important information and frequent tasks, including the ability to configure according to contextual needs.

- *SMART8: Cater for diverse mobile environments* – Diverse environments consist of different types of context of use, such as poor lighting conditions and high ambient noise are common issues that mobile users have to face every day. Cater for these potential issues, for example by allowing users to change interface brightness and sound settings.
- *SMART9: Facilitate effortlessness input* – Mobile devices are difficult to use from a content input perspective. Ensure users can input content accurately by displaying keyboard buttons that are as large as possible, as well as allowing multimodal input.
- *SMART10: Make good use of sensors* – Utilize the complex sensors available as much as possible to provide users with a more interesting and stimulating experience.
- *SMART11: Create an aesthetic and identifiable icon* – An icon for a mobile application should be aesthetic and identifiable as this is what a user sees when searching the device interface for the application they wish to launch and when scanning through app stores it will be the first item they see before the application title, description and screenshots.

4.2.3 *Summary*

The focus of this section was to better understand which factors should be included within a set of mobile application usability heuristics. To do this, a linear approach comprising of six steps was utilized. One of the steps was to conduct a review of Nielsen's (1994) heuristics to consider their applicability to the mobile domain. From this review, an initial set of heuristics for mobile applications was defined. The next chapter focuses on evaluating the set of mobile application usability heuristics.

Chapter 5. HOW MIGHT A SET OF MOBILE APPLICATION USABILITY HEURISTICS BE EVALUATED?

The second research question focused on evaluating the set of mobile application heuristics defined as part of the first research question. This research question was also split into two sub-research questions. The first sub-research question measured the attitudes towards the mobile application usability heuristics, after which updates were made to the mobile application heuristics. The second sub-research question evaluated the updated set of mobile application heuristics by comparing them to two other sets of usability heuristics, a traditional set from Nielsen (1994) and an alternative set of mobile application heuristics that had been defined by Bertini et al. (2006).

5.1 WHAT IS THE ATTITUDE OF THOSE KNOWLEDGEABLE IN HCI TO THE MOBILE APPLICATION USABILITY HEURISTICS?

Having conducted the explicative stage of the protocol defined by Rusu et al. (2011), which considered the factors that should be included in a set of mobile application usability heuristics, the next step was 'Step 5 (Validation stage)'. This stage ensured that the new set of mobile application usability heuristics did what they are intended to do. Rusu et al. (2011) recommended that a heuristic evaluation is conducted of the same artefact using the newly defined heuristics and a set of traditional heuristics, after which the results should be compared to ensure that the newly defined heuristics were more applicable in terms of the number and types of issues found. To increase the validity of the findings, a step was added whereby participants knowledgeable in HCI had an opportunity to rate the usefulness of each of the initial set of mobile application usability heuristics. Further, participants knowledgeable in HCI were able to offer their own thoughts on anything that was missing which they deemed important to include.

5.1.1 *Method*

There were multiple ways of gathering this feedback. Examples of which were a questionnaire, a series of interviews, or focus groups. The latter two may have taken quite some time to schedule, conduct and analyse. Additionally, the sample size would undoubtedly have been too small, which would have impacted the diversity of thoughts gathered from participants. For this extension of the validation stage, it was decided to design and conduct a questionnaire. There were several existing tools to enable this. While it was far more time-consuming to create a bespoke tool, it was decided to develop a database-driven Web 2.0 application. Not only did this afford a learning opportunity, this approach also allowed the ability to create a custom questionnaire without the limitations of current toolsets. The full application design can be found in detail in Appendix D.

5.1.1.1 Participants

Participants were recruited by emailing 210 authors of articles related to usability and evaluation, in particular the usability evaluation of mobile applications. Sixteen email invites bounced as the email addresses in question were no longer in use. While participants were not paid for this study, potential participants were informed that for each response received £2 was donated to the Sands charity in the United Kingdom (*About Sands*, n.d.).

5.1.1.2 Research Quality

As quantitative research methods were used, research quality focused on reliability and validity. Reliability was increased by using Kendall's *W* (Kendall &

Gibbons, 1990) in SPSS v.26 to determine if there was agreement between raters. Participants statistically significantly agreed in their ratings of the heuristics, $W = .238$, $p < .0005$. Further, the research effort was fully transparent, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. In addition, questionnaire items were clearly written, as were instructions and a clear section within the questionnaire ensured that participants knew what was being asked of them. To that end, other researchers can, if they so wished, replicate the study. Internal validity was increased in a number of ways. First of all, the study had high construct validity by recruiting those knowledgeable in the field who were familiar with the concept of usability, which ensured that instruments measured what was intended. Secondly, instrumentation was consistent throughout the study, even though the questionnaire ran for several weeks. In terms of external validity, the results might be transferable to different people and different settings given that the sample of participants were drawn from those knowledgeable in field of HCI and mobile HCI. Finally, objectivity was increased whereby data was reviewed as many times as was necessary to ensure that participants' perspectives were clearly and accurately presented.

5.1.2 *Results and Discussion*

Sixty participants (n=60) from eighteen countries completed the questionnaire (the full list of countries and number of participants from each country can be found in Appendix E). The study was conducted between November 15th and 27th, 2012 (ethics protocol number: 1112/299) (Joyce, 2013). The majority of the participants were HCI researchers (Table 5.1):

Table 5.1. Roles of participants (n=60)

Role	Number of Participants
HCI Researcher	46
HCI Educator	5
HCI Practitioner	5
Other	4

Four participants listed their role as 'Other', these were participants that were familiar with the field of usability and/or were working in the mobile industry. The majority of participants had up to ten years of experience in HCI (Table 5.2):

Table 5.2. Years' spent in their current role (n=60)

Length of time in role	Number of Participants
Less than 5 years	20
5-10 years	20
11-15 years	13
More than 15 years	7

Each participant read about what was being asked of them, then reviewed the initial set of mobile application usability heuristics. The results of the questionnaire are listed in Table 5.3:

Table 5.3. Rating the usefulness of the mobile application heuristics (n=60)

	Not useful (1)	**Less useful (2)**	**Neutral (3)**	**Useful (4)**	**Very useful (5)**	**Mode**
SMART1	0	2	5	21	32	5
SMART2	0	0	6	25	29	5
SMART3	0	0	4	20	36	5
SMART4	0	12	22	16	10	3
SMART5	1	2	6	19	32	5
SMART6	0	1	5	20	34	5
SMART7	0	3	20	28	9	4
SMART8	3	5	8	24	20	4
SMART9	1	0	7	29	23	4
SMART10	2	5	23	18	12	3
SMART11	1	10	14	17	18	5

This data could have been represented visually in a number of ways, such as a stacked bar chart. However, a stacked bar chart does not have a common baseline, therefore it can be difficult to interpret the results (Robbins et al., 2011). To visualize the results of the questionnaire in a clearer way, a centred stacked bar chart was used, whereby the Likert scale responses equal to 'Neutral (3)' were removed. This approach created a central line equal to zero, thus separating positive and negative responses, which allowed differences to be clearly seen (Petrillo et al., 2011) (Figure 5.1):

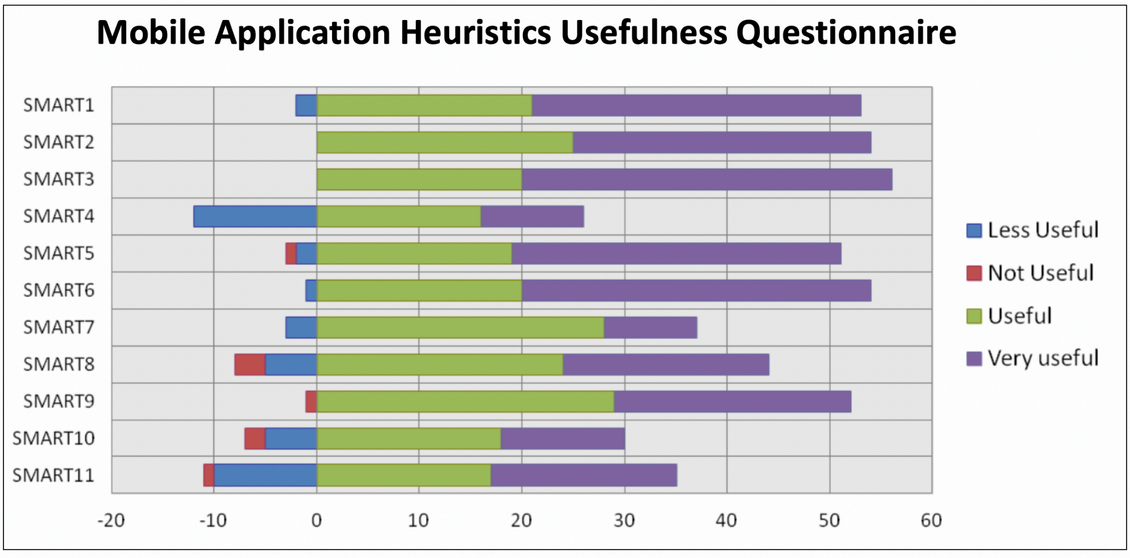


Figure 5.1. Usefulness of mobile application heuristics

The data visualization showed that the majority of participants rated the initial set of mobile application usability heuristics as either useful or very useful. The quantitative results were coupled with encouraging qualitative feedback, such as:

- *"To have a limited set of heuristics to evaluate smartphone applications usability is very interesting. Congratulations for your efforts.", P6*
- *"Great stuff! Please post when it's final, so we can use.", P8*
- *"Interesting topic you have, looking forward to googling the results of this survey later.", P28*
- *"Good work; I'll be attending its results.", P43*
- *"Good idea to revisit these heuristics. Very pleasing survey experience.", P46*
- *"It looks really promising.", P53*
- *"Very interesting study, I will look forward to hearing the outcomes.", P56*
- *"In my opinion, good heuristics based off Nielsen's.", P59*

The quotes above were positive in nature, yet that did not mean that the initial set of mobile application usability heuristics were perfect. The most critical piece of feedback was from P38, who mentioned “All these are good heuristics, but none seem to be exclusive to the limited screen size and bandwidth of mobile devices. You need heuristics pertaining to loading time, network accessibility, size of text and images on the screen and similar factors that are particularly critical in the mobile context”. Initially, this appeared to be an obvious approach for a set of mobile application usability heuristics. Yet, mobile user interfaces, by their very nature, share attributes of non-mobile user interfaces, most of which are well covered by Nielsen’s (1994) heuristics. To that end, it should be no surprise that there would be overlap. The alternative was to create a set of heuristics exclusive to mobile applications, whereby HCI researchers and practitioners would need to evaluate a mobile application using multiple sets of heuristics, such as Nielsen’s (1994) set first, followed by the heuristics defined within this programme of research. For those that already view heuristic evaluation as slow to yield usability insights, this could be problematic. Nonetheless, this critical feedback was neither incorrect nor unwelcome. To that end, care was taken to update the heuristics defined within this programme of research to ensure that it was more transparent that the heuristics were defined for mobile applications.

Shifting the focus back to the quantitative feedback, there was significantly more positive feedback as a whole. Yet, the results also clearly showed that a number of participants considered mobile application heuristics 4, 5, 7, 8, 9, 10 and 11 as less useful without modification. Based on the comments accompanying the ratings, the following changes were made:

‘*SMART4: Use a welcome mat for first-time users*’—it seems the term ‘welcome mat’ was not familiar to all participants. This was encapsulated by a comment from one participant: “Welcome mat - I wasn't exactly sure what this meant. I imagine the first screen should be simple and let users discover what is available”, P8. Another participant commented “Welcome mat should be available also for later use”, P28.

This certainly made sense whereby a user discovers several features they wish to use and would like to see the welcome mat to learn more about the features.

'*SMART5: Employ a simplistic, focused, glanceable, visually pleasing, intuitive interface*'—while this heuristic received a high score in the survey results, some of the comments suggested that this heuristic tried to cover too much. For instance, "SMART5 heuristics would be divided into more heuristics, because they relate to several issues", P6 and "SMART5: Totally agree with this, but it feels like a very broad heuristic!", P8. Subsequently, it was decided to split this heuristic into more atomic parts 'Design a visually pleasing interface' and 'Intuitive interfaces make for easier user journeys'.

'*SMART7: Allow configuration options and shortcuts*'—most power-users use configuration settings, but there are a significant majority of users that rarely, if ever, want to see a complex listing of features for possible configurations. This was reflected in several comments from participants.

'*SMART8: Cater for diverse mobile environments*'—participants pointed out that there are far too many contexts to be able to cater for them all. Yet, there was an opportunity to cater for the most common mobile contexts of use by pointing out usability issues during a standard laboratory-based heuristic evaluation. This idea was tested in the next phase of this work. Other participants pointed out that it was the responsibility of the operating system to adjust screen and sound settings to the environment. Subsequently, this heuristic was updated to reflect the comments pointed out by participants.

'*SMART9: Facilitate effortlessness input*'—for heuristic number 9, one participant commented "I think no. 9 should have a different title - effortless - is probably an unobtainable ideal.", P46. This heuristic was changed to reflect the participant's comment.

‘*SMART10: Make good use of sensors*’—several participants pointed out that sensors and other smartphone functions, including the camera, may not be needed for most mobile applications. Indeed, these may only be needed for certain types of mobile applications. To several participants, the heuristic in its current form was perceived as obligatory in that sensors and the camera must be utilized. Consequently, evaluators using these heuristics would need to highlight the lack of sensor use as a usability problem even if the mobile application being evaluated did not require their use. This is problematic. A modification was necessary that allowed for a more accurate evaluation so evaluators could better determine if the camera and the sensors were used appropriately, if they were needed at all.

‘*SMART11: Create an aesthetic and identifiable icon*’—there were no comments specifically about creating an aesthetic and identifiable icon. Yet, it was clear from the results that several participants did not deem this important or felt it would be difficult to evaluate as the feedback could essentially come down to individual tastes.

5.1.2.1 Updated Mobile Application Heuristics

Following the aggregation, analysis and implementation of quantitative and qualitative feedback received during the questionnaire, the next iteration of the mobile application heuristics was prepared. Each was re-numbered sequentially, and the SMART designation was replaced with an M for readability. The modified mobile application usability heuristics follow:

- *M1: Provide immediate notification of application status.* Ensure the mobile application user is informed of the application status immediately and as long as is necessary. Where appropriate do this non-intrusively, such as displaying notifications within the status bar.

- *M2: Use a theme and consistent terms, as well as conventions and standards familiar to the user.* Use a theme for the mobile application to ensure different screens are consistent. Also create a style guide from which words, phrases and concepts familiar to the user will be applied consistently throughout the interface, using a natural and logical order. Use platform conventions and standards that users have come to expect in a mobile application such as the same effects when gestures are used.
- *M3: Prevent problems where possible; Assist users should a problem occur.* Ensure the mobile application is error-proofed as much as is possible. Should a problem occur, let the user know what the problem is in a way they will understand, and offer advice in how they might fix the issue or otherwise proceed. This includes problems with the mobile network connection, whereby the application might work offline until the network connection has been re-established.
- *M4: Display an overlay pointing out the main features when appropriate or requested.* An overlay pointing out the main features and how to interact with the application allows first-time users to get up-and-running quickly, after which they can explore the mobile application at their leisure. This overlay or a form of help system should also be displayed when requested.
- *M5: Each interface should focus on one task.* Being focusing on one task ensures that mobile interfaces are less cluttered and simple to the point of only having the absolute necessary elements onscreen to complete that task. This also allows the interface to be glanceable to users that are interrupted frequently.

- *M6: Design a visually pleasing interface.* Mobile interfaces that are attractive are far more memorable and are therefore used more often. Users are also more forgiving of attractive interfaces.
- *M7: Intuitive interfaces make for easier user journeys.* Mobile interfaces should be easy to learn, whereby next steps are obvious. This allows users to more easily complete their tasks.
- *M8: Design a clear navigable path to task completion.* Users should be able to see right away how they can interact with the application and navigate their way to task completion.
- *M9: Allow configuration options and shortcuts.* Depending on the target user, the mobile application might allow configuration options and shortcuts to the most important information and frequent tasks, including the ability to configure according to contextual needs.
- *M10: Cater for diverse mobile environments.* Diverse environments consist of different types of context of use such as poor lighting conditions and high ambient noise are common issues that mobile users have to face every day. While the operating system should allow the user to change the interface brightness and sound settings, developers can assist users even more for example by allowing them to display larger buttons and allowing multimodal input and output options.
- *M11: Facilitate easier input.* Mobile devices are difficult to use from a content input perspective. Ensure users can input content more easily and accurately by, for instance displaying keyboard buttons that are as large as possible, as well as allowing multimodal input and by keeping form fields to a minimum.

- *M12: Use the camera, microphone and sensors when appropriate to lessen the user's workload.* Consider the use of the camera, microphone and sensors to lessen the users' workload. For instance, by using GPS so the user knows where they are and how to get where they need to go, or by using optical character recognition (OCR) and the camera to digitally capture the information the user needs to input, or by allowing use of the microphone to input content.

5.1.3 *Summary*

Within this section, sixty participants offered feedback on the initial set of heuristics for mobile applications via a questionnaire. Following the analysis, changes were made to the set of heuristics. The next section focuses on comparing the mobile application usability heuristics to other sets of heuristics.

5.2 How Might the Set of Mobile Application Usability Heuristics be Compared to Other Heuristic Sets?

Continuing with the validation stage, Rusu et al. (2011) recommended that a heuristic evaluation should be conducted of the same artefact using two sets of heuristics, the newly defined heuristics and a set of traditional heuristics. On completion, the results can be compared to ensure that the newly defined heuristics are more applicable in terms of the number and types of issues found.

5.2.1 *Method*

The method applied followed the recommendation from Rusu et al. (2011) in that a heuristic evaluation was conducted of the same artefact, whereby the recommended number of evaluators was three to five (Nielsen & Molich, 1990). It was decided to keep the sample size for this study close to that recommended by Nielsen & Molich

(1990). This differed to the approach taken by several researchers, such as Inostroza et al. (2016), who recruited more than twenty evaluators, generally without HCI experience, to uncover a statistical difference between heuristic evaluation results.

To increase the validity of results with a smaller sample size, the approach taken within this study differed in three ways. Firstly, experienced HCI practitioners were recruited. Secondly, the mobile application usability heuristics defined within this programme of research were compared to two sets of heuristics, not just one. This helped to determine if the mobile application heuristics defined as part of this programme of research could find a higher number and more applicable usability issues than the other two sets. In addition to Nielsen's (1994) traditional heuristics, the other set used was from Bertini et al. (2006) as it was clear during the literature review that this set differed to Nielsen's (1994) set more than other mobile application heuristics. Lastly, immediately after the heuristic evaluation of a mobile application using the three sets of heuristics, an evaluation of heuristics was conducted by the same participants. From the evaluation of heuristics, information was collected in regard to how each heuristic set compared in terms of ease of use, ease of learning, ease of understanding, as well as appropriateness for the usability evaluation of mobile applications.

5.2.1.1 Participants

As three sets of heuristics were used in the evaluation, it was decided to recruit six evaluators in order to more easily randomise the order of heuristics. Several HCI practitioners I was acquainted with from attending UX Conferences were emailed in January 2014. Not all of those who received the initial communication were able to take part, although they were able to recommend others who were subsequently contacted. To that end, the sampling techniques used were non-probability methods, namely purposeful sampling and snowball sampling. The study was conducted between February 24th and March 16th, 2015 (ethics protocol number: COM/PG/UH/00084) with the following participants (2 male, 4 female; Years' experience in HCI: Range=1-20; Mean=7.5; SD=6.92; Years' experience in mobile HCI: Range=0-6; Mean=2.83; SD=2.22) (Table 5.4):

Table 5.4. Heuristic evaluation and evaluation of heuristics participants (n=6)

	Job role	**Gender**	**Years' experience (HCI/Mobile HCI)**	**Personal mobile device**
P1	UX Designer	Male	1/0	iPhone 6 (iOS8)
P2	UX Designer	Female	6/2	iPhone 5S (iOS8)
P3	Director of UX	Male	20/4	Samsung Galaxy S3 (Android 4.0)
P4	UX Designer	Female	2/1	Samsung S5 (Android 5.0)
P5	Lead UX Designer	Female	6/4	iPhone 6 (iOS8)
P6	Senior UX Designer	Female	10/6	Samsung Galaxy S5 (Android 5.0)

5.2.1.2 Materials

Each participant had different types of personal mobile devices with different versions of operation systems. Should each participant install the mobile application selected for the study, each might have had different versions of the application. Further, as the study was conducted over the span of several weeks, different versions of the mobile application of choice may have been used due to updates from the app stores, even if the same participants had the same device type and OS version. To that end, one device was selected for the study, as well as the same version of the mobile application of choice, to mitigate the risk of confounding variables impacting the study results. With the above variables and potential risks in mind, it was decided to use an LG G2 mobile device running Android 4.4.2 for all participants. The mobile application that was evaluated was chosen from the travel category, namely Tripadvisor, version 9.6.1 build 90060034 (*Tripadvisor Android App*, n.d.). Updates on the mobile device were turned off to ensure that all participants used the same version of the Tripadvisor mobile application even though the study was conducted over several weeks. The full set of device and non-device variables for the study can be found in Appendix F.

5.2.1.3 Protocol

Phase 1-Heuristic evaluation: All participants used the same three sets of heuristics, namely Nielsen (1994), Bertini et al. (2006), and Joyce & Lilley (2014). The latter set was defined within this programme of research. All three full sets of heuristics can be found in Appendix G. Each set was labelled with a letter, namely A, B, and C respectfully without the researchers' names that defined each set nor the year of publication. This ensured that participants were less likely to know which set had been defined by each researcher, although some participants may have been familiar with Nielsen's (1994) set of heuristics given their popularity. The study was conducted in-person and time was allowed at the start of the study to ensure that participants were fully familiar with each heuristic set. Table 5.5 shows the order of the heuristic sets which was randomised to mitigate recency bias and learning effect during the evaluation of heuristics:

Table 5.5. Order of heuristic sets used by participants (n=6)

Random order 1	**Random order 2**	**Random order 3**
Bertini et al. (2006) (Set B)	Nielsen (1994) (Set A)	Joyce & Lilley (2014) (Set C)
Joyce & Lilley (2014) (Set C)	Bertini et al. (2006) (Set B)	Nielsen (1994) (Set A)
Nielsen (1994) (Set A)	Joyce & Lilley (2014) (Set C)	Bertini et al. (2006) (Set B)
Used by participants	**Used by participants**	**Used by participants**
Participant 1: Feb 24th, 2015	Participant 2: Feb 26th, 2015	Participant 3: Mar 1st, 2015
Participant 6: Mar 16th, 2015	Participant 5: Mar 15th, 2015	Participant 4: Mar 2nd, 2015

Participants were asked to complete travel-related tasks within the Tripadvisor mobile application. To place participants into a travel mindset, they were given the following scenario: "You are a first-year Bachelor of Science student at the University of Hertfordshire and are new to the Hatfield area. You have decided to install the Tripadvisor application on your smartphone to find hotels and restaurants, including reviews of restaurants you have not been to in the Hatfield area, as well as to find a flight to visit your family in your home country for a few days early in 2015". The related tasks that each participant had to attempt were:

1. Find a hotel near your current location using GPS for one adult that is available within the next two weeks.
2. Find a return flight for one adult in economy class from London Heathrow to Paris on February 20th, 2015, returning on February 24th, 2015.
3. Read at least one review of a restaurant in Hatfield, marking a review you have read as helpful.

Once a participant attempted a task, they would list any issues found using the first set of heuristics. The participant would then attempt the next task, after which they would list any issues found using the first set of heuristics. Once all tasks were attempted and all issues listed under the first set of heuristics, each participant would start the first task once more, and would list issues found using the second set of heuristics and so on. Any issues found would be assigned a severity rating:

- *Minor:* Causes some hesitation or irritation
- *Moderate:* Causes occasional task failure for some users or causes delays and moderate irritation
- *Critical:* Leads to task failure or causes extreme irritation

Phase 2- Evaluation of heuristics: At this point, the approach defined by Rusu et al. (2011) moved into 'Step 6 (Refinement Stage)', whereby the heuristics for a specific domain would be modified based on the results of a heuristic evaluation when compared against traditional usability heuristics. For this study, another sub-step was added in order to increase the validity of the research findings. As discovered from the interviews at the start of this programme of research, the applicability of a set of usability heuristics to mobile applications is just the beginning. The heuristic set may also need to work well within fast-paced Agile environments and/or for HCI teams that have resource constraints. To that end, a set of mobile application heuristics need to be easy to understand, easy to use, and easy to learn. Further, the heuristics should be future-proofed, without the need to be modified constantly as new mobile technologies are released. Finally, those that use the mobile application usability heuristics need to have confidence that the set of heuristics actually work, not just for the mobile application evaluated, but also for other mobile applications. Subsequently, the sub-step added to this study gathered data via a questionnaire immediately after the heuristic evaluation from the same participants. The questions are listed below:

1) Which set of heuristics did you consider to be the most useful in finding usability issues within mobile applications, A, B or C?
2) With regard to heuristic set A (followed by heuristic set B and heuristic set C respectfully), please say whether you agree or disagree with the following statements, using the scale provided, by ticking the most appropriate box:
    a. I felt the set of heuristics was able to locate all the problems in the mobile application used.
    b. The heuristic set would be able to locate most of the problems likely to occur in any mobile application.
    c. The heuristic set would be able to capture problems related to the most recent developments in mobile applications.
    d. I would be confident in using this heuristic set to evaluate usability within mobile applications. in a professional context.

e. The set of heuristics could be applied to a range of mobile devices and screen resolutions.

3) With regard to heuristic set A (followed by heuristic set B and heuristic set C respectfully), please say whether you agree or disagree with the following statements, using the scale provided, by ticking the most appropriate box:

   a. I felt the set of heuristics were easy to use.

   b. I felt the set of heuristics were easy to learn.

   c. I felt the set of heuristics were easy to understand.

4) In the context the evaluation of the usability of mobile applications:

   a. What, if anything, might be changed to heuristic set A to improve it?

   b. What, if anything, might be changed to heuristic set B to improve it?

   c. What, if anything, might be changed to heuristic set C to improve it?

5) Please rank the heuristic sets in term of their usefulness in the context of mobile application usability evaluations.

5.2.1.4 Research Quality

As qualitative research methods were used, research quality focused on credibility, transferability, dependability, and confirmability. Credibility was established by prolonged engagement with six experienced HCI practitioners. Furthermore, time was spent at the start of each heuristic evaluation and evaluation of heuristics to train evaluators fully and effectively. Transferability was established by purposive sampling of a representative set of HCI practitioners and sufficiently detailing the context of the study. From this, readers can establish the applicability of findings with other contexts. Transparency was linked to dependability, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. Confirmability was established by clearly and accurately presenting participants' perspectives, whereby transcriptions were revisited as many times as necessary to ensure that themes remained true to participants' accounts. Selection by maturation was protected against by randomising the order of heuristics, thus mitigating the impact of learning effect or recency bias. The threat of artificiality, in particular experimenter expectancy (Maruyama & Ryan, 2014), was lessened by remaining friendly, yet neutral at all times during the study. Another type of artificiality threat, namely demand characteristics (McCambridge et al., 2012), was reduced by labelling each set of heuristics with a letter, not by the researcher(s) names that had defined each set nor the year of publication.

5.2.2 *Results and Discussion*

Phase 1-Heuristic evaluation: Six participants found a total of 145 usability issues (Mean=48, SD=9) within the Tripadvisor mobile application using the three sets of heuristics. The set of heuristics defined within this programme of research uncovered the highest number of usability issues in absolute terms (Figure 5.2):

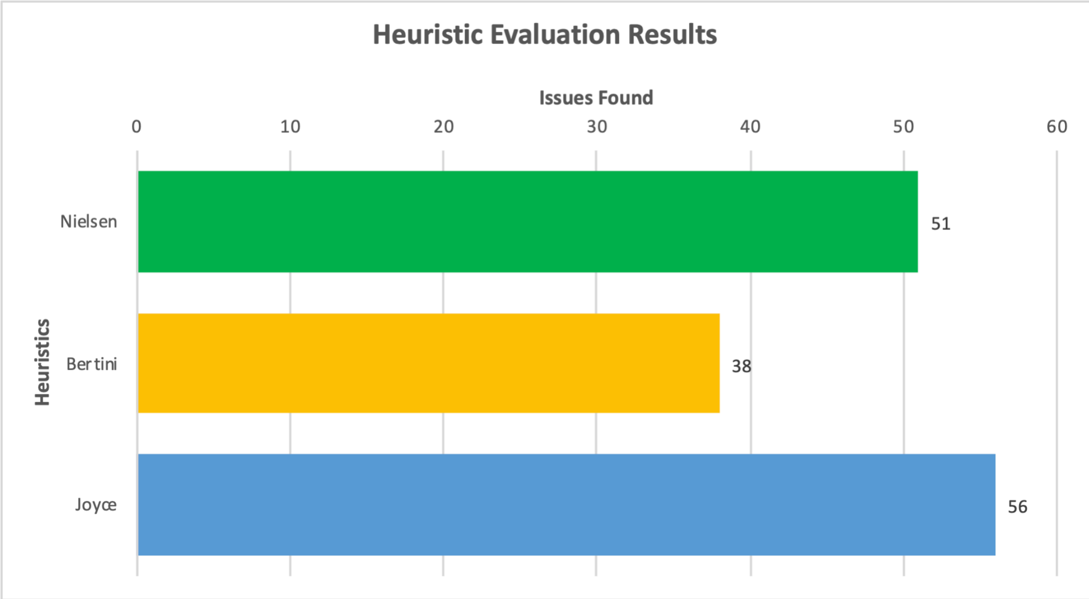


Figure 5.2. Total number of issues found (n=6)

In terms of severity, heuristic evaluation found quite a few minor issues, which is common for the method (Jeffries & Desurvire, 1992). More importantly, from Table 5.6, it can be seen that the set of heuristics defined within this programme of research found the most critical issues:

Table 5.6. Severity of issues found (n=6)

	Nielsen's (1994) (Set A)	**Bertini et al. (2006) (Set B)**	**Joyce & Lilley (2014) (Set C)**
Minor	25	19	32
Moderate	20	17	15
Critical	6	2	9
Total	51	38	56

However, not all heuristics found usability issues. Table 5.7 shows the number and severity of issues found from each heuristic across all three sets:

Table 5.7. Number and severity of issues found per heuristic (n=6)

	H1	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**H8**	**H9**	**H10**	**H11**	**H12**
Nielsen (1994) (Set A)	Crit: 1 Mod: 2 Min: 2	Crit: 1 Mod: 2 Min: 0	Crit: 0 Mod: 0 Min: 1	Crit: 0 Mod: 3 Min: 2	Crit: 0 Mod: 4 Min: 5	Crit: 0 Mod: 2 Min: 1	Crit: 0 Mod: 0 Min: 9	Crit: 0 Mod: 2 Min: 3	Crit: 1 Mod: 1 Min: 0	Crit: 3 Mod: 4 Min: 2		
Bertini et al. (2006) (Set B)	Crit: 1 Mod: 1 Min: 1	Crit: 1 Mod: 2 Min: 0	Crit: 0 Mod: 2 Min: 1	Crit: 0 Mod: 3 Min: 4	Crit: 0 Mod: 3 Min: 4	Crit: 0 Mod: 2 Min: 5	Crit: 0 Mod: 3 Min: 4	Crit: 0 Mod: 1 Min: 0				
Joyce & Lilley (2014) (Set C)	Crit: 2 Mod: 3 Min: 1	Crit: 1 Mod: 3 Min: 4	Crit: 2 Mod: 2 Min: 1	Crit: 2 Mod: 1 Min: 7	Crit: 0 Mod: 0 Min: 1	Crit: 0 Mod: 1 Min: X	Crit: 0 Mod: 0 Min: 0	Crit: 0 Mod: 0 Min: 0	Crit: 0 Mod: 1 Min: 6	Crit: 0 Mod: 0 Min: 5	Crit: 0 Mod: 2 Min: 0	Crit: 2 Mod: 2 Min: 4

Going beyond numbers of usability issues found, it was clear that Nielsen's (1994) heuristics were not suited to mobile applications in many cases. For example, the heuristic 'User control and freedom' was originally designed to capture issues that allowed users to 'leave the unwanted state without having to go through an extended dialogue', as well as a call for the support of undo and redo. Yet, an evaluator in this study used it to log a mobile-specific issue, as no other heuristics within that set were suitable:

- *"The Done button is on the top-right corner which makes it hard to reach.", P2*

Additionally, mobile application users often expect in-context help and tutorials as opposed to traditional online help and documentation (Bertini et al., 2006). Furthermore, when a mobile application tutorial is available, a substantial proportion of first-time users of a mobile application appear to interact with the mobile application tutorial either briefly or in-depth (Inbar et al., 2009; Tokárová &

Weideman, 2013). Yet, some evaluators in this study suggested or inferred that help needed to exist because the heuristic said so:

- *"[I] suggest in-context help.", P5*
- *"I have no idea where the help is, or where to find it.", P6*

Conversely, the heuristics from Bertini et al. (2006) found issues that were more relevant to mobile applications than Nielsen's (1994) heuristics. For example, the heuristic 'Ease of input, screen readability and glancability' from Bertini et al. (2006) found issues related to the ability to read information while on the go:

- *"You can only see two results at a time. It would be nice to see more results above the fold.", P1*
- *"I would need to stop for a moment to read more carefully.", P6*

The heuristic 'Aesthetic, privacy and social conventions' from Bertini et al. (2006) found an issue about privacy that is becoming more important to mobile users depending on the type of data collected (Martin & Shilton, 2016):

- *"I have connected via my Facebook account, so I am assuming my interactions are being tracked somewhere, so not private.", P6*

While the heuristic 'Good ergonomics and minimalist design' from Bertini et al. (2006) could have been clearer in terms of software, not hardware ergonomics, several evaluators brought up related issues:

- *"I couldn't swipe to navigate even though it looks like it's swippable.", P2*
- *"On the map, hotel markers were close together and difficult to target.", P4*

Bertini et al.'s (2006) heuristic 'Flexibility, efficiency of use and personalization' was deemed important for mobile users, in particular the concept of personalisation. For instance, one evaluator mentioned:

- *"I don't see a list of most recently selected airports, or travel routes. It would be nice if the app recognized my regular travel and asked me if I wanted to start from that point.", P6*

The heuristics from Joyce & Lilley (2014) were designed specifically for mobile applications, which was reflected in the issues found. For instance, mobile devices nowadays employ technology that can reduce the burden on users. A heuristic from Joyce & Lilley (2014) recognised this with the heuristic 'Use the camera, microphone and sensors when appropriate to lessen the user's workload'. Several evaluators found usability issues related to this heuristic:

- *"The ability to use microphone to input the name of locations might be helpful.", P1*
- *"GPS...never told me where I was located." and "[If] I could take a photo of the location and add to the review - that would be neat.", P6*

As previously stated, traditional help and documentation are less suited to mobile application users. Related usability issues were raised with the heuristic from Joyce & Lilley (2014) 'Display an overlay pointing out the main features when appropriate or requested':

- *"It would be ideal to have some contextual help indicators for date range selector.", P5*
- *"There was no tutorial on how the Reviews work.", P6*

Given the small size of today's mobile devices, designers should strive to 'Facilitate easier input' according to Joyce & Lilley (2014). Several related usability issues were found in regard to this heuristic:

- *"The keyboard "feels" small, and there is no auto-correct.", P2*
- *"Target for Room selector seems too small.", P5*

The heuristic 'Cater for diverse mobile environments' from Joyce & Lilley (2014) recognised that mobile applications are often used in changing contexts of use. While it was difficult to consider different mobile use scenarios when evaluating a mobile application in a usability laboratory, one evaluator mentioned the following issue:

- *"I saw no indication of the app change to adapt to various scenarios.", P6*

Subsequently, it was considered that in this case the set of heuristics defined within this programme of research compared at least as well or better than other sets of heuristics in finding usability issues within a mobile application; yet the heuristics clearly needed improvement. Ideally, a statistical approach would have been applied, and statistical significance achieved in order to ensure that this result was not by chance. Indeed, as demonstrated within the literature review, this approach has been employed by other researchers. In the case of this study, G*Power (Faul et al., 2009) suggested that at least twenty-eight evaluators would be more appropriate for a parametric test, whereby all evaluators used all three sets of usability heuristics, the significance level was set to 5% ($\alpha = 0.05$), a medium effect size was sought (Cohen's f $= 0.25$), and statistical power was high ($1-\beta = 0.8$). However, with only six evaluators, a one-tailed Friedman test (Friedman, 1940) ($\alpha = 0.05$) suggested that the actual difference between the number of usability issues found across all three sets was not significant at $X^2(2) = 3.739$, $p = 0.077$. This is not to say that there wasn't a difference, it only implies that there was not enough evidence to suggest there was a difference, potentially due to the study being underpowered. This issue is highlighted by Field (2013), whereby statistical significance is easier to achieve with larger sample sizes.

5.2.2.1 Time Needed to Conduct a Heuristic Evaluation

As heuristic evaluation was perceived as a time-consuming approach during a series of interviews conducted as part of this programme of research, the time taken to conduct each evaluation using each set of heuristics was collected. This enabled a better understanding in regard to heuristic evaluation being used within organisations that employ a fast-paced Agile SDLC. It was noted that the heuristic evaluation using Nielsen's (1994) heuristics took an average of 35.6 minutes, the heuristics from Bertini et al. (2006) took an average of 35.3 minutes, and the heuristics from Joyce & Lilley (2014) took an average of 37.6 minutes to complete. The analysis took two days; however, this was for all three sets of heuristics. Thus, conducting a heuristic evaluation was relatively quick, leading to the belief that using a pre-defined set of heuristics for mobile application usability evaluation was suited to a fast-paced Agile SDLC.

Phase 2-Evaluation of heuristics: The second phase of this study was designed to better understand if the set of mobile application heuristics were easy to understand, easy to use, easy to learn, future-proofed, and that participants had confidence that the heuristics work for the application evaluated and most other mobile applications. It was clear from the figures and tables below (Figures 5.3 to 5.5 / Tables 5.8 to 5.10) that the mobile application usability heuristics defined within this programme of research were, for the most part, easy to use, learn and understand, although it could have been argued that they were no easier to use, learn and understand than the two other sets of heuristics:

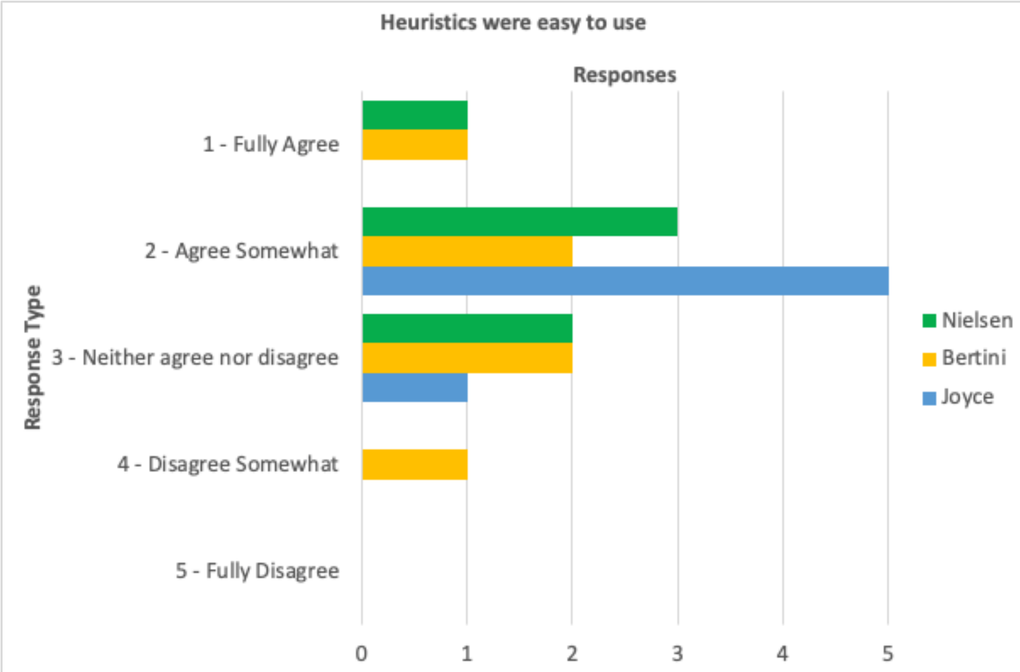


Figure 5.3. Ease of use of the heuristic sets (n=6)

Table 5.8. Ease of use of the heuristic sets (n=6)

	1 – Fully Agree	**2 – Agree Somewhat**	**3 – Neither agree nor disagree**	**4 – Disagree Somewhat**	**5 – Fully Disagree**
Nielsen (1994) (Set A)	1	3	2	0	0
Bertini et al. (2006) (Set B)	1	2	2	1	0
Joyce & Lilley (2014) (Set C)	0	5	1	0	0

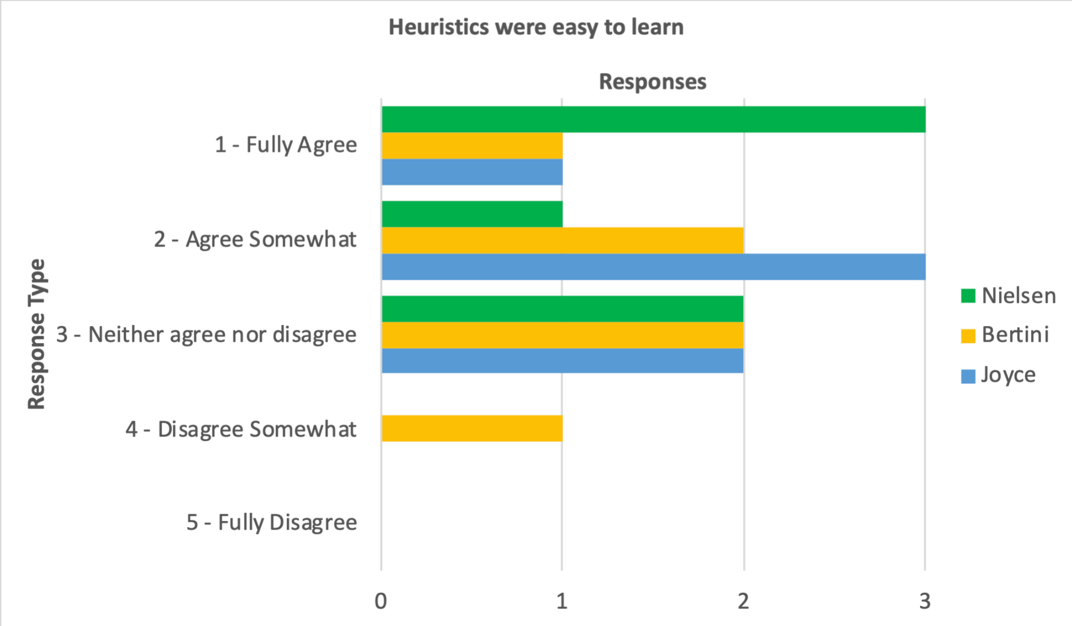


Figure 5.4. Ease of learning of the heuristic sets (n=6)

Table 5.9. Ease of learning of the heuristic sets (n=6)

	1 - Fully Agree	**2 - Agree Somewhat**	**3 - Neither agree nor disagree**	**4 - Disagree Somewhat**	**5 - Fully Disagree**
Nielsen (1994) (Set A)	3	1	2	0	0
Bertini et al. (2006) (Set B)	1	2	2	1	0
Joyce & Lilley (2014) (Set C)	1	3	2	0	0

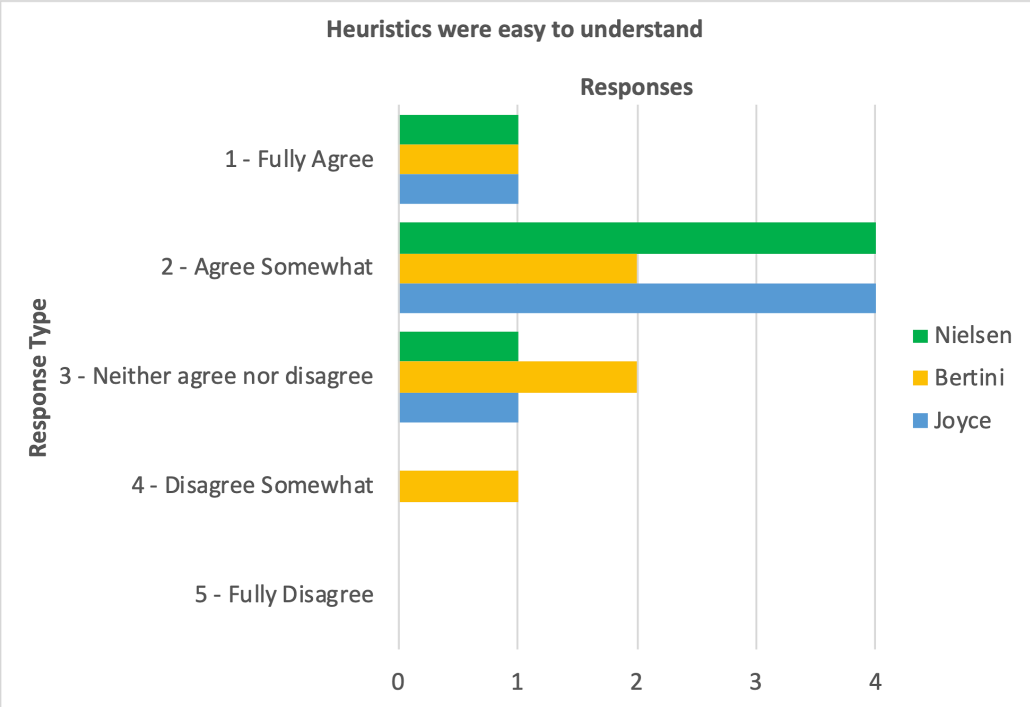


Figure 5.5. Ease of understanding of the heuristics sets (n=6)

Table 5.10. Ease of understanding of the heuristics sets (n=6)

	1 – Fully Agree	**2 – Agree Somewhat**	**3 – Neither agree nor disagree**	**4 – Disagree Somewhat**	**5 – Fully Disagree**
Nielsen (1994) (Set A)	1	4	1	0	0
Bertini et al. (2006) (Set B)	1	2	2	1	0
Joyce & Lilley (2014) (Set C)	1	4	1	0	0

When asked if they would be confident in using these heuristic sets to evaluate usability within mobile applications in a professional context, participants seemed to have more confidence in the mobile application usability heuristics defined within this programme of research (Figure 5.6 / Table 5.11):

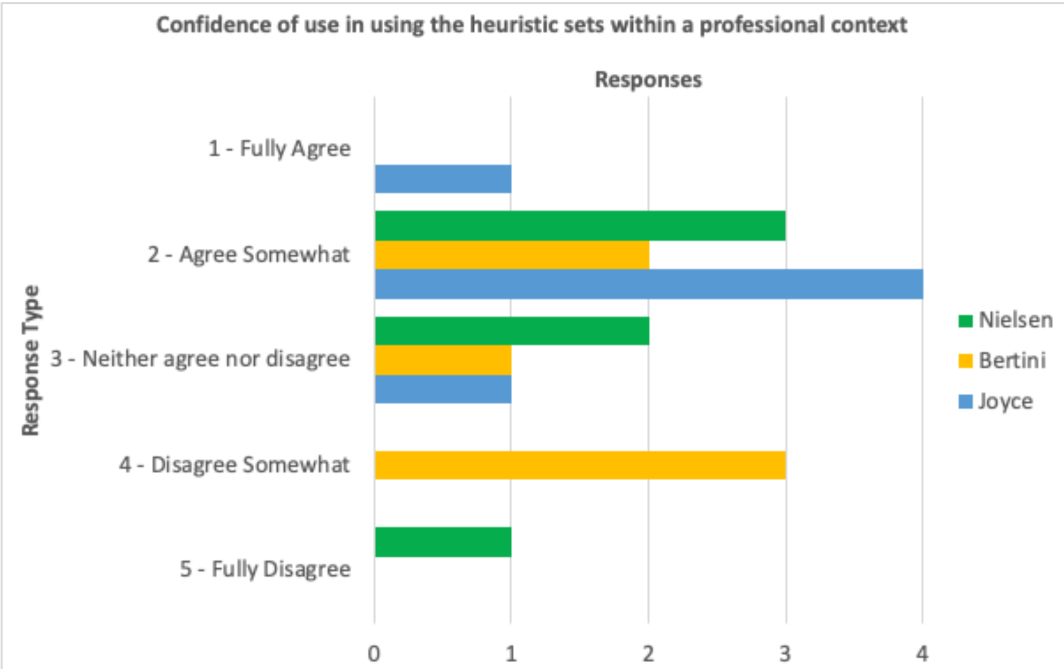


Figure 5.6. Confidence in using the heuristic sets in a professional context (n=6)

Table 5.11. Confidence in using the heuristic sets in a professional context (n=6)

	1 - Fully Agree	**2 - Agree Somewhat**	**3 - Neither agree nor disagree**	**4 - Disagree Somewhat**	**5 - Fully Disagree**
Nielsen (1994) (Set A)	0	3	2	0	1
Bertini et al. (2006) (Set B)	0	2	1	3	0
Joyce & Lilley (2014) (Set C)	1	4	1	0	0

From Figure 5.7 and Table 5.12, it can be surmised that participants were confident that the mobile application heuristics defined within this programme of research would find all of the problems in the mobile application used for the study:

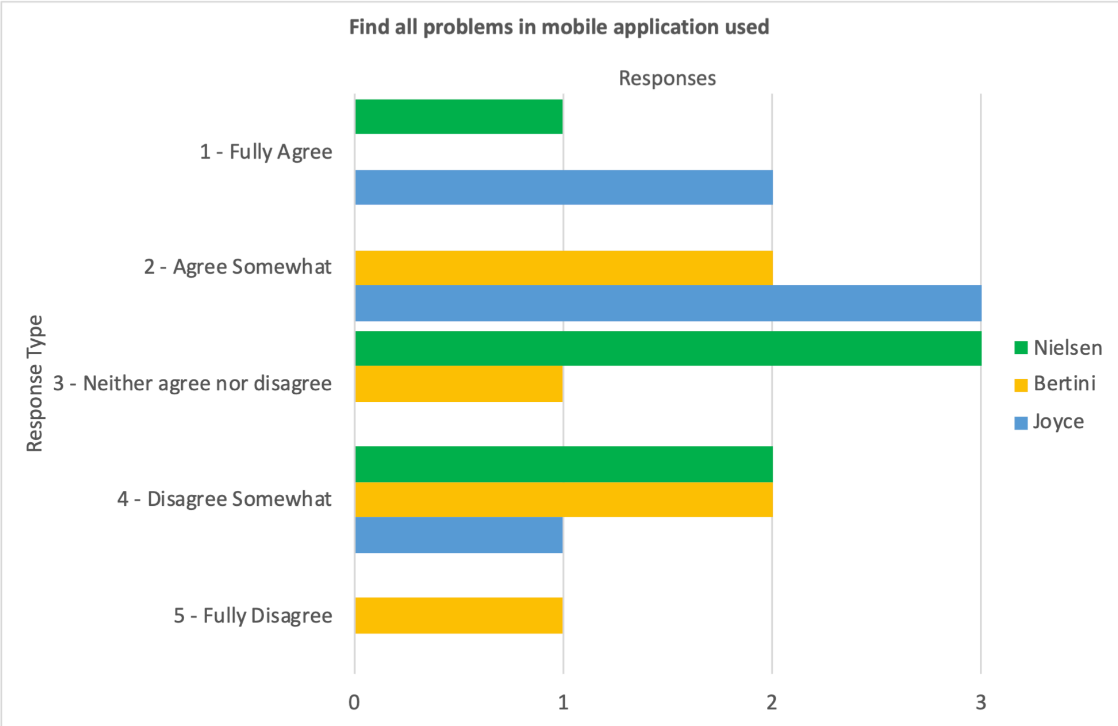


Figure 5.7. Confidence that the heuristics found all of the usability issues in the mobile application used for the study (n=6)

Table 5.12. Confidence that the heuristics found all of the usability issues in the mobile application used for the study (n=6)

	1 - Fully Agree	**2 - Agree Somewhat**	**3 - Neither agree nor disagree**	**4 - Disagree Somewhat**	**5 - Fully Disagree**
Nielsen (1994) (Set A)	1	0	3	2	0
Bertini et al. (2006) (Set B)	0	2	1	2	1
Joyce & Lilley (2014) (Set C)	2	3	0	1	0

It was also important to consider mobile applications beyond the current study, including future developments, lest the mobile application usability heuristics defined within this programme of research became obsolete within a few months or years. To that end, participants were asked if the heuristics sets could be applied to a range of mobile devices and screen resolutions, if the heuristic sets could find most problems in any mobile application, and if the heuristic sets could capture recent developments in mobile applications. It was clear from the results that participants believed that the mobile application usability heuristics defined within this programme of research were the most applicable in terms of application to a range of mobile devices and screen resolutions (Figure 5.8 / Table 5.13) and that the mobile application usability heuristics defined within this programme of research could find most problems in any mobile application (Figure 5.9 / Table 5.14). Participants also believed that the mobile application usability heuristics defined within this programme of research would be able to capture recent developments in mobile applications (Figure 5.10 / Table 5.15):

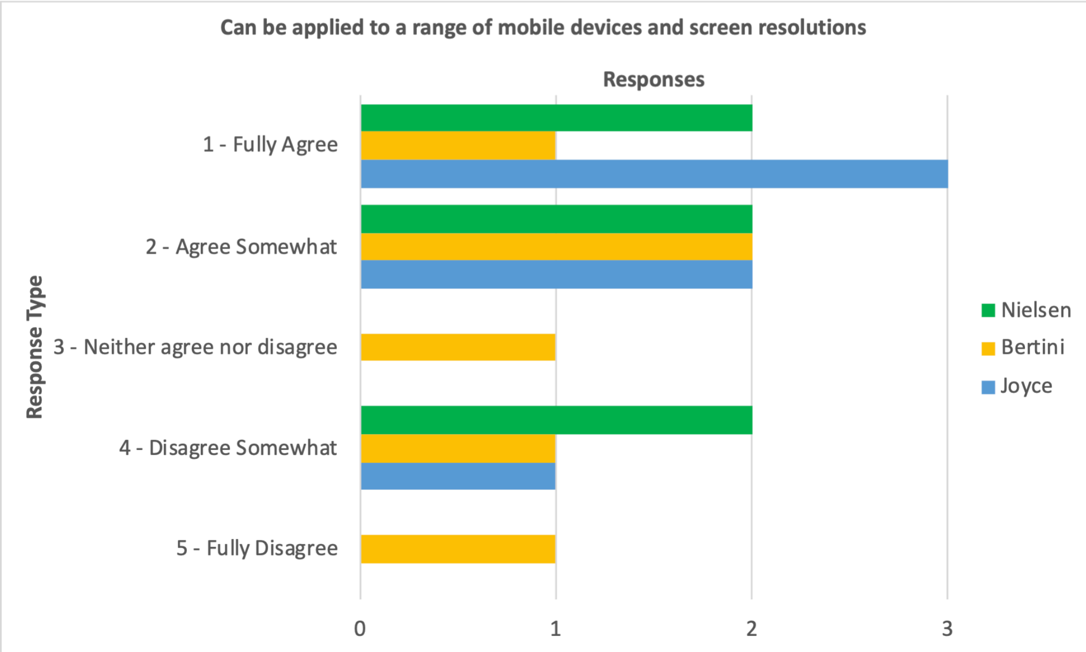


Figure 5.8. Confidence that the sets of heuristics within this study can be applied to a range of mobile devices and screen resolutions (n=6)

Table 5.13. Confidence that the sets of heuristics within this study can be applied to a range of mobile devices and screen resolutions (n=6)

	1 - Fully Agree	2 - Agree Somewhat	3 - Neither agree nor disagree	4 - Disagree Somewhat	5 - Fully Disagree
Nielsen (1994) (Set A)	2	2	0	2	0
Bertini et al. (2006) (Set B)	1	2	1	1	1
Joyce & Lilley (2014) (Set C)	3	2	0	1	0

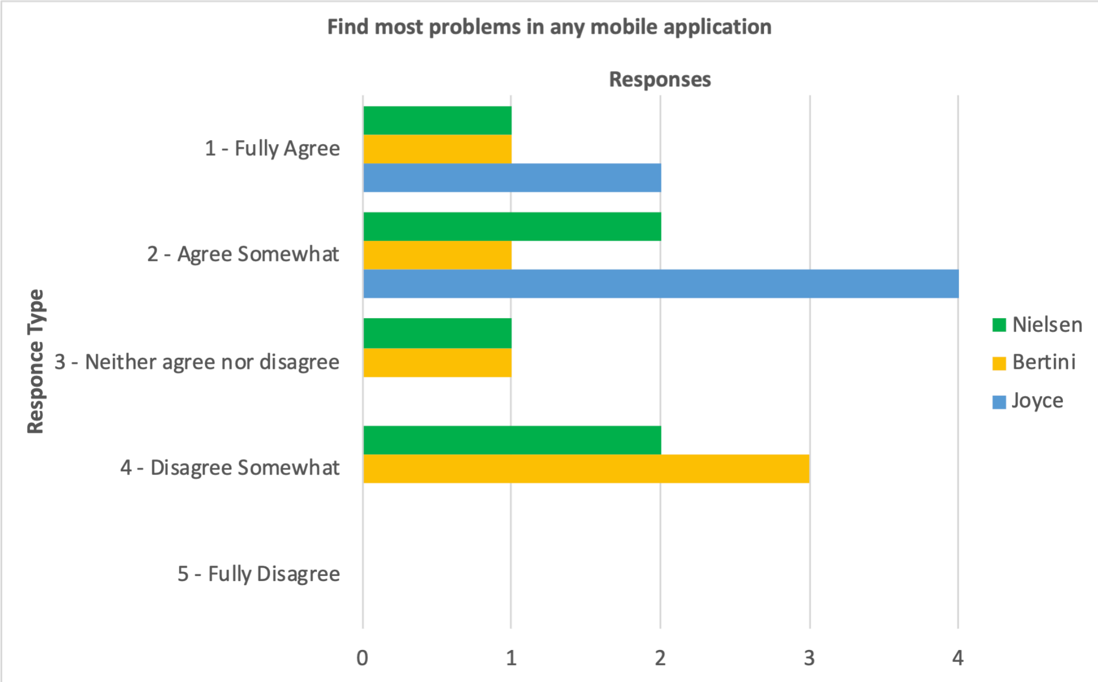


Figure 5.9. Confidence that the heuristics will find most problems in any mobile application (n=6)

Table 5.14. Confidence that the heuristics will find most problems in any mobile application (n=6)

	1 - Fully Agree	2 - Agree Somewhat	3 - Neither agree nor disagree	4 - Disagree Somewhat	5 - Fully Disagree
Nielsen (1994) (Set A)	1	2	1	2	0
Bertini et al. (2006) (Set B)	1	1	1	3	0
Joyce & Lilley (2014) (Set C)	2	4	0	0	0

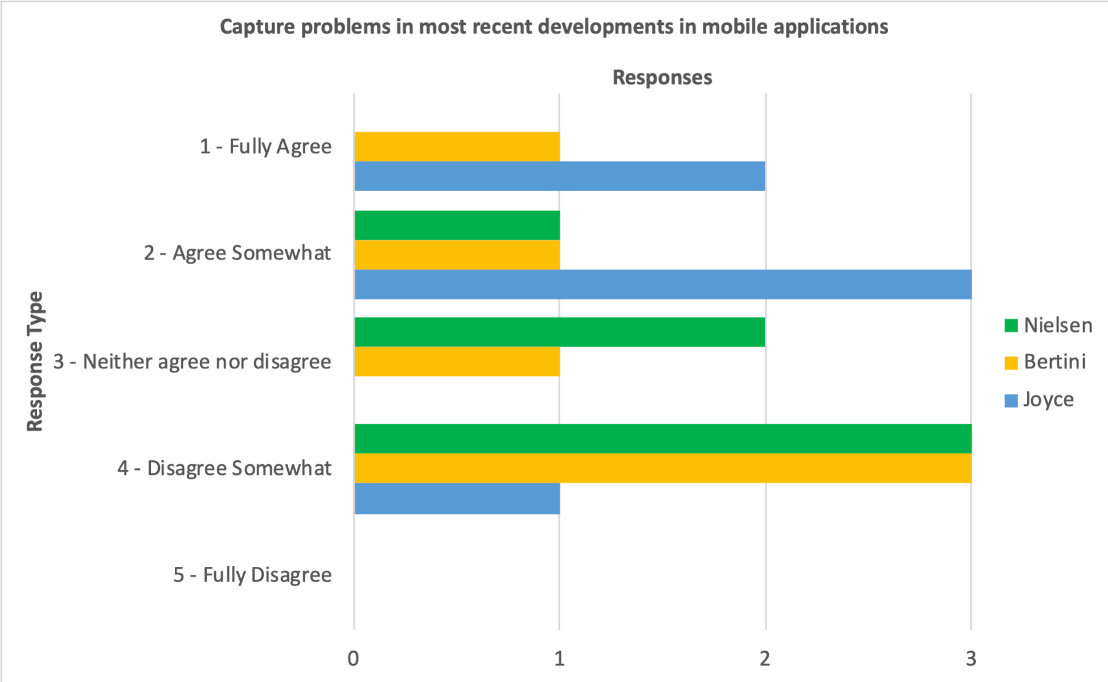


Figure 5.10. Confidence that the heuristics will capture problems in the most recent developments in any mobile application (n=6)

Table 5.15. Confidence that the heuristics will capture problems in the most recent developments in any mobile application (n=6)

	1 - Fully Agree	**2 - Agree Somewhat**	**3 - Neither agree nor disagree**	**4 - Disagree Somewhat**	**5 - Fully Disagree**
Nielsen (1994) (Set A)	0	1	2	3	0
Bertini et al. (2006) (Set B)	1	1	1	3	0
Joyce & Lilley (2014) (Set C)	2	3	0	1	0

When asked to rank heuristics sets in terms of usefulness when evaluating the usability of mobile applications, four participants ranked the usefulness as Joyce & Lilley (2014), Nielsen (1994), then Bertini et al. (2006). The remaining two participants chose Joyce & Lilley (2014), Bertini et al. (2006), then Nielsen (1994). No other combinations were selected by participants. Thus, all participants ranked Joyce & Lilley's (2014) heuristics as the most useful when evaluating the usability of mobile applications.

Finally, when asked what they might change to improve heuristic set C (Joyce & Lilley, 2014), if anything, participants responded:

- *"Think about a way to decrease the number of principles and offer a similar completeness. The description for each heuristic is a bit long. If there was a way to describe each heuristic in one sentence, the set would be much easier to go through and understand.", P1*
- *"Two too many heuristics. If possible, a set of ten works much better (remove or combine). Explanations are a bit too long. It requires extra work (cognitive load) for the users to understand set C. If would be much better if the title (wording) of the heuristic itself is easy to understand, not to mention some may not read through everything. Overall, set C covers essential evaluations for mobile applications.", P2*
- *"I'd use set C and run 100 reviews on a variety of mobile apps and devices to see if further modifications to the heuristics are required. It's hard to tell having just run one test.", P3*
- *"Heuristic 2 could be simplified to say: Develop a pattern or system that resonates with the user. Use consistent language and standards. Heuristic 4 might be too narrow. This suggests that overlays are a*

necessary feature to correctly implement onboarding. I would make this broader to say. Ability to learn the interface with ease. Then, in the details ask: 'Can the user learn features with minimal assistance?'. If onboarding is implemented, will the user be able to accomplish the tasks quickly?", P4

The results of the heuristic evaluation and evaluation of heuristics had practical significance, which in this case consisted of three parts. First and foremost, the perception that a set of heuristics was more applicable toward a specific domain could increase the sense that the heuristics contribute more substantially "to the validity, reliability, and credibility of an evaluation", as per Scriven's (2005, p. 4) theory. In this study, the results of the post-evaluation questionnaire indicated that evaluators considered the heuristic set from Joyce & Lilley (2014) as being most applicable for mobile applications.

Secondly, even if statistical significance was not achieved, any difference between the number of usability issues found could have been detected by effect sizes. As illustrated by Vacha-Haase & Thompson (2004, p. 473), effect sizes should be reported for every statistical study, even for results that are not statistically significant, as "statistical testing cannot evaluate result importance". This is supported by Durlak (2009, p. 917) who contend that "There is no straightforward relationship between a p-value and the magnitude of effect". To that end, a post-hoc analysis with Wilcoxon signed-rank tests (Wilcoxon, 1992) was conducted after the Friedman test previously discussed. A Bonferroni adjustment was applied, which was set to $p = 0.025$, as only the results of Bertini et al. (2006) vs. Joyce & Lilley (2014), as well as Nielsen's (1994) vs Joyce & Lilley (2014) were of interest. Based on Pearson's correlation coefficient r (Sedgwick, 2012), a large effect size was evident between the number of usability issues found by Bertini et al. (2006) vs. Joyce & Lilley (2014), $Z = 1.802$, $p = 0.036$, r = -

0.520. In contrast, the effect size between Nielsen's (1994) vs Joyce & Lilley (2014) was small, $Z = 0.422$, $p = 0.337$, $r = -0.122$. While this study may have been underpowered, the effect size regarding the difference in the number of usability issues between Bertini et al. (2006) vs. Joyce & Lilley (2014) may have indicated that the choice of heuristic set was an important decision to make for an HCI researcher or practitioner as it may impact the results of a heuristic evaluation.

As noted in the results, the types of issues found were an essential consideration. Given the universal nature of Nielsen's (1994) heuristics, evaluators found several issues that might be relevant to any type of user interface. However, evaluators also documented several issues that were associated more so with desktop-based software and websites than mobile applications. In addition, several issues applicable to mobile applications were missed, which were found by the other two sets of heuristics. Moreover, in some cases, evaluators using Nielsen's (1994) heuristics catalogued mobile-specific issues within the closest heuristic, even if that heuristic was a poor fit.

Other feedback focused on simplifying the language and shortening the heuristic descriptions. As per previous studies, there seemed to be potential issues associated with the heuristic about mobile application tutorials. It was also clear from the heuristic evaluation results that several heuristics were not effective. For instance, heuristics 7 and 8 in the mobile application usability heuristics that were defined within this programme of research found no usability issues. Further, while heuristic 10 'Cater for diverse mobile environments' found five minor issues, it could be argued that a heuristic in this form was less effective within an environment with no distractions, and with low ambient noise and good lighting. It was concluded that a heuristic requesting evaluators to consider the potential impact of varying contexts of use on mobile application usability was not an effective approach. However, a heuristic regarding distractions remained in the set as it was important to remind future evaluators that mobile application users are often distracted (Ferreira et al., 2014; Jiang et al., 2018). It is also relatively easy to simulate being distracted in a usability laboratory; therefore, this heuristic can be useful during a standard heuristic evaluation.

There were three steps taken based on this study. The first step was to update the mobile application heuristics based on the results of the heuristic evaluation and evaluation of heuristics (Joyce et al., 2016c). The second step was to change the order of the heuristics to reflect order of importance. Nonetheless, the order of the heuristics was subjective and may change based on the type of mobile application being evaluated. The modified set of heuristics follows:

- *M1 Interaction:* Ensure that tasks can be completed quickly and easily on mobile apps by focusing on specified user goals, minimizing data input, using device capabilities, smart defaults, appropriately sized tap targets, and offering clear affordances.
- *M2 Micro-usage:* Ensure that the mobile application is designed for micro-usage as the user might be frequently distracted.
- *M3 Readability:* Ensure that all elements, including graphics and text, on small mobile application screens are big enough to be readable in portrait and/or landscape modes.
- *M4 Simplicity:* Ensure that the elements, including graphics and text, on each mobile application screen are only those required to achieve a specified goal.
- *M5 Consistency:* Ensure that elements, including graphics and text, used on each mobile application screen are consistent across the application,

while conforming to platform and industry conventions familiar to the user.

- *M6 Errors:* Ensure that the mobile application is tolerant of errors, allowing undo and redo, and using real-time validation of form field input.
- *M7 Responsive:* Ensure that interactions with the mobile application are acknowledged instantly, even if an operation may take time to complete.
- *M8 Learnability:* Ensure that the mobile application is easy to learn, using tutorials or on-boarding if necessary. Tutorials should only focus on critical areas and have a quick way to exit.
- *M9 Personalization:* Ensure that the mobile application can be personalized, customized, and/or configured to suit the user.

During the evaluation of heuristics, it was pointed out by participants that they expected to see ten heuristics in the list, which is in line with the recommendation from Sharpe et al. (2007). Consequently, this was an opportunity to add one more heuristic, which was part of the original mobile characteristics list—that was the ability for a mobile application to be used as part of a wider ecosystem with other devices, such as laptops, TV's, and wearables. In such a scenario, a user could start a task on one device and continue on another device, including mobile devices. This is not critical for all mobile applications, hence the reason why the heuristic was placed last in the set. However, this functionality was becoming more common in cloud-based mobile applications, such as Amazon, Netflix, Spotify, Chrome and others (Jokela et al., 2015). To that end, a tenth heuristic that dealt with this topic was added:

- *M10 Continuity:* Where it makes sense to do so, ensure that the mobile application is part of a continuous experience across other device types.

Finally, the third step was to extend, yet decouple, the mobile application heuristics from elements of context of use given the relative weakness in results when including a heuristic focused on context of use. By extending yet decoupling the heuristics from context of use, the mobile application heuristics can be used as a standalone method within laboratory-based studies, and also use the extended heuristics outside of the laboratory to better understand the impact of context of use on mobile application usability (Joyce et al., 2017b). Consequently, the next two research questions focus on how context of use might be considered when evaluating mobile application usability.

5.2.3 *Summary*

This section documented a study whereby six participants took part in a heuristic evaluation and an evaluation of heuristics, which compared the mobile application usability heuristics defined within this programme of research to two other sets of heuristics. It was surmised that in this case the set of heuristics defined within this programme of research compared at least as well or better than other sets of heuristics in finding usability issues within a mobile application. Based on feedback received during the study, the mobile application usability heuristics were updated. The next chapter focuses on context of use, specifically how the topic might be considered in relation to mobile application usability.

Chapter 6. HOW MIGHT CONTEXT OF USE BE CONSIDERED WHEN EVALUATING MOBILE APPLICATION USABILITY?

Having defined a set of mobile application usability heuristics that might be used within a laboratory environment, the next two research questions focused on the impact of context of use on mobile application usability. The first research question contemplated how context of use might be considered when evaluating mobile application usability. This research question was divided into two sub-research questions, the first of which investigated how context of use was considered by those knowledgeable in HCI when evaluating mobile application usability. The second sub-research question focused on the development of a satisfactory protocol that allowed for the consideration of context of use when evaluating mobile application usability from the perspective of extending the mobile application heuristics defined as part of this programme of research.

6.1 HOW IS CONTEXT OF USE CURRENTLY CONSIDERED WHEN EVALUATING MOBILE APPLICATION USABILITY?

As observed from background research, consideration of context of use is critical within the mobile domain. During the previous study, a heuristic was included to consider the potential impact of context of use when evaluating the usability of mobile applications. That approach proved ineffective, yet the problem cannot be overlooked. Eshet & Bouwman (2014) attempted to discover which methods and tools were being used to understand the impact of context of use. Through the use of a questionnaire with 150 responses predominantly from UX designers (33%), project managers (15%), project owners (9%) and software developers (7%), the authors found that methods used were more suitable to understanding contexts of use at an early stage of a mobile application development project. Yet, such methods were less useful when considering the temporal impact of context of use. The authors followed up a year later with fifteen interviews conducted with HCI practitioners. Concluding this work, Eshet & Bouwman (2015) urged HCI researchers to define new approaches that allowed for a better understanding of the impact of context of use on mobile applications over time. Yet, the authors cautioned HCI researchers working on this problem "to be more thoughtful of the complexity, uncertainty and value conflicts in the problems faced by professional practitioners" (Eshet & Bouwman, 2015, p. 515).

A potential concern was that the questionnaire sent by Eshet & Bouwman during their 2014 study may or may not have been completed by those with experience in mobile HCI. The authors stated that 42% of participants had at least ten years' experience in user-centred design (UCD), and that this level of experience "suggests that they are knowledgeable about developments in UCD, particularly with regard to mobile computing" (Eshet & Bouwman, 2014, p. 7). This is not necessarily the case as not all organisations are likely to develop mobile applications. For those organisations that have mobile applications, it is also possible that HCI practitioners may work on other projects, not on mobile applications. This was evident in the disparity between years of HCI experience versus years of mobile HCI experience in the vast majority of participants that contributed to this programme of research. Thus, while one may assume the results from Eshet & Bouwman (2014) were valid, it was deemed appropriate to conduct a similar study, ensuring that participants stated how much experience they had specifically within mobile HCI.

6.1.1 *Method*

Semi-structured interviews were considered and discounted as such an approach would have collected data from few participants. The approach taken was to use a questionnaire with open-ended and closed questions enabling a large amount of data collection from a higher number of those knowledgeable in HCI within a shorter amount of time. Using this format, participants were able to enter one or more methods that they used to consider context of use when designing or evaluating mobile applications. Responses were counted and categorized.

6.1.1.1 Materials and Procedure

Learning only about the methods used would not have resulted in understanding of the attitudes of those knowledgeable in HCI toward the importance of the potential impact of context of use on the usability of mobile applications. A previous draft of the questionnaire asked, 'Which context of use elements do you feel are important for the usability of mobile applications?'. A concern with this approach was that participants might only bring up the environment and user activity. As such, the final version asked specifically about these two elements, then asked participants 'Other than Environment and User Activity, are there other context of use elements that you feel are important for the usability of mobile applications?'. This allowed participants to think beyond the more obvious aspects of context of use. The study was conducted from January 22nd to March 31st, 2016 (ethics protocol number: aCOM/PG/UH/00107(1)). The questions asked were:

1) Within the field of mobile application usability, how important is the consideration of the following context of use elements:
    a. Environment (Ambient Sound, Surrounding Light Level, Privacy, Interruptions etc.)
    b. User Activity (Lying down, Sitting, Walking, Taking Public Transport etc.)
2) Other than Environment and User Activity, are there other context of use elements that you feel are important for the usability of mobile applications?
3) Do you consider context of use when designing, or evaluating the usability of mobile applications?
4) Please list the specific research tools and methods that you use when considering mobile application context of use (if any)
5) What is your job role?
6) How many years' experience do you have in HCI/UX?
7) How many years' experience do you have in mobile HCI/UX?

6.1.1.2 Sampling

Purposive sampling (Taherdoost, 2016) was used to target those knowledgeable in HCI via LinkedIn groups. LinkedIn groups tend to be active conversations about a specific topic. By requesting that people within those groups take the questionnaire, this helped to ensure that a representative sample was recruited.

6.1.1.3 Analysis

During the analysis, Likert scale data was counted, whereas open question data was thematically analysed. As the analysis was being conducted, it was noted that quite a few answers from participants would not have been considered should closed questions have been used instead of open-ended questions.

6.1.1.4 Research Quality

As quantitative research methods were used, research quality focused on reliability and validity. Reliability was improved by ensuring that the research effort was fully transparent, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings (Nunnally, 1978). As such, another researcher could replicate the study if they so wished. Internal validity was increased in a number of ways. First of all, the study had high construct validity by recruiting those knowledgeable in the field who were familiar with the concepts of usability and context of use, which ensured that instruments measured what was intended. Secondly, instrumentation was consistent throughout the study, even though the questionnaire ran for several weeks. In terms of external validity, the results might be transferable to different people and different settings given that the sample of participants were drawn from those knowledgeable in the field of HCI and mobile HCI. Finally, objectivity was increased whereby data was reviewed as many times as was necessary to ensure that participants' perspectives were clearly and accurately presented.

6.1.2 *Results and Discussion*

There were 156 responses, of which 149 responses were deemed valid and complete—seven respondents were removed from the analysis as those participants stated that they had no experience designing or evaluating a mobile application. The majority of participants were HCI designers (Table 6.1):

Table 6.1. Roles of participants (n=149)

Role	**Number of participants**
HCI Designer	53
HCI Researcher	40
HCI Student	11
HCI Educator	5
Other	40

Forty participants selected 'Other', all of whom were deemed to be familiar with the field of usability and/or were working in the mobile industry, such as mobile consultant, product manager, experience strategist, and mobile applications tester. The majority of participants had between one- and five-years' experience in HCI (Table 6.2), and between one- and five-years' experience in mobile HCI (Table 6.3). As such, one of the critical aspects of this questionnaire, as opposed to that from Eshet & Bouwman (2014), was that it was clear that participants had experience in mobile HCI:

Table 6.2. Years' experience in HCI (n=149)

Years' experience in HCI	**Number of participants**
Less than 1 year	15
1 to 5 years	49
6-10 years	40
11-15 years	21
More than 15 years	24

Table 6.3. Years' experience in mobile HCI (n=149)

Years' experience in HCI	**Number of participants**
Less than 1 year	20
1 to 5 years	88
6-10 years	30
11-15 years	6
More than 15 years	5

The first question focused on how important the consideration of the environment and user activity was within the field of mobile application usability. This question was split into two parts to ensure that it was not double-barrelled. Thus, part a) focused on the environment (e.g. ambient sound, surrounding light level, privacy, interruptions and so on) and part b) focused on the users' activity (e.g. lying down, sitting, walking, taking public transport and so on). In total, 126 of the 149 participants considered the environment's impact on mobile application usability as either very important or moderately important (Figure 6.1):

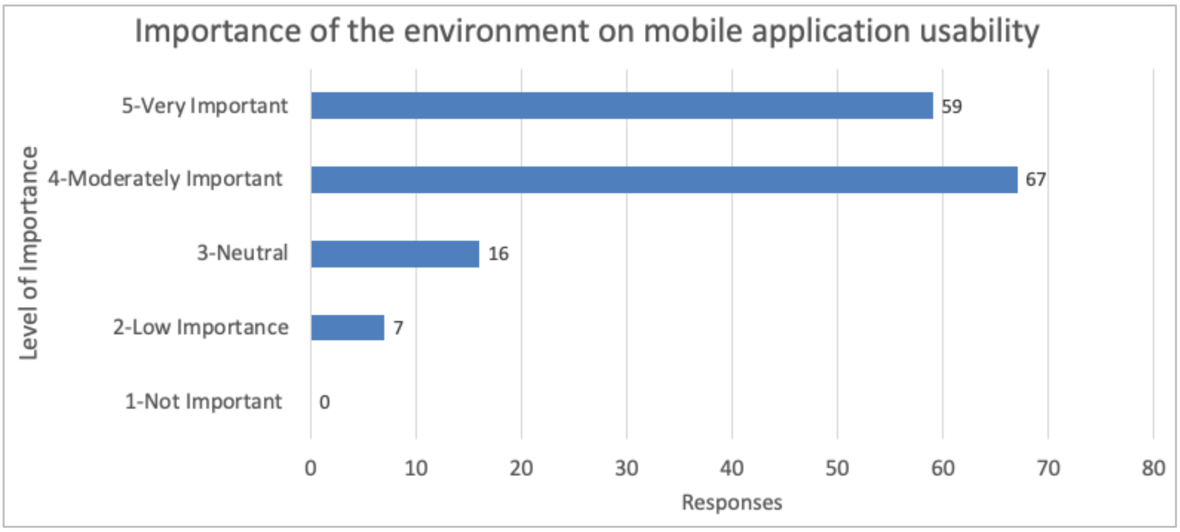


Figure 6.1. Importance of the environment on mobile application usability (n=149)

The responses to part b) were somewhat similar in that the majority of participants felt that users' activities were also important to mobile application usability. In such a scenario, someone that was walking or on public transport may perceive a mobile applications usability as being different to when they were sitting down. In total, 128 of the 149 participants considered the impact of users' activities on mobile application usability as either very important or moderately important (Figure 6.2):

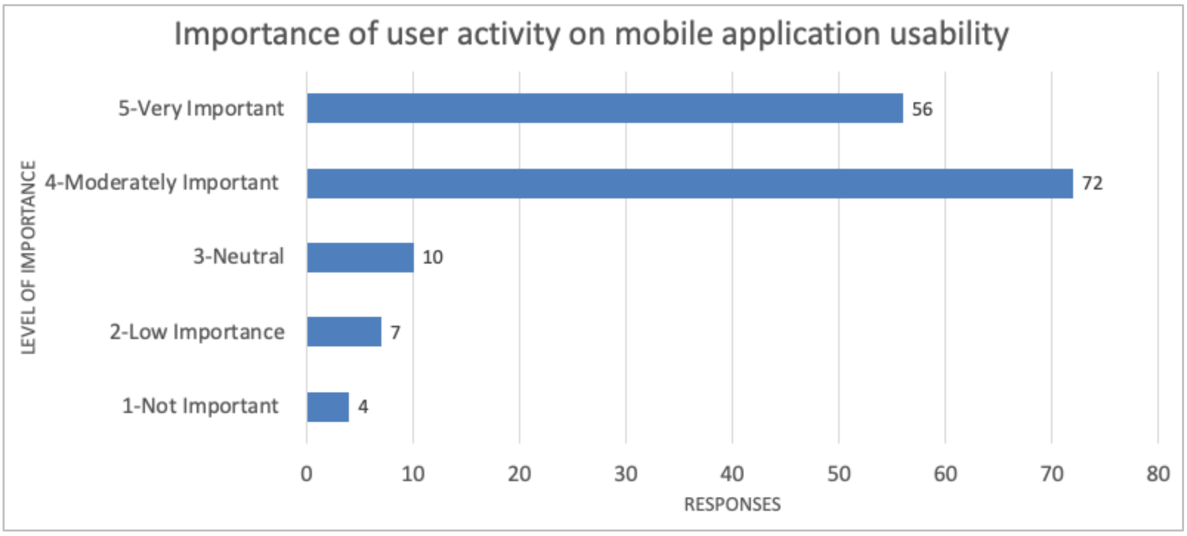


Figure 6.2. Importance of user activity on mobile application usability (n=149)

The next question asked participants about other context of use elements, other than environment and user activity, that participants felt were important to the usability of mobile applications. In total, 213 valid responses were received to this question. These were thematically analysed and categorized into twenty categories:

1) Cognitive load
2) Cultural differences
3) Regulations and policies
4) Users' level of training
5) Distractions and interruptions
6) Time of day/night
7) Network availability
8) Device characteristics
9) Location of users
10) Type of application
11) Interaction with the device or application
12) Users' emotional state

13) Users' characteristics and capabilities
14) Users' goals/tasks
15) Users' specific situations
16) Social context
17) Using the device or application with others
18) Users' expectations
19) Mobile device and application interacting with other devices and applications
20) Time available to use an application or complete a task

Clearly, there was more to consider than just the environment and user activity alone. This list, which supplemented the literature, raised a question regarding the frequency and approach in which HCI practitioners considered these contextual elements. Eshet & Bouwman (2015) claimed that most HCI practitioners do not leave the laboratory when evaluating the usability of mobile applications, and therefore they cannot accurately take the impact of context of use into account. To learn more about this, the next question within the questionnaire was: 'Do you consider context of use when designing or evaluating the usability of mobile applications?'. While the focus was on mobile application evaluation, not design, context of use needs to be considered early in the design phase and not be an afterthought left to the evaluation stage. To this, most participants indicated that they either always or often considered context of use when designing or evaluating mobile applications (Figure 6.3):

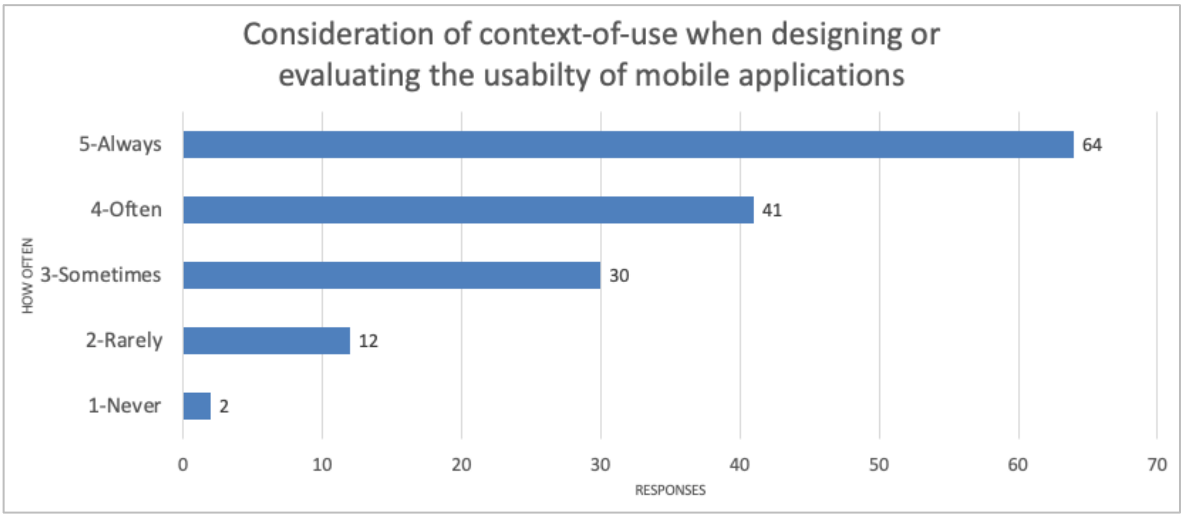


Figure 6.3. Consideration of context of use when designing or evaluating the usability of mobile applications (n=149)

It was a surprise to see that most participants felt that they always or often considered context of use when designing or evaluating the usability of mobile applications. One potential reason for this was that the question was double-barrelled. If the questionnaire was conducted once more, it might be better to either be explicit in that it was wished to learn only about usability evaluation and not design, or to separate these as two questions. The next question asked participants about the specific research tools and methods that they used when considering mobile application context of use, if any. The top ten approaches are listed in Table 6.4 with the full list in Appendix H:

Table 6.4. Top ten tools and methods used to consider context of use when designing and evaluating the usability of mobile applications (n=149)

Tool/Method	**Number of instances within responses**
Ethnographic research	51
Usability testing (total responses/responses that mentioned in-context usability testing)	39/9
Interviews	33
Sketching/Prototyping	18
User stories/Use cases	16
Questionnaires	14
Diary studies	13
Focus groups	9
Market research	8
Personas	8

Table 6.4 above demonstrated the considerable variety in terms of the tools and methods used to consider context of use when designing and evaluating the usability of mobile applications. Furthermore, the list clearly revealed that the majority of approaches are qualitative in nature, which in turn inferred that small sample sizes would be used. Additionally, most of the methods, such as ethnographic research, interviews, questionnaires, diary studies, focus groups, market research and so on,

were best utilised during the research stage of the SDLC, not later in the SDLC when the potential impact of context of use might be better understood. Consequently, an HCI researcher or practitioner may know more about the context of use that mobile application users regularly find themselves in, yet they would have little comprehension regarding the real-world impact of various contextual elements on the usability of a mobile application.

One might argue that some of these methods, such as ethnographic research, interviews, guerrilla testing, and usability testing, can also be conducted when a mobile application is in beta stage or has already been released. With an ethnographic approach, an HCI researcher or practitioner may shadow users to better understand how certain elements of context of use may impact their perception of mobile application usability. However, there are a number of potential problems with this approach. First of all, by its very nature ethnographic research is time-consuming. Thus, only a handful of users can be shadowed during a project, which reduces the transferability of the research findings to other users in the same settings. Should a researcher report that some users of a study within a small sample size were impacted by a certain elements of context of use, that is unlikely to be enough evidence for an HCI team to assign limited resources and funds to tackle a problem that may not exist within the greater population when other high-priority projects are also vying for the same scarce resources and funds. In addition, shadowing users could lead to the Hawthorne effect (Roethlisberger & Dickson, 1939), whereby users may act differently to the way they normally do, biasing the research, which would also impact guerrilla testing and moderated in-context usability testing. To resolve this, one might argue that interviews might be conducted with users to better understand how certain elements of context of use may have impacted their perception of mobile application usability. This approach is not impacted by the Hawthorne effect as users are not shadowed while they use a mobile device. However, this approach lacks ecological validity as it relies on users accurately recalling when, how, and where they used a mobile application, and what, if any, elements of context of use impacted their perception of mobile application usability. Human memory is too fallible for such an approach whereby it would be highly improbable that participants could accurately recall this information hours, days or weeks later (Bannon, 2006).

Several participants mentioned that they would make assumptions; this is not an approach that should be encouraged. One participant responded with 'Social Media'—this is one of the downsides of the questionnaire method, in that it was unsure exactly what the participant meant and there was no way to probe as there might have been had interviews been used. Other approaches uncovered during the questionnaire included heuristic evaluation, context-aware sensors, and eye tracking. As noted, heuristic evaluation is not suitable when attempting to understand the impact of context of use on mobile application usability. Context-aware sensors can miss or misinterpret contexts. Eye-tracking can be conducted when a mobile application is at a prototype or a fully developed stage. In this case, a researcher needs to calibrate eye-tracking glasses for each participant, then shadow the participant as they use the mobile application in a natural setting (Pérez-Edgar et al., 2020). This is not only costly; it is also time-consuming. Further, the output will tell a researcher what the participant looked at, but not necessarily how much context of use had an impact. One might combine eye-tracking within multiple scenarios to see what the differences are between contexts of use and combine the approach with interviews. This might allow the researcher to better understand the impact of context of use, but again, it is unlikely that the results can be transferred to other settings or users as only a handful of users will be able to take part given how time-consuming the approach is. This study demonstrated that there were few effective approaches being used to understand how context of use impacted the perception of mobile application usability, which supported the findings from Eshet & Bouwman (2014).

6.1.3 *Summary*

This section moved away from the mobile application usability heuristics and focused on context of use. A total of 149 valid responses were collected using a questionnaire, which sought to understand the areas of importance in relation to understanding context of use. In addition, it was learned that few effective methods were in use that allow for an understanding of the impact of context of use on mobile application usability. The next section focuses on defining a satisfactory protocol that allows for the consideration of context of use when evaluating the usability of mobile applications.

6.2 What is a Satisfactory Protocol That Allows for the Consideration of Context of Use When Evaluating Mobile Application Usability?

The second sub-research question contemplated how to define a satisfactory protocol that allowed for the consideration of context of use when evaluating mobile application usability. Yet, understanding the impact of context of use in relation to mobile usability is an inherently complex topic. One potential approach was to reduce complexity, where possible, when defining the protocol.

6.2.1 *Method*

The method used to define the protocol was to build upon the work of others and on work conducted within this programme of research, including the mobile application usability heuristics. An example of the work of others whereby complexity was reduced was the Agile SDLC. This differs from the waterfall SDLC, a distinct, phase-bound approach that results in complex software being developed over several years, often resulting in software that does not meet customers' needs (Schwaber & Beedle, 2002). An Agile SDLC changes the distinct Waterfall phases into bite-sized user stories, which allows for faster software development releases, which in turn enables more frequent customer feedback. One of the primary ways this is accomplished is through the use of Agile user stories, which follows the format: As a <type of user that wants to accomplish something>, I want to <what the type of user wants to accomplish>, so that <why the type of user wants to accomplish that thing>. The user story and Agile SDLC as a whole have endeared themselves to many organisations around the world. A questionnaire conducted in 2020 by CollabNet reported that 95% of 1,121 respondents practiced Agile software development, 34% of whom stated that they have been using Agile for the past three to five years, and 27% of whom stated that they have been using Agile for more than five years (*14th Annual State of Agile Development Survey*, 2020).

By following a similar approach, and by taking findings within this programme of research into account, the needs of a satisfactory protocol could be defined. To that end, the proposed protocol needed to meet the following criteria:

- *Reduce complexity:* The protocol needed to reduce the complexity of understanding the impact of context of use on mobile application usability;
- *Be flexible:* The protocol needed to consider the disparate areas of context of use and to build upon the mobile application usability heuristics previously defined within this programme of research;
- *Better use of resources:* As the HCI practitioners that took part in earlier interviews used only one or two resources per heuristic evaluation, the protocol needed to use a small number of HCI resources. Ideally, this

entailed only one person with an HCI background as a moderator. This meant that participants needed to be recruited, which is not a normal practice within a standard heuristic evaluation;

- *Relatively fast:* The protocol needed to enable the gathering of insights relatively quickly, ideally within several hours at a minimum or several days at most, in order to inform the design of a mobile application (Joyce et al., 2016a). This would suit both Agile and non-Agile SDLC environments;
- *Offer directional and/or statistical insights:* The protocol needed to offer at least directional insights and potentially statistical insights. This ensured that any negative impact of context of use on mobile application usability was resolved as much as possible once an HCI team deemed that enough evidence had been gathered.

In addition to the criteria listed above, it was important to state what was effectively outside of the scope of the protocol. The protocol allows those knowledgeable in HCI to better understand how context of use may impact the perception of mobile application usability. Conversely, the protocol was not designed to discover the actual contexts within which a mobile application is used. This is a different use case. It is important that such research is conducted in advance of designing the mobile application to better understand how, where and by whom the mobile application might be used. This early research allows HCI teams to understand actual contexts of use, but it cannot directly measure the potential impact of contextual elements on mobile application usability.

6.2.2 *Results and Discussion*

To build the foundation of the protocol, which was named the contextual usability evaluation protocol, the traditional Agile user story format was re-framed as a testable hypothesis: When <attempting a task>, does <an element of context of use> have an impact on <mobile application usability>? While it is up to each HCI team to define what is important to test, Table 6.5 to 6.7 demonstrates examples of tasks, contextual elements and context of use areas that might be used as a guide and extended upon. The mobile application usability heuristics defined within this programme of research were integrated within contextual usability evaluation as part of the mobile application usability segment. These contextual usability evaluation tables were not meant to be exhaustive—different HCI teams working on different mobile applications can create their own tables or extend these tables for their own use:

Table 6.5. Examples of tasks

When <attempting a task>
• Making a call • Composing a text message • Texting a friend • Viewing a photo/video • Hailing an uber • Transferring funds • Sending an email • Reading stories on Medium • Saving files to DropBox • Looking at photos on Instagram • Buying an item on Amazon • Watching a video on YouTube • Finding directions on Google Maps • *Extend as needed…*

Table 6.6. Examples of contextual elements

Does <an element of context of use>		
Location	*Environment*	*Activity*
• Usability lab	• Sound	• Lying Down
• Home	• Light	• Sitting
• Office	• Space	• Walking
• Street	• Distractions	• Driving
• Pub	• Other people	• Eating
• Airport	• Privacy	• Juggling groceries
• Railway Station	• Weather	• Waiting for public transport
• University	• Construction	• Taking public transport
• Café	• Pollution	• Waiting in line
• Hospital	• *Extend as needed…*	• On an elevator
• Gym		• Right, left-Handed or cradling
• Store		• *Extend as needed…*
• *Extend as needed…*		

Connection	*User Attribute*
• Offline	• Accessibility:
• Speed	• Colour-blind
• Reliability	• Motor disability
• *Extend as needed…*	• Poor eyesight
	• Continuous/full
	• Continuous/partial
	• Intermittent/full
	• Intermittent/partial
	• Age
	• Experience
	• Confidence
	• Emotion
	• High stress level
	• *Extend as needed…*

Table 6.7. Examples of mobile application usability elements

Have an impact on <mobile application usability>
• M1 Interaction • M2 Micro-usage • M3 Readability • M4 Simplicity • M5 Consistency • M6 Errors • M7 Responsive • M8 Learnability • M9 Personalization • M10 Continuity • *Extend as needed…*

From these tables, HCI teams can define testable contextual usability evaluation user stories, such as:

- When reading the news, does standing on the subway at rush hour have an impact on readability of all screen elements in landscape and/or portrait modes?
- When waiting on a text message, does being surrounded by friends in a noisy environment have an impact on knowing the status of the text message?
- When changing music stations on Spotify, does walking on a busy street have an impact on completing the task quickly and easily?

6.2.2.1 Measuring Impact on Usability

Two parts of the contextual usability evaluation were measurable, i.e. the intensity of a contextual element and the resulting impact on usability. For example, should an HCI team be interested in the impact of certain elements of the environment, such as sound, they might ask participants within a study to rate the intensity of this element.

An example might be offered given the subjectivity of these ratings. For instance, sound-related Likert scale items might be measured as:

- 1-Faint (Leaves rustling)
- 2-Soft (Quiet library)
- 3-Moderate (Normal conversation, moderate rainfall)
- 4-Loud (Traffic, alarm clock)
- 5-Very loud (Rock concerts, Lawnmowers, Hair dryers, Blenders)

Labelling each Likert item increased objectivity and would work well with a limited number of variables, yet naturally occurring contexts of use can be complex. This would need to be done for each and every variable, including variables that are rarely, if ever, evaluated. Before too long, the protocol that had originally been envisioned would also become quite complex. One option was to stay true to the roots of the protocol by measuring only one variable—the perceived impact on mobile application usability. In such a scenario, the complexity of the protocol was reduced. However, the subjectivity and complications of context of use cannot be overlooked. An approach was to consider such complications during the early phase of research using Maguire's (2001a) context of use protocol, within which the author allows for the specification of varying degrees of context within a 'User requirements or test conditions' column.

Consequently, only the perceived impact on mobile application usability needed to be measured. This was done on a Likert scale from 0 (no impact) to 7 (severe impact). The decision to use an 8-point unipolar Likert item was decided upon as Schwarz et al. (1991) claimed such a Likert item was reliable, whereas Sauro & Lewis (2020) maintained that labelling only items 0 and 7 did not decrease the validity of subsequent findings.

6.2.2.2 Contextual Usability Evaluation Protocol

The following steps were put in place to ensure that a contextual usability evaluation was carried out correctly:

- *Step 1 - Create contextual usability evaluation user stories:* A single HCI researcher or practitioner sets up the study by creating contextual usability evaluation user stories. These are informed by prior research, such that only the most critical contextual elements are selected for evaluation. The HCI researcher or practitioner will not accompany participants during the study. This reduces the potential for the Hawthorne effect (Roethlisberger and Dickson, 1939) and allows the HCI researcher or practitioner to conduct the study with many participants simultaneously;
- *Step 2 - Recruit participants:* Ensure participants can conduct the requested tasks using their personal smartphones within specified naturally occurring contexts of use. Recruiting less than twenty-nine participants will likely offer directional insights. Recruiting at least thirty participants will likely offer enough power to conduct a statistical test to better ensure that the results do not occur by chance;
- *Step 3 – Conduct contextual usability evaluation:* Instruct participants in what is expected of them, then conduct the study, whereby all participants conduct all specified tasks within all specified conditions. Data can be sent back to the moderator via a survey system on a mobile device. Two additional items may be requested from participants:
    - *Proof:* As this is a remote, potentially crowdsourced study, a photo or short video may be requested from each participant to ensure that the participant was truly subjected to the specified contextual

element. Photos and videos can be copied to a cloud-based file location, such as Dropbox;

  - *Comment:* Given the complexity of context of use, a test result may be impacted by one or more contextual elements. For instance, a task might request that participants type a text message while walking in direct sunlight. If participants state that this has an impact, then it is more difficult for an HCI researcher or practitioner to understand which is causing the impact, walking, direct sunlight, or both. To that end, allowing participants to enter comments helps to better understand the cause of the impact. In addition, it is also possible to ask participants to measure the impact of each contextual element separately. For example, they might use the 7-point Likert item twice, once for walking and once for direct sunlight;

- *Step 4 - Create visualization:* The results of contextual usability evaluation can be visualized in a number of ways, such as a dot plot, that can offer directional insights for the HCI researcher or practitioner and their respective stakeholders;
- *Step 5 - Calculate statistics (optional):* Conducting an optional statistical analysis allows an HCI researcher or practitioner to ensure that the difference in perceived impact between contextual usability evaluation user stories, if any, is not by chance. If the results of two contextual elements are compared, and approximately thirty participants have been recruited, it is possible to conduct a statistical analysis of the results using a Mann-Whitney U test. If three or more contextual elements are compared, and approximately thirty participants have been recruited, a Friedman test with post-hoc Wilcoxon signed-ranks tests is appropriate;

- *Step 6 - Report results:* Once the visualization (and statistics if these were completed) are ready, they need to be presented to key stakeholders quickly, especially in Agile SDLC's.

6.2.3 *Summary*

This section aimed to define a protocol that allowed an understanding of the impact of context of use on mobile application usability. To do this, the complex topic of context of use was reduced by modelling the protocol on the user story format commonly used within Agile SDLC's. The next chapter focuses on the evaluation of the contextual usability protocol defined within this chapter.

Chapter 7. HOW MIGHT A PROTOCOL THAT CONSIDERS THE IMPACT OF CONTEXT OF USE ON MOBILE APPLICATION USABILITY BE EVALUATED?

A number of approaches might have been used to evaluate the protocol, such as to conduct a study using the protocol. However, such an approach would only have offered a visualization and a statistical result. While that might have been useful for an HCI researcher or practitioner, the robustness of the protocol in terms of the consideration of real-world, naturally occurring contexts of use would be uncertain. It was decided that a more effective evaluation was a three-step process. Firstly, the protocol was evaluated in terms of being robust enough to consider diverse contexts of use within which mobile applications are used. Secondly, the attitudes of those knowledgeable in HCI towards the protocol were gathered, following which any concerns uncovered were addressed. Finally, a workshop was conducted to understand how those knowledgeable in HCI might determine the scope of a typical contextual usability evaluation.

7.1 How Well Does the Protocol Consider the Actual Contexts of Use that Mobile Applications are Used in?

The previous chapter focused on defining the contextual usability evaluation protocol. To ensure that the protocol was able to handle diverse user stories created from real-world tasks and contexts of use, it was decided to learn about the types of mobile applications that people use, the types of tasks they attempt using these mobile applications, and the types of contexts of use they find themselves in when using their mobile device.

7.1.1 *Method*

The method chosen for this study was thirty semi-structured interviews, primarily because the approach allowed for probing on areas of interest. This number of participants allowed the collection of a diverse range of perspectives. It may have been possible to speak with the same number of participants in two or three focus groups. This approach would have been less time-consuming, however there was a potential risk of group conformity (Rosenbaum et al., 2002).

7.1.1.1 Materials

The interview guide follows—once data was collected, a thematic analysis (Blaxter, 2010) of the interview data was conducted:

1) What type of mobile phone do you own?
    a. Make
    b. Model
    c. OS
2) Which mobile applications do you use often?
    a. Allow the participant to check their mobile device if needed
3) What types of things do you do with those applications?
4) When and where do you do those things?

7.1.1.2 Participants

Recruiting those that were not knowledgeable in HCI was more problematic than recruiting those that were. Incentivising participants was not permissible according to university guidelines and attempting to schedule interview time with those that were busy, yet offer no payment proved to be quite difficult. With those knowledgeable in HCI, recruiting proved to be easier as there was an intrinsic motivation to help create mobile application usability heuristics and a contextual usability evaluation protocol. It was decided to use this intrinsic motivation to recruit four participants with HCI experience as part of a pilot study to ensure that none of the questions within this study were ambiguous or difficult to answer. The pilot study was conducted between January 16th, 2018 to March 10th, 2018 with four females, two of whom were between 25-34 years old, one was between 45-54 years old, and one was between 55-64 years old (ethics protocol number: aCOM/PGR/UH/03021). All questions were understood by participants; therefore, no changes were made to the questions for the main study.

For the main study, to work around the issue regarding the lack of incentive, participants were selected at random from Tripadvisor's active directory with permission from the Director of User Experience Research. Thus, a purposive sample of those that had a very low probability of having HCI knowledge were recruited from departments, such as business development, marketing, finance, customer relationship management, search engine optimization and so on. Several participants were known to the principal researcher at a superficial level, most were not known at all. An email was sent out to each person on the list, letting each potential participant know that this was an academic study, thus not work-related, whereby I wished to speak with them about the mobile device they owned, the common day-to-day tasks they completed on their mobile device, and when/where they tended to complete those tasks. All participants accepted and interviews were conducted between June 13th, 2018 and August 14th, 2018 (ethics protocol number: aCOM/PGR/UH/03021(1)). Of the thirty participants, nine were male and twenty-one were female, while twenty-four used iPhones and six used Android. The majority of participants were between 25-34 years old (Table 7.1):

Table 7.1. Age range of participants (n=30)

Age in years	Number of participants
18-24 years	5
25-34 years	19
35-44 years	5
45-54 years	1

7.1.1.3 Research Quality

As qualitative research methods were used, research quality focused on credibility, transferability, dependability, and confirmability. Credibility was established by prolonged engagement with thirty participants that regularly used mobile devices. Transferability was established by purposive sampling of a representative set of participants and sufficiently detailing the context of the study. From this, readers can establish the applicability of findings with other contexts. Transparency was linked to dependability, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. Confirmability was established by clearly and accurately presenting participants' perspectives, whereby transcriptions were revisited as many times as necessary to ensure that themes remained true to participants' accounts. Finally, the threat of artificiality, in particular experimenter expectancy (Maruyama & Ryan, 2014), was mitigated by remaining friendly, yet neutral at all times during the study.

7.1.2 *Results and Discussion*

Participants primarily spoke about six contextual areas of focus within which tasks on mobile applications were attempted. These contextual focus points were when they were at home before work or during days off work, traveling to/from work, at work, arriving back home from work, outside locally and on holiday:

7.1.2.1 At Home Before Work or During Days Off Work (n=30)

When participants were at home, they tended to use mobile devices throughout the day. Many participants used the mobile device as their alarm. After the phone alarm woke them, participants often stayed in bed and checked a number of applications as they began to wake up—these included sending and reading texts on iMessage or the default text message software, messages on Slack and HipChat (refer to the glossary in Appendix A for more details regarding company and product names), emails on Gmail and Outlook, checking weather for the day on Weather.com, Weather Underground, and Yahoo weather, reading the news on Chrome, Safari, and DuckDuckGo browsers or news mobile applications, reviewing their calendar, and looking through their social media accounts to see what they missed as they slept. As the day continued, much of the daily use of mobile applications was deemed 'mindless scrolling' on social media.

7.1.2.2 Traveling to/from Work (n=30)

When traveling to/from work, participants often ordered coffee using the Starbucks mobile application, used Google Maps or Waze to navigate and learn about traffic congestion, as well as Spotify to listen to music. While driving, participants limited their mobile device usage knowing the dangers involved. However, the times that they needed to use their mobile devices were often difficult. This included trying to change songs or podcasts while driving, and while using navigation applications should they take a wrong turn or needed to change a destination. Several participants were starting to use voice commands, an imperfect technology, to assist with these tasks.

If participants used other means to commute, such as a bus or light rail, they tended to use other applications, such as reading articles on Medium or Reddit, reading books on Kindle, reading the news on NY Times, CNN, AP, Espresso, BBC, The Economist, Business insider, Washington Post, and Boston.com, text messaging, listening to podcasts or books on Audible, listening to music on Spotify or Pandora, checking social media, including Pinterest, Instagram, Twitter, Facebook and LinkedIn. In these situations, some participants would commence work-related activities, such as reading and answering emails. However, this was difficult for other participants as they often experienced motion sickness when trying to read as the light rail car or bus moved around. Should this be an issue, participants tended to listen to music and/or podcasts. A participant that did not experience motion sickness found it difficult to type messages or emails on a tiny keyboard on a moving bus. She reduced the number of errors by using a pop-socket, which she attached to her mobile device. Other participants mentioned that trying to use a mobile device with one hand can be annoying. An example of this was while holding a handrail on a light rail train with one hand and attempting to type and send a message on a tiny keyboard with your other hand. While not limited to underground train travel, dead zones were a major annoyance for several participants.

7.1.2.3 At Work (n=30)

When at work, some participants made a concentrated effort to avoid continued 'mindless scrolling' on social media. This often meant putting the phone out of arms reach and turning the device face down so as not to be distracted. Participants would use mobile applications, such as email and calendar when they were walking between meetings and otherwise away from their desks. Deviations from this practice were periodic, such as when the participant was in the restroom/toilet or every few hours at their desk, where they tended to just check for and respond to personal messages. Further, participants often had to change their notification preferences as many applications sent constant notifications of non-important events to get a user's attention, which was often irritating.

7.1.2.4 Arriving Back Home from Work (n=30)

When participants arrived home from work, they often actively avoided work-related mobile applications, which differed to their morning routine. Instead, they would catch up on personal messages, social media, any news they might have missed, and the weather for the next day, so they could think about which clothes to wear. Several participants referred to this as 'zoning-out'.

Participants used their mobile devices to assist with dinner, some to find recipes for dinner using apps such as GoKitchen, Hello Fresh, and the NY Times Cooking app, some to order food using GrubHub and Seemless, while others reserved a table at local restaurants using OpenTable. One of main issues raised by participants was when they were cooking and following a recipe on a mobile application, only for that application or the mobile device to automatically log them out for 'non-usage' when their hands are 'icky'. Following their evening meal and subsequent clean-up, many participants checked their phones once more for any notifications they might have missed.

As participants settled down for the evening and watched TV, many used mobile applications, such as Netflix, Xfinity, IMDB, ESPN, and Amazon Prime Video, generally to cast to a larger smart TV. Should a TV show have adverts/commercials, many participants checked social media once more, to which they often referred to once again as ‘mindless scrolling’ that ‘added no value’.

Should participants need to shop or book travel, they tended to use mobile applications, such as Amazon, Macy’s, Target, CVS, Wayfair, Myntra for shopping and Tripadvisor, United Airlines, Kayak, and Yelp for travel. In some cases, participants switched to the internet on a laptop for these activities to better view photos of products or places. During the evening, finances, paying bills, checking stock prices etc. tended to be conducted using a series of mobile applications, including Venmo, Chase, Bank of America, Venmo, Saffire, Mint, Citizens Bank, DCU, MyBluebird, Avidia, Capital One, Splitwise, Apple Stocks application, and Yahoo finance. Other applications that were inclined to be used at this time were dating apps, such as Bumble and Hing.

7.1.2.5 Outside Locally (n=30)

When participants were outside their home or work, but within their neighbourhood or city, they used different mobile applications depending on the situation. For example, when participants were walking or running, they often used health-related applications, such as Runkeeper, Nike run club, Rappa, FitBit, and MapMyRun, while using Spotify to listen to music. Once actively walking or running, participants found it difficult to start a new mobile application or change a song etc. unless they stopped moving. Not only did stopping make it easier to complete a task, participants were also more aware of their environment, which reduced the likelihood of having an accident. Participants also used mobile devices outside during inclement weather, such as in the rain or snow. Such weather conditions hampered the usage of mobile devices, even with touchpad-friendly gloves. Additionally, inclement weather was considered potentially damaging to a mobile device. Further, participants were hesitant to use their mobile device at night in urban areas—this had the potential to lead to an unsafe situation due to the lack of awareness of their immediate surroundings.

Another example of being outside the home or work was attending social events. When going out with family, friends, or colleagues, participants often made an effort to not use their mobile devices as they felt that it would be inappropriate. This was often enforced with rules, such as 'the first person to pick up their phone buys a round of drinks for everyone'. If participants needed to use their mobile devices, they would do so quickly. Examples include organizing transportation using Uber or Lyft, splitting bills using Venmo, using Snapchat or text message to tell others where they were, using their camera with applications, such as Instagram, to take a photo of the meal and/or social event, as well as Google or Guinness book of records to find the answer to a question someone posed to them or to the group. When attending social events, surrounding environments can often be noisy. In such cases, should participants need privacy, they had to create workarounds that ensured they could hear a caller, but no one else nearby could hear the conversation, such as cupping their hand over the microphone on the mobile device.

7.1.2.6 On Holiday (n=30)

Participants also spoke about using mobile applications while on holiday, whereby they often used their mobile devices for photography, to find things to do and places to eat, find directions to/from points of interest, and to post to social media. Some mobile applications that participants used regularly were used differently while on holiday. One example was Instagram, which was usually used to share photos, whereas on holiday, it was also used for inspiration to find interesting places to go nearby based on hashtags. Several participants mentioned that trying to use a mobile device in the bright sun, such as at the beach, was close to impossible. In such situations, mobile devices also tended to overheat, making them difficult to hold or touch. The bright sun also caused other issues, such as trying to take a photo in landscape mode with polarized sunglasses, whereby it was impossible to see the mobile device screen. In bright conditions, it was also more difficult to show something on the mobile device screen to someone as they could not see the item on the screen from an angle.

When traveling outside of the United States, many participants did not want to pay extra for an International data plan, so their mobile phone usage was often limited. Some participants embraced this by taking more notice of their destination and to cherish being away from their mobile device. Other participants, or their friends, did not embrace the inability to use their mobile devices as they normally could and would regularly seek out a free Wi-Fi signal. The latter was often difficult to find as many Wi-Fi providers wanted to collect an email address before Wi-Fi could be used. Several participants contemplated the lack of a data plan and free Wi-Fi in advance of their travel dates by saving screenshots of directions from their vacation rental or hotel to nearby points of interest and restaurants, or bringing paper maps, as well as printing itineraries before they left for their destination. Should participants have an International data plan, they tended to use travel-related mobile applications more often, such as airline applications, baggage trackers, Yelp, Foursquare, and Tripadvisor. In such cases, it could sometimes be difficult to use applications that did not have well-designed in-built capabilities. As such, participants needed to switch between applications to complete a single task. Examples included using a travel application to find a place to eat, then switching to Google Maps to see how far away it was while trying to remember the name and address of the restaurant. Even if participants paid more for an International data plan, some mobile phone service providers throttled data resulting in slower speeds when participants exceeded a certain limit, even on unlimited data plans. This made mobile application usage while outside of participants' home country extremely frustrating.

7.1.2.7 Considering the Robustness of the Contextual Usability Evaluation Protocol

Following the interviews, it was important to ensure that the protocol defined as part of this work allowed for the real-world tasks and contexts of use. Below are ten examples of contextual usability evaluation user stories that originated from the interviews. While more might have been defined from the data, a diverse list of thirty-four contextual usability evaluation user stories that originated from the interviews can be found in Appendix I:

1) When writing an iMessage, does walking on a busy street, have an impact on the perception that the mobile application is tolerant of errors?
2) When reading messages quickly on WhatsApp, does being with friends or family at a social event, have an impact on the perception that the mobile application feels responsive?
3) When writing an email and attaching a photo, does walking from one meeting to another at work, have an impact on the task being completed quickly and easily?
4) When sending a Slack message with a video attachment, does keeping an eye out for your shuttle bus, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?

5) When reviewing upcoming events on Fantastical Calendar, does walking from one meeting to another at work, have an impact on readability of all screen elements in portrait mode?
6) When checking the weather for the day on Yahoo weather, does being sleepy having just woke up, have an impact on perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?
7) When quickly reading the latest posts on LinkedIn, does attempting to meet an upcoming deadline at work, have an impact on readability?
8) When looking for inspiring things to do nearby on Instagram, does being in an unfamiliar location while on holiday, have an impact on the perception that the mobile application is personalized to suit the user?
9) When posting to Twitter, does standing on a busy underground train while holding an overhead rail, have an impact on the perception that the mobile application is tolerant of errors?
10) When requesting an Uber ride, does keeping an eye on your surroundings while on a dark street in the city, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?

7.1.3 *Summary*

This section focused on ensuring the contextual usability evaluation protocol could handle diverse user stories created from real-world tasks and contexts of use. To do this, it was decided to learn about the types of mobile applications that people use, the types of tasks they attempt using these mobile applications, and the types of contexts they find themselves in when using their mobile device. The method selected was thirty semi-structured interviews of those not knowledgeable in HCI. From the analysis, it was clear that participants primarily spoke about six contextual areas of focus within which tasks on mobile applications were attempted. Sample contextual usability evaluation user stories were created based on the findings to illustrate how the protocol was robust enough to consider real-world contexts. The next section focuses on gathering the attitudes of those knowledgeable towards the contextual usability evaluation protocol.

7.2 What is the Attitude of those Knowledgeable in HCI to the Contextual Usability Evaluation Method?

With the protocol defined, it was important to gather the attitudes of those knowledgeable in HCI towards the protocol. Only then was it possible to improve the protocol, if needed, and to have a better understanding if the protocol would be used by those knowledgeable in HCI or discounted as unimportant or infeasible.

7.2.1 *Methods*

The methods chosen for this study were semi-structured interviews, a demonstration of the protocol, followed by an in-person questionnaire. This approach allowed the opportunity to probe on areas of interest. The number of interviews decided upon was twenty, which facilitated the capture of diverse viewpoints that would help to improve the protocol. The demonstration of the protocol included how an evaluation might be carried out as an example (Appendix J). This example evaluation, which included a dot plot, was hypothetical and had not been conducted, it was included only to offer a practical demonstration to participants as to how the protocol might be used by HCI researchers and practitioners.

7.2.1.1 Participants

There were two potential problems when recruiting those knowledgeable in HCI for this study. Firstly, university regulations did not permit incentivisation of participants. Secondly, given that both the principal researcher and each participant would be working full-time, interviews would need to be conducted before 8am and after 7pm, which further complicated recruitment efforts. To address these issues, a representative purposive sample was selected at random from Tripadvisor's active directory with permission from the Director of User Experience Research. Ten of those selected had 'design' in their job title and ten had 'research'. Once selected, a potential participant was sent an 'invitation to participate' email. This approach increased the probability that participants had HCI knowledge, while also ensuring that there would be a good balance of designers and researchers across the participant pool. Tripadvisor did not have a single HCI team; designers and researchers were split across many products and business units within the organization. Therefore, while some participants were known to the principal researcher at a superficial level, most were not known. The study was conducted between November 11th, 2019 and December 16th, 2019 (ethics protocol number: ECS/PGR/UH/03864).

In addition to separating the participant pool evenly between HCI designers and researchers, participants also had varying degrees of seniority and experience levels (8 male, 12 female; Years' experience in HCI: Range=3-25; Mean= 9.3; SD=5.58; Years' experience in mobile HCI: Range=1-14; Mean=4.25; SD=3.99) (Table 7.2):

Table 7.2. HCI design and research participants (n=20)

Participant	**Gender**	**Job role**	**Years' experience (HCI/Mobile HCI)**
P1	Female	UX Researcher	3/1
P2	Male	Lead UX Researcher	11/1
P3	Female	Senior UX Researcher	11/2
P4	Female	Senior UX Researcher	4/1
P5	Female	Senior UX/UI Designer	14/14
P6	Male	Senior UX/UI Designer	11/8
P7	Female	Principal UX Designer	11/9
P8	Male	Lead UX Researcher	10/1
P9	Female	Sr. Manager, UX Research	25/2
P10	Female	UX Researcher	4/1
P11	Female	Dir. Product Design & UX	10/6
P12	Male	Dir. Product Design & UX	15/10
P13	Male	Lead UX Researcher	5/2
P14	Male	Senior UX Designer	15/6
P15	Female	UX/UI Designer	4/2
P16	Female	UX Researcher	4/1
P17	Male	Assoc. Dir. Product Design	10/10
P18	Female	UX Designer	3/1
P19	Female	Senior UX Researcher	4/2
P20	Male	Senior UX Researcher	12/5

7.2.1.2 Materials and Procedure

The email sent to participants was concise and laid out the motivation for the study, that participation was voluntary, participants were informed how long the interview would take, and that the interview would be broken up into three phases. Participants were informed the first phase involved an interview with a focus on how the participant considered the impact of context of use on mobile application usability, if at all. A demonstration of the contextual usability evaluation protocol followed, after which an evaluation of the protocol using a short questionnaire would be conducted. All interviews were recorded with participants' permission using a high-quality audio recorder. No problems were noted with audio quality. The interviews were transcribed in Microsoft Word. The interview guide follows:

1) What is your job role?
2) How many years' experience do you have in UX?
3) How many years' experience do you have in mobile UX?
4) In relation to mobile applications, what does context of use mean to you?
5) When researching for and/or designing mobile applications in your current role or in a previous role, how have you considered the potential impact of context of use, if at all?

Following a demonstration of the contextual usability evaluation protocol using Microsoft PowerPoint, a questionnaire was posed to participants:

- I felt that contextual usability evaluation would be able to consider the impact on mobile application usability within most contexts
- I felt that contextual usability evaluation was easy to use
- I felt that contextual usability evaluation was easy to understand
- I felt that contextual usability evaluation was easy to learn
- How might contextual usability evaluation, including the heuristic set, be improved, if at all?

All questions, apart from the last, used Likert scale choices between 1 - Fully Agree to 5 - Fully Disagree. The final question was addressed using free text. A thematic analysis of the interview data was conducted (Blaxter, 2010).

7.2.1.3 Research Quality

As qualitative research methods were used, research quality focused on credibility, transferability, dependability, and confirmability. Credibility was established by prolonged engagement with twenty participants knowledgeable in HCI. Transferability was established by purposive sampling of a representative set of experienced HCI practitioners and sufficiently detailing the context of the study. From this, readers can establish the applicability of findings with other contexts. Transparency was linked with dependability, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. Confirmability was established by clearly and accurately presenting participants' perspectives, whereby transcriptions were revisited as many times as necessary to ensure that themes remained true to participants' accounts. Finally, the threat of artificiality, in particular experimenter expectancy (Maruyama & Ryan, 2014), was mitigated by remaining friendly, yet neutral at all times during the study.

7.2.2 *Results and Discussion*

7.2.2.1 Interview Results

In relation to mobile application user experience, the term 'context of use' meant different things to different participants. This is no different to the literature, whereby context of use has many definitions. Examples were:

- *"The place and the situation that the user is using the app in", P3*
- *"The time, space, environment that the user is engaging with the app", P10*

- *“What is the user’s current point of view when looking at an app, what are they trying to get done”, P14*
- *“The greater environment in which the device is being used”, P16*
- *“Time and place they will use the mobile app”, P20*

While the vast majority of participants had previously stated that they had to evaluate the usability of mobile applications in their current or previous roles, it became clear that many participants had difficulties with this task. This supported the findings from a previous study within this programme of research in that effective methods used to consider the impact of context of use were limited. For instance, P2 mentioned that they could not do quantitative research within a true context, so they had to fall back on a questionnaire, which is problematic. P3 on the other hand stated that it was difficult to emulate the real world, so the impact of context of use was often overlooked, which undoubtedly impacted usability. P9 stated that there was no method to consider the impact of context of use effectively, so once again the topic was overlooked. P10 and her team made assumptions in a previous role, whereby they created storyboards from their own limited experiences of the impact of context of use without speaking with any users. P16 and P19 also stated that their teams would discount the impact of context of use on mobile application usability given the challenges involved. Knowing that there were few, if any, options available in understanding the impact of context of use on mobile application usability, P20 attempted to solve the issue by using GPS sensor data. That approach was not successful, so the project they were working on was subsequently cancelled.

7.2.2.2 Questionnaire Results

Following a demonstration of the contextual usability evaluation protocol, participants answered a questionnaire. As the questionnaire was conducted in-person, the principal researcher was able to probe on interesting answers from participants. The first question that participants were asked in the questionnaire was in regard to their perception around the ease of use of the protocol. The majority of participants agreed somewhat that the protocol would be easy to use (Figure 7.1). There were five main concerns from participants in regard to ease of use. P3 was most concerned with wording the contextual usability evaluation user story. This is understandable as there were three distinct parts to the story, and it was a new format that had not been encountered before. P5, on the other hand, felt that the open-ended questions asked of those participating in an actual contextual usability evaluation may be difficult to answer in some contexts. P9, P11, and P17 mentioned that it might be difficult to place thirty participants within various contexts should statistical significance be important. P13, P15, P16, and P18 all brought up the fact that the real world is complex—there might be a lot of different contexts to consider and they would not know which to prioritize. Finally, P7 and P20 were concerned with the learning curve associated with using a new approach.

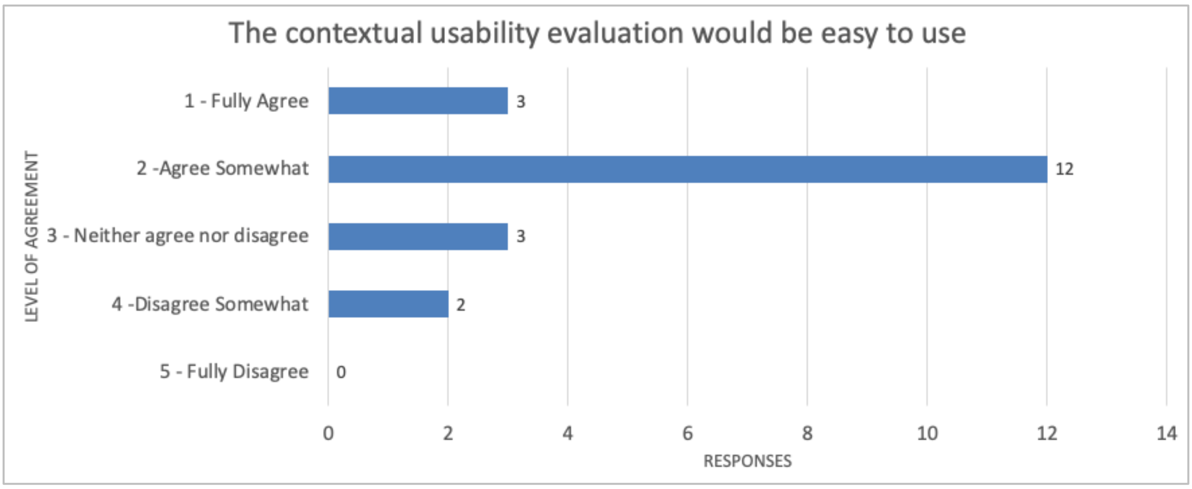


Figure 7.1. Context usability evaluation would be easy to use (n=20)

The next question focused on ease of understanding, whereby the majority of participants fully agreed or agreed somewhat (Figure 7.2). Four areas of concern were raised when participants were probed on their answers. P4 liked the one sentence Agile-like format and wished to see more examples of contextual usability evaluation user stories, including more complicated scenarios. P5, P16 and P17 felt that the example dot plot (Appendix J) was difficult to understand. While P19 disagreed somewhat that the contextual usability evaluation was easy to understand, the participant felt that the scale was straightforward, whereby the participant mentioned “having one metric is great” and “the format and structure makes a lot of sense”. The point of contention from P19 was to ensure that the baseline was the same for everyone as a contextual usability evaluation was being conducted.

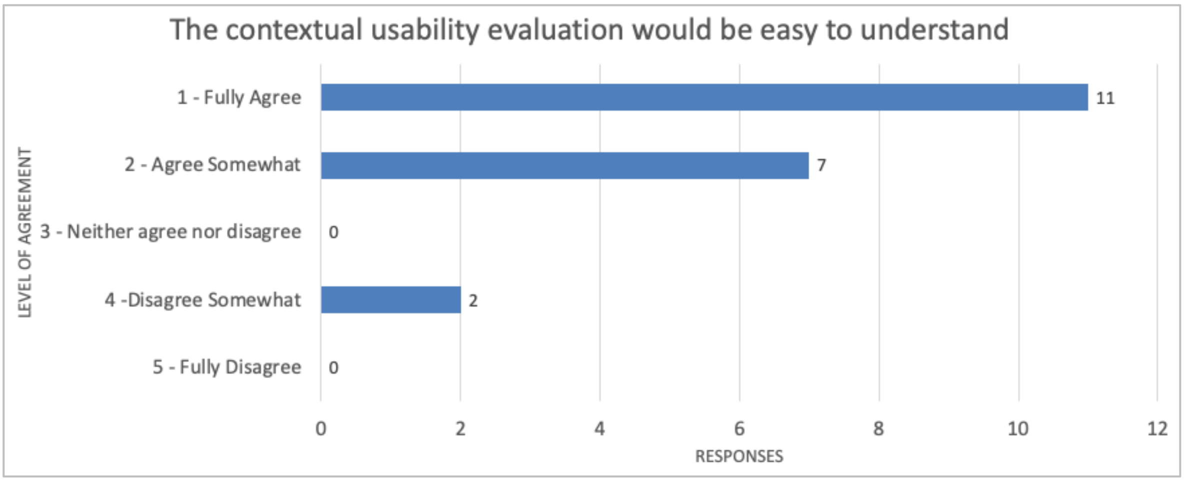


Figure 7.2. Context usability evaluation would be easy to understand (n=20)

The majority of participants also fully agreed or agreed somewhat when asked if the contextual usability evaluation protocol would be easy to learn (Figure 7.3). The concerns raised were from P7, P9, and P16. P7 again brought up the learning curve in having to learn something new. P9 felt that it would take practice in better understanding which contexts to consider, while P16 needed a refresher on statistics. Conversely, P2 mentioned “I can use it right now as it is easy to use and understand, even without the slides I can explain this to other researchers”.

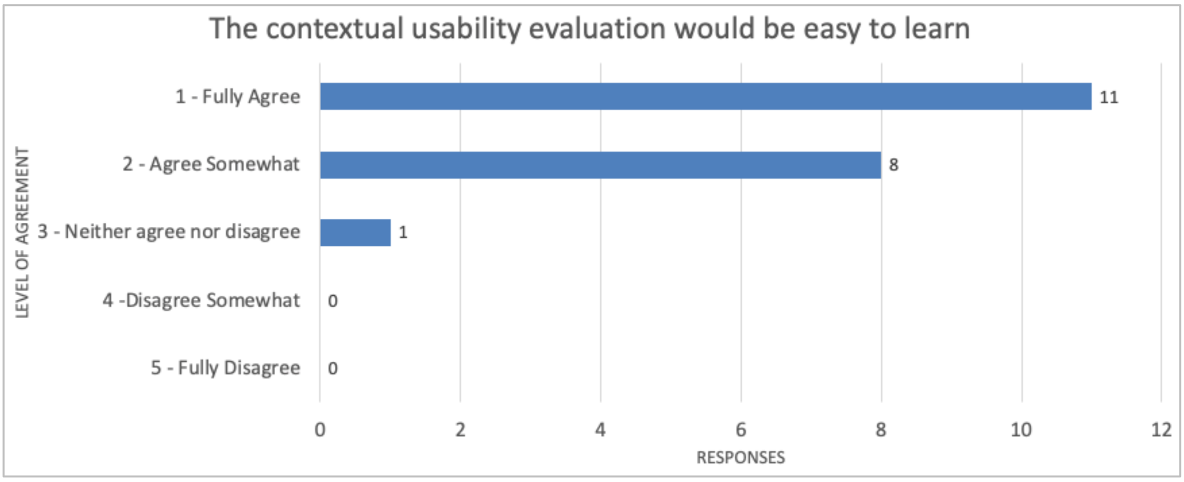


Figure 7.3. Context usability evaluation would be easy to learn (n=20)

The next question asked about the ability of the protocol to uncover differences in the perception of usability of mobile applications within different contexts. Again, the majority of participants answered fully agree or somewhat agree (Figure 7.4). P13 agreed somewhat, and again brought up the issue of prioritization, as well as potentially adding a screener to ensure that the correct target sample was recruited. P18 was more concerned with conducting a contextual usability evaluation, only to find that the results were neutral, whereby they would still be unsure about the changes, if any, that needed to be made to a mobile application design.

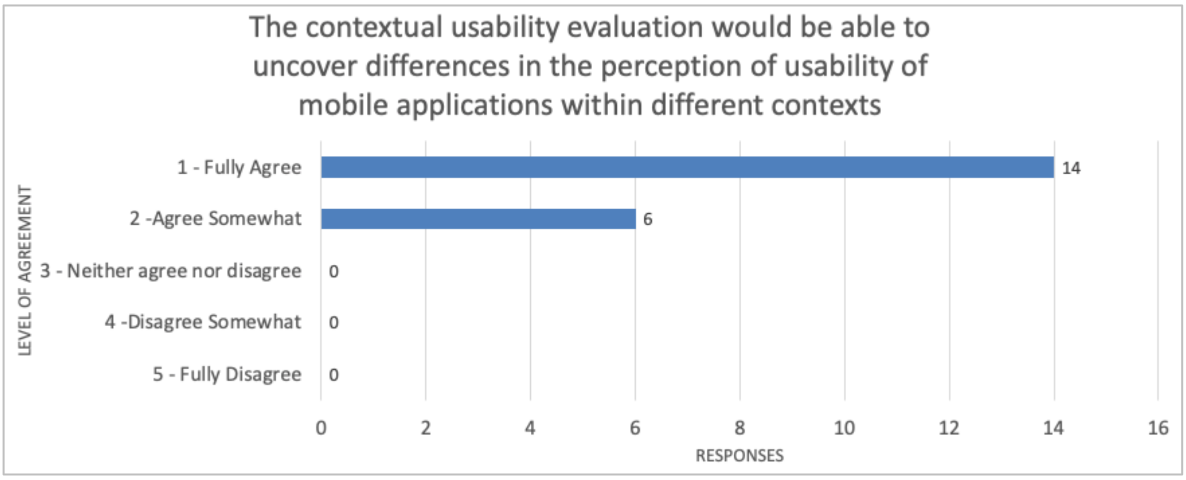


Figure 7.4. Context usability evaluation would be able to uncover differences in the perception of usability of mobile applications within different contexts (n=20)

It was important for the protocol to be easily modified to account for a broad range of contexts of use given the complexity of real-world mobile application use. The next question gathered the attitudes of participants on this issue, all of whom felt that the protocol could be easily modified by fully agreeing or agreeing somewhat (Figure 7.5). Some of the common concerns previously discussed were once again mentioned at this stage. P4, for instance, wondered how well the protocol could adapt to multiple elements, such as "It's dark, I'm tired and I'm on the T [MBTA - Boston's underground subway train system]". P10 was more concerned with how she would get enough users into specific contexts, while P11 felt that there could be a lot of contexts of use to consider.

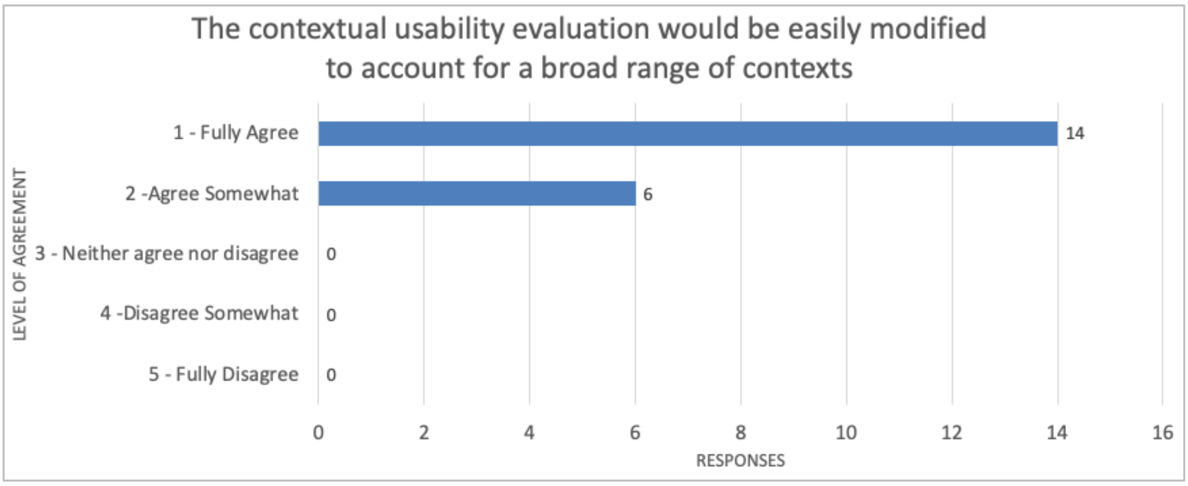


Figure 7.5. Context usability evaluation would be easily modified to account for a broad range of contexts (n=20)

The final question requested that participants consider how the protocol could be improved, if at all. The majority of participants re-iterated previous concerns without offering a better approach. P4 brought up an additional concern, whereby the participant mentioned that the word 'impact' in the question presented to those taking part in a contextual usability evaluation might be leading. A small number of participants considered potential amendments that might address some of the concerns. P10 and P12 mentioned that an additional component, possibly through the use of a participant screener, could take demographic information in account, which would be helpful in ensuring that the sample selected would be representative. P13 mentioned adding a context screener to allow for user story prioritization. For instance, the context screener might list minimum viable conditions for use, without necessarily considering each and every context of use within which the usability of a mobile application might be evaluated. Furthermore, the context screener might also include the degree to which specific context(s) of use were associated with key use cases defined for mobile application usage. In such a case, the mobile application would only be evaluated within certain contexts of use that have a high probability of occurring. That probability of occurrence might be defined by an HCI team following an ethnographic research session(s). P5 suggested using a set of canned responses or 'mad libs' in a dropdown to allow participants within a study to more easily answer the open-ended question within the protocol.

In response to the potential issue of recruiting enough participants and placing them into the disparate contexts, P9 did not directly address the issue, yet spoke about having to recruit less often. In such a scenario, an HCI team may need to wait until a fully interactive prototype or even the first version of the mobile application was ready before conducting a contextual usability evaluation. This would lend the protocol to a summative evaluation more so than a formative evaluation. Some of the problems raised by participants of a contextual usability evaluation could then be addressed in the mobile application if their level of effort is low, while more complex issues might be addressed for the next release. P9 also suggested that crowdsourcing might be an option that could increase the chances of recruiting enough participants for a study. Finally, P16 considered alternative ways to visualize the output from a contextual usability evaluation study for easy comparison, to that end moving away from the example dot plot and instead using different bubble sizes, call outs of critical information, and using the ubiquitous colours green-yellow-red. P16 also mentioned that consideration should be taken in how much data needs to be shown to reduce cognitive load for the viewer; for instance, thirty data points might be averaged to one data point. All of the points raised were addressed in the next section.

7.2.2.3 Updated Mobile Application Heuristics

While contextual usability evaluation user stories were considered straightforward, one participant felt that the wording of each story might take some practice to get right. Arguably, the first two parts of a contextual usability evaluation user story was not difficult to comprehend, namely the task to complete and the context(s) of use within which a task should be attempted. Conversely, the latter part of a contextual usability evaluation user story was less straightforward. To ensure that the creation of contextual usability evaluation user stories was easier, story endings were added to the end of each heuristic within square brackets:

- *M1 Interaction:* Ensure that tasks can be completed quickly and easily on mobile apps by focusing on specified user goals, minimizing data input, using device capabilities, smart defaults, appropriately sized tap targets, and offering clear affordances [...have an impact on the task being completed quickly and easily?]
- *M2 Micro-usage:* Ensure that the mobile application is designed for micro-usage as the user might be frequently distracted [...have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?]
- *M3 Readability:* Ensure that all elements, including graphics and text, on small mobile application screens are big enough to be readable in portrait

and/or landscape modes [...have an impact on the readability of all screen elements in landscape and/or portrait modes?]

- *M4 Simplicity:* Ensure that the elements, including graphics and text, on each mobile application screen are only those required to achieve a specified goal [...have an impact on the perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?]
- *M5 Consistency:* Ensure that elements, including graphics and text, used on each mobile application screen are consistent across the application, while conforming to platform and industry conventions familiar to the user [...have an impact on the perception that all elements, including graphics and text, are consistent and familiar throughout the mobile application?]
- *M6 Errors:* Ensure that the mobile application is tolerant of errors, allowing undo and redo, and using real-time validation of form field input [...have an impact on the perception that the mobile application is tolerant of errors?]
- *M7 Responsive:* Ensure that interactions with the mobile application are acknowledged instantly, even if an operation may take time to complete [...have an impact on the perception that the mobile application feels responsive?]
- *M8 Learnability:* Ensure that the mobile application is easy to learn, using tutorials or onboarding if necessary. Tutorials should only focus on critical areas and have a quick way to exit [...have an impact on the perception that the mobile application is easy to learn?]

- *M9 Personalization:* Ensure that the mobile application can be personalized, customized, and/or configured to suit the user [...have an impact on the perception that the mobile application can be modified, such as personalized, customized or configured, to suit the user?]
- *M10 Continuity:* Where it makes sense to do so, ensure that the mobile application is part of a continuous experience across other device types, such as desktop and TV, thereby allowing the user to continue a task on the mobile application that was started on another device, or continue a task on another device that was started on the mobile application [...have an impact on the ability to continue a task on the mobile application that was started on another device, or continue a task on another device that was started on the mobile app?]

7.2.2.4 Updated Contextual Usability Evaluation Protocol

Prior to discussing the updates to the protocol, it was important to note a change that was not made, namely the addition of participant or context screeners. These are outside of the scope of a contextual usability evaluation, which is conducted after the customer types have been defined, their needs have been understood, their context(s) of use have been mapped, and once a mobile solution is already in place, be it a prototype or a fully developed solution.

Regarding changes that were made, there was a concern that the word 'impact' within a contextual usability evaluation question might be leading. One way to rectify this was to change the unit of analysis from impact to task success or failure. To that end, participants would log if they were successful in completing a task (1) or if they failed to complete a task (0) during a contextual usability evaluation. Such a change was beneficial in other ways. In terms of visualisation of results, the use of a dichotomous task success/failure model could be visualized on a bar chart with green for task success and red for task failure, and/or using patterns should colour-blindness be a concern. The use of a task success/failure model also allowed changes to be made to the statistical analysis method, moving away from a Friedman test to Cochran's Q (Cochran, 1950) with follow-up McNemar tests (McNemar, 1947). The key question at this stage was how many participants were sufficient for a statistical analysis using Cochran's Q. According to Hintze (1998), a sufficiently large sample size is $n \geq 4$ and $nk \geq 24$, where n is the number of subjects and where the match set responses (nk) are not all 0's or 1's. Consequently, as few as eight study participants marking either task success or task failure for one or more tasks within three contexts meets the $nk \geq 24$ condition.

Another concern raised the potential difficulty for participants when writing notes within some contexts, an example of which might be standing on a moving subway train without having something to hold onto. In such a scenario, quickly tapping a large radio button to rate an experience as task success or failure would be relatively easy, more so than answering an open-ended question. However, should notes be written later by participants, important details might be overlooked. To mitigate this issue, P5 suggested using a set of canned responses or mad libs in a dropdown format. This approach is not new; Microsoft's desirability toolkit, for instance, is a classic approach that allows HCI study participants to select from a range of canned responses (Benedek & Miner, 2002). While an interesting idea, weaknesses in the approach suggested by P5 are that limited feedback would be gathered from study participants as such methods are best suited to qualitative research where a study moderator can probe on the words chosen. Additionally, there is a potential for recency bias, whereby participants select words that are higher on the list (Choi & Pak, 2005). To that end, a preferable approach would be the snippet technique (Brandt et al., 2007). This technique was designed specifically to allow study participants capture just enough information under mobile conditions that do not often allow for fuller written text entries. The approach requests that study participants capture words and other bits of text, audio if they are comfortable doing so, and photos, all of which can be created within a few seconds. As soon as study participants are in a comfortable and safe position to do so, they would revisit the saved snippet(s) to add more information to the study notes. In this scenario, a study moderator will receive the required insights to learn why tasks were more difficult within various contexts of use, while mitigating the risk of details being forgotten by participants due to the fallibility of human memory.

Based on the feedback from participants within this study, the procedure for a contextual usability evaluation study was modified as follows:

1) *Prioritize tasks and contexts of use:* Create a research plan that prioritizes key tasks and contexts of use from research conducted earlier in the mobile application software development life cycle;
2) *Define user stories:* Within the research plan, define contextual usability evaluation user stories that should be evaluated. Contexts of use can be unidimensional or multidimensional;
3) *Recruit participants:* Recruit at least eight representative participants that are not usability experts;
4) *Conduct evaluation:* Request that participants attempt specific tasks within a specified context, whereby the order of contexts of use are randomised;
5) *Task Outcome:* Following an attempt at each task, each participant marks the task as either task success (1) or task failure (0). Notes are written or

snippets are created, which may include reasons for any minor or moderate delays or frustrations even if the task is considered successful;

6) *Next context:* Participants move onto the next context of use and attempt the same specific tasks. This cycle continues until all tasks are attempted within all specified contexts of use;
7) *Display output:* Following a statistical analysis using Cochran's Q and follow-up McNemar tests, differences in the impact of context of use on mobile application usability between contexts of use are determined. Results are displayed in statistical format and on a bar chart using green for task success and red for task failure, and/or using patterns should colour-blindness be a concern.

7.2.3 *Summary*

To enable a successful implementation of the contextual usability evaluation protocol by those knowledgeable in HCI, it was vital to gather the attitudes towards the protocol. The methods chosen for this study were semi-structured interviews with ten HCI designers and ten HCI researchers, a demonstration of the protocol, followed by an in-person questionnaire. The protocol was well received, any concerns raised were addressed, and the protocol was updated. With the mobile application usability heuristics and the contextual usability evaluation protocol defined, the next chapter concludes the programme of research. Additionally, implications to theory and practice, how the mobile application usability heuristics and the contextual usability evaluation protocol can be applied by HCI researchers and practitioners, as well as areas for future research, are considered.

7.3 How might those knowledgeable in HCI determine the scope of a typical contextual usability evaluation?

Within the previous section, the proposed contextual usability evaluation protocol was modified based on the attitudes of those knowledgeable in HCI. As a final step in in evaluating the protocol, participants knowledgeable in HCI determined the scope of a typical contextual usability evaluation in order to provide insights into how to apply the method in practice.

7.3.1 *Method*

The method chosen for this study was a workshop. This approach allowed participants to work through a practical example of contextual usability evaluation planning, considering for instance, the number of users, number of stories, number of elements of context, and other general practicalities.

7.3.1.1 Participants

To recruit a representative purposive sample, an email was sent to eBay's Design Research team requesting nine volunteers for the study. eBay, a global ecommerce marketplace, was the place of employment of the principal researcher at the time of the study. The study was conducted with permission from the Head of Design Research at eBay. Of twenty-three team members, ten members volunteered, from which nine were randomly selected. On the day of the workshop, one of the selected volunteers stated that they were unable to attend due to an ongoing eBay-related research project. The volunteer that had not originally been selected was requested to join. Shortly before the start of the workshop, another volunteer stated that they were unable to attend due to their workload. The remaining eight participants were asked for the following information prior to the workshop: Occupation, location, years' of experience in HCI/UX, and years' of experience in mobile HCI/UX. The workshop was conducted on January 14th, 2021 via Zoom (ethics protocol number: SPECS/PGR/UH/04311). Participants were located in California, New York, and Seattle, and had varying degrees of seniority and experience levels (1 male, 7 female; Years' experience in HCI: Range=2-20; Mean=6.56; SD=5.83; Years' experience in mobile HCI: Range=0-5; Mean=2.43; SD=1.54) (Table 7.3):

Table 7.3. HCI research participants (n=8)

Participant	Gender	Job role	Years' experience (HCI/Mobile HCI)
P1	Female	Design Research Lead	20/5
P2	Female	Design Researcher	6/3
P3	Male	Design Researcher	8/2
P4	Female	Design Researcher	2/1.5
P5	Female	Senior Design Researcher	7/4
P6	Female	Design Researcher	3/2
P7	Female	Design Researcher	3/2
P8	Female	Senior Design Researcher	3.5/0

7.3.1.2 Materials and Procedure

An email sent to potential participants included the motivation for the study, stated that participation was voluntary, informed potential participants how long the study would take, and that the study would be broken up into two phases. Phase one involved the principal researcher sending a one-page sheet to participants demonstrating the contextual usability evaluation method, including the creation of stories and the use of statistics. This sheet was sent several days in advance of the workshop as the time available during the workshop was limited. Phase two was the workshop, which was split into three parts: The first eight minutes was dedicated to a recap of the main points from the sheet that was sent in advance to volunteers. Following which all participants were given the same practical example of a task:

“Imagine your 3-person research team is tasked with conducting a contextual usability evaluation study on the eBay mobile (iOS or Android) app widget. How would you determine the scope of the study, including number of participants, number of contextual stories, elements of context, and general practicalities? Is there anything else that would need to be determined before conducting this research project?”. A widget is larger than an icon on a smartphone and displays key information at a glance. Participants were split into three breakout rooms consisting of randomised groups of participants, thus reducing the risk of group conformity. Breakout rooms consisted of the following participants:

- Breakout room 1: P3, P2, P8
- Breakout room 2: P4, P5, P7
- Breakout room 3: P1, P6

Each group was unable to see or hear other groups during this stage, which lasted for twenty-two minutes. When the breakout time had elapsed, all participants returned to the main room. A spokesperson from each group took up to ten minutes to readout how their group would determine the scope of a contextual usability evaluation. However, all participants were encouraged to discuss their individual points of view should they wish to do so.

The workshop was conducted with the Zoom video conferencing platform as all participants worked remotely due to the COVID-19 pandemic. Notes were transcribed in Microsoft Word following which a thematic analysis was conducted (Blaxter, 2010).

7.3.1.3 Research Quality

As qualitative research methods were used, research quality focused on credibility, transferability, dependability, and confirmability. Credibility was established by prolonged interaction with eight participants knowledgeable in HCI. Transferability was established by purposive sampling of a representative set of experienced HCI practitioners and sufficiently detailing the context of the study. From this, readers can establish the applicability of findings with other contexts. Transparency was linked with dependability, whereby there was a clear description of the research process from the initial outline through the development of the methods and reporting of findings. Confirmability was established by clearly and accurately presenting participants' perspectives, whereby transcriptions were revisited as many times as necessary to ensure that themes remained true to participants' accounts.

7.3.2 *Results and Discussion*

When determining the scope of a typical contextual usability evaluation, P3, the spokesperson for breakout room 1, stated that his or her group would initially consider existing primary and secondary research. Such research would inform the design of a contextual usability evaluation, yet the research may not have been conducted by the researcher running the contextual usability evaluation. Should relevant research not be available, researchers had two main options. The first option would be to conduct generative research, such as ethnographic studies, to inform the design of a contextual usability evaluation. The second option was to work with stakeholders, including product managers and product designers, to guess what the dimensions for the study would be, such as the types of contexts that users of the mobile applications would find themselves in. The latter option was not the best choice as guesses could be incorrect, yet this approach does occur within product teams as has been pointed out earlier in this chapter. Aside from the types of tasks, different types of contexts of use, and elements of usability which make up a standard contextual usability evaluation, participants mentioned that they would also need to consider the following general practicalities:

- Time available for the evaluation
- Types of participants
- Time needed to recruit participants
- Number of participants
- Remote or in-person
- Participant incentivization
- Non-Disclosure Agreement (NDA)
- Business metrics

The time available for the contextual usability evaluation was the "largest forcing function" according to P3, which would normally be agreed upon with key stakeholders. Having more time may allow for more tasks being evaluated within more contexts, if that was needed for a contextual usability evaluation. Even with the same task given to different groups within this study, different participants felt that a contextual usability evaluation would take different amounts of time—breakout room 1 stated that they would like to have a full financial quarter to conduct a contextual usability evaluation, breakout room 2 felt that they only needed about two weeks, while breakout room 3 needed between two to four weeks. It is likely that HCI teams would take less time for contextual usability evaluations over time as they became more familiar with the approach.

Participants within all the breakout rooms mentioned the need to consider the type of participants for a contextual usability evaluation. These needed to be the target users for a proposed or existing mobile application. From an eBay perspective, this meant a focus on Business to Consumer (B2C) sellers, Consumer to Consumer (C2C) sellers, or buyers.

The number of contextual usability evaluation participants was also important. Workshop participants agreed that eight contextual usability evaluation participants were sufficient, which is recommended within the protocol. However, they would recruit more than eight to account for no-shows or mortality—the latter being participants that left a study before it was finished. P3 mentioned that he or she would recruit up to twelve participants, then would adjust as necessary. P4, the spokesperson for breakout room 2, was comfortable recruiting nine participants, accounting for only one no-show.

An interesting outcome of the study was that participants were aware that they could conduct a contextual usability evaluation remotely, yet those in breakout room 1 and 2 stated that they might conduct the studies in-person if they could. When asked why, P3 mentioned that conducting a study in-person may “enrich the data collected during the evaluation”. However, P4 stated that he or she would conduct the study remotely if he or she had to travel outside of the local area, if the study was being conducted during the pandemic, or if he or she felt unsafe, such as a need to conduct a contextual usability evaluation in San Francisco at night.

The next three topics for consideration were other practical areas that HCI/UX practitioners needed to consider, namely ensuring that they had the budget for, and knew how best to pay, for participants’ time. Ensuring that participants signed NDA’s was also important in that company secrets would less likely be shared with competitors. P1, the spokesperson for breakout room 3, revealed that the user-centric metric of a contextual usability evaluation, namely task success and failure, may not be enough. It would also be essential to consider business-centric metrics, such as how much money the organization might make should the project be successful.

During the readout, it became apparent that participants had not fully considered the number of contextual usability evaluation stories they would create. None of the participants in the workshop specialised in B2C sellers on the eBay mobile application, therefore following a prompt from the principal researcher regarding the number of stories, only C2C sellers and buyers were considered. P4 and P5 focused on buyers, in particular an existing Persona used on the eBay Design Research team named Samia, a busy, working mom who buys items on eBay. Samia might conduct two primary tasks with the eBay mobile application widget. First of all, she might glance at the widget to check the status of the auction of an item on which she has placed a bid. Secondly, she might tap on the widget to place a higher bid if her bid is no longer the highest and if the price of the item is still within her budget. Samia may do this in two primary contexts, one of which is where she is at home and relaxing when her children are in bed, and secondly when she is making lunch for her children when there is a lot of background noise. After some discussion, P4 and P5 were most interested in measuring the success/failure of two elements of mobile application usability, namely M3 Readability when glancing and M1 Interaction when tapping. To that end, study participants that represent Samia would be recruited and asked to conduct these tasks within their natural contexts, following which they would mark the success or failure of the tasks they are requested to conduct. As such, four contextual user stories would be created:

- When Samia is glancing at the eBay mobile application widget to check the price and status of the auction does being at home and relaxing when her children were in bed have an impact on the readability of the widget?
- When Samia is tapping on the eBay widget to place a higher bid if her bid is no longer the highest and if the price of the item is still within her budget does being at home and relaxing when her children were in bed have an impact on the task being completed quickly and easily?

- When Samia is glancing at the eBay mobile application widget to check the price and status of the auction does making lunch for her children with a lot of background noise have an impact on the readability of the widget?
- When Samia is tapping on the eBay widget to place a higher bid if her bid is no longer the highest and if the price of the item is still within her budget does making lunch for her children with a lot of background noise have an impact on the task being completed quickly and easily?

Regarding C2C sellers, P3 and P8 anticipated that such a seller might glance at the eBay mobile application widget to obtain the status of a payment. The C2C seller may also glance at the eBay mobile application widget to find that a buyer has left a poor review, which could impact the seller's future sales. Two contexts were referred to, namely relaxing at a barber shop, where many people get their local news in the United States, and during a period of potential high stress, such as seasonal or key retail moments, such as Valentine's day and Christmas. Finally, P3 and P8 were most interested in two elements of usability: M3 Readability and M4 Simplicity. In total, this would equal eight contextual user stories, however these were simplified to four stories. While the elements of usability were combined for simplicity, participants would mark the success or failure of each:

- When a C2C seller glances at the eBay mobile application widget to obtain the status of a payment does relaxing at a barber shop have an impact on the readability of the widget *and* on the perception that only the elements required to achieve a specified goal are on the widget?
- When a C2C seller glances at the eBay mobile application widget to find that a buyer has left a poor review does relaxing at a barber shop have an

impact on the readability of the widget *and* on the perception that only the elements required to achieve a specified goal are on the widget?

- When a C2C seller glances at the eBay mobile application widget to obtain the status of a payment does selling during Christmas have an impact on the readability of the widget *and* on the perception that only the elements required to achieve a specified goal are on the widget?
- When a C2C seller glances at the eBay mobile application widget to find that a buyer has left a poor review does selling during Christmas have an impact on the readability of the widget *and* on the perception that only the elements required to achieve a specified goal are on the widget?

While two elements highlighted by participants may have led to changes to the contextual usability evaluation protocol, namely types of users and business metrics. However, no changes were made as types of users and business metrics were important considerations for eBay, yet that may not be the case for all HCI teams.

7.3.3 *Summary*

This section considered how participants knowledgeable in HCI determined the scope of a typical contextual usability in order to provide insights into how to apply the method in practice. The section followed two other sections, one of which sought to understand if the proposed protocol was able to handle diverse user stories based on real-world contexts of use, and another that gathered the attitudes of those knowledgeable in HCI towards the proposed protocol. In all, these sections helped to enable a successful implementation of the contextual usability evaluation protocol.

The method chosen for this study was a workshop with eight participants knowledgeable in HCI. From the study, it was clear that past research was important as it would inform a contextual usability evaluation. Besides this, the types of tasks, contexts of use, and elements of usability in the form of mobile applications heuristics needed to be decided upon. Once these decisions were made, participants had to contemplate a number of areas as they determined the scope of a contextual usability evaluation. These areas included the time available for the evaluation, the types of participants and the time needed to recruit participants, the number of participants, if the study would be conducted remotely or in-person, how to pay participants, how to protect business secrets, and finally to measure business metrics, not just user-centric metrics. When reflecting on the number of contextual usability evaluation stories that would be created, participants considered C2C sellers and buyers that might use the eBay mobile application widget. For buyers, four contextual usability evaluation stories were defined, whereas for C2C sellers, eight contextual usability evaluation stories were defined, which was simplified to four.

Workshop participants were able to successfully determine the scope of a typical contextual usability evaluation using the protocol and no major concerns or limitations were uncovered.

Chapter 8. CONCLUSION

Usability issues that occur on mobile applications often differ from those that occur on desktop applications. This is no surprise, given that users of traditional desktop applications do not have to contend with small screens, difficult to use input methods, limited network connectivity, rapidly changing environments and other types of factors that mobile application users regularly contend with. Yet, much of the evaluation carried out on mobile applications relies on usability methods that were created for desktop applications. The aim of this programme of research was to fill this gap in knowledge by investigating how a method originally defined for desktop applications, namely heuristic evaluation, might be adapted and extended to uncover the primary usability issues applicable to mobile applications, including the potential impact of changing contexts of use.

Creating a set of mobile application usability heuristics and attempting to address the potential impact of changing contexts of use are not new. Past attempts, however, have not addressed these related issues simultaneously—instead, researchers have treated these related gaps in knowledge as separate issues. Additionally, past research has focused on different areas, such as the ergonomics of the mobile device (Bertini et al., 2006), while simultaneously overlooking key areas. Further, important points of guidance from evaluation theory, such as Stufflebeam's (2000a) CDC and Scriven's (2005) comlist criteria, which underpin this very topic, have rarely been fully considered within past research efforts. The approach taken to adapt and extend heuristic evaluation in order to address the gaps in knowledge was to define four research questions. The first two research questions focused on adapting heuristic evaluation for mobile applications, while the second set of research questions focused on extending heuristic evaluation for mobile applications by addressing the impact on mobile application usability from changing contexts of use. To better address each of the four main research questions, each was split into two sub-research questions, a summary of which can be found in the next section.

8.1 Addressing Research Questions

This section focuses on how each research question was addressed.

8.1.1 *RQ1: How Might a set of Mobile Application Usability Heuristics be Developed?*

The first research question focused on developing a set of mobile application heuristics. This research question was split into two sub-research questions that considered how mobile application usability heuristics were currently being used, which factors should be included in a set of mobile application usability heuristics.

8.1.1.1 SRQ1-1: How are Mobile Application Usability Heuristics Currently Being Used?

An important theme across this programme of research was to better understand and meet the needs of HCI researchers and practitioners. To that end, the first step in developing a set of mobile application usability heuristics was to learn more about the current usage of heuristic evaluation by those knowledgeable in HCI, especially in terms of evaluating the usability of mobile applications. To better understand this usage, thirteen participants knowledgeable in HCI took part in the sixteen interviews, with three participants being interviewed twice, which can be found in section 4.1.

The primary insight from this study was that heuristic evaluation was often perceived as not needed when evaluating the usability of mobile applications. Additionally, heuristic evaluation was perceived as slow to conduct. The main reasons for this were that participants believed they knew about the primary mobile usability issues, which combined with good design practices meant that heuristic evaluation was not needed. A reliance on usability testing of mobile application strengthened the perception that heuristic evaluation was not required. In addition, when heuristic evaluation was conducted, potentially in the interests of saving time and scarce HCI resources, only one or two evaluators tended to take part.

Yet, even though mobile design practices have matured in recent years and usability testing is faster and easier with tools such as usertesting.com, it would be incorrect to assume that there is no need for heuristic evaluation. It is unlikely that HCI researchers and practitioners know it all nor that the narrow focus of usability testing can find all issues. Rather, it is important to recognize that the holistic nature of heuristic evaluation complements existing methods, such as usability testing, and ensures that even more seasoned HCI researchers and practitioners do not overlook usability issues. Heuristic evaluation might also be conducted relatively quickly should HCI researchers and practitioners review various sets of heuristics that have already been defined prior to creating a repository from which they draw upon. Not including the analysis stage, heuristic evaluations conducted within this programme of research took about 35 minutes to complete, which is relatively quick. Finally, regarding the use of one or two evaluators, not three to five as recommended, this could be something that may never change due to the need to release mobile applications quickly. In such cases, the limitations of the approach should be made clear to product teams.

8.1.1.2 SRQ1-2: What Factors Should be Included in a set of Mobile Application Usability Heuristics?

Heuristics are often defined without a formal methodology (Quiñones et al., 2016), which has led to some researchers making questionable decisions. These decisions have been critiqued within the literature review. Consequently, when choosing what factors should be included in a set of mobile application usability heuristics, a methodology defined by Rusu et al. (2011) was utilised for this purpose. Four steps of the six-step process allowed for the gathering of literature related to mobile application usability characteristics, following which all aspects that needed to be considered were collated. Using this list of mobile characteristics, an investigation of Nielsen's (1994) heuristics was initiated to consider how those heuristics may change and where the gaps continue to exist when applied to mobile applications.

The primary finding was that Nielsen's (1994) heuristics were not suitable for the usability evaluation of mobile applications as they did not consider many of the factors gathered from the literature, such as users that were frequently distracted nor the use of sensors to lessen the burden on users. In addition, several of Nielsen's (1994) heuristics were not relevant to today's mobile applications. For example, the last heuristic from Nielsen (1994) requests that evaluators ensure that 'Help and documentation' exists. Conversely, mobile application users expect to see help that is interactive and non-distractive (Bertini et al., 2006). From this review, an initial set of heuristics for mobile applications was defined, which can be found in section 4.2:

- *SMART1: Provide immediate notification of application status* – Ensure the mobile application user is informed of the application status immediately and as long as is necessary.
- *SMART2: Use a theme and consistent terms, as well as conventions and standards familiar to the user* – Use a theme for the mobile application to ensure different screens look alike. Also create a style guide from which words, phrases and concepts familiar to the user will be applied consistently throughout the interface, using a natural and logical order. Use platform conventions and standards that users have come to expect in a mobile application such as the same effects when gestures are used.

- *SMART3: Prevent errors where possible; Assist users should an error occur* – Ensure the mobile application is error-proofed as much as is possible. Should an error occur, let the user know what the error is in a way they will understand, and offer advice in how they might fix the error or otherwise proceed.
- *SMART4: Use a welcome mat for first-time users* – A welcome mat displaying the main features and how to interact with the application allows first-time users to get up-and-running quickly, after which they can explore the mobile application at their leisure.
- *SMART5: Employ a simplistic, focused, glanceable, visually pleasing, intuitive interface* – Main interfaces should be easy to learn whereby next steps are obvious, focused on one task, be simple to the point of only having the absolute necessary elements to complete that task which will allow access to vital information while users are interrupted frequently and are themselves mobile, yet the interface should still be attractive and memorable.
- *SMART6: Design a clear navigable path to task completion* – Users should be able to see right away how they can interact with the application and navigate their way to task completion.
- *SMART7: Allow configuration options and shortcuts* – The mobile application should allow configuration options and shortcuts to the most important information and frequent tasks, including the ability to configure according to contextual needs.
- *SMART8: Cater for diverse mobile environments* – Diverse environments consist of different types of context of use such as poor lighting conditions and high ambient noise are common issues that mobile users have to face every day.

Cater for these potential issues, for example by allowing users to change interface brightness and sound settings.

- *SMART9: Facilitate effortlessness input* – Mobile devices are difficult to use from a content input perspective. Ensure users can input content accurately by displaying keyboard buttons that are as large as possible, as well as allowing multimodal input.
- *SMART10: Make good use of sensors* – Utilize the complex sensors available as much as possible to provide users with a more interesting and stimulating experience.
- *SMART11: Create an aesthetic and identifiable icon* – An icon for a mobile application should be aesthetic and identifiable as this is what a user sees when searching the device interface for the application they wish to launch and when scanning through app stores it will be the first item they see before the application title, description and screenshots.

8.1.2 *RQ2: How Might a Set of Mobile Application Usability Heuristics be Evaluated?*

The second research question focused on evaluating the set of mobile application heuristics defined as part of the first research question. This research question was also split into two sub-research questions. The first sub-research question measured the attitudes of those knowledgeable in HCI towards the mobile application usability heuristics defined within this programme of research. The second sub-research question compared the results of a heuristic evaluation, which was conducted using three sets of heuristics, the mobile application usability heuristics defined within this programme of research, a traditional set from Nielsen (1994), and an alternative set of mobile application usability heuristics defined by Bertini et al. (2006).

8.1.2.1 SRQ2-1: What is the Attitude of those Knowledgeable in HCI to the Mobile Application Usability Heuristics?

Continuing with the methodology defined by Rusu et al. (2011), the fifth step ensured that the set of mobile application heuristics was validated. While Rusu et al. (2011) recommended that a heuristic evaluation against traditional heuristics be conducted, an additional step was included to increase the validity of findings. To that end, a questionnaire was completed by sixty participants from eighteen countries who offered their feedback on the initial set of heuristics. The majority of participants rated the initial set of mobile application heuristics as either useful or very useful, however several heuristics were deemed less useful without modification. For example, heuristic 4 referred to a 'welcome mat' for an onboarding tutorial. While this was a term used in the literature, it was not familiar to all participants. Heuristic 5 was perceived as attempting to cover too much. Heuristic 9, on the other hand, was focused, yet the title regarding 'effortless input' was deemed unobtainable in regard to mobile applications. Several participants also pointed out that sensors and other smartphone functions, including the camera, may not be needed for standard mobile applications. Furthermore, it was clear from the results that several participants did not deem an identifiable icon as important or they felt it would be difficult to evaluate as feedback could essentially come down to individual tastes.

Following the aggregation, analysis and implementation of quantitative and qualitative feedback received during the questionnaire, the next iteration of twelve mobile application heuristics was prepared, which can be found in section 5.1. These included modified terminology, such as a move away from 'welcome mat' to 'overlay', i.e. 'Display an overlay pointing out the main features when appropriate or requested'. Heuristic 5 was split into multiple heuristics, including 'M5: Each interface should focus on one task', 'M6: Design a visually pleasing interface', and 'M7: Intuitive interfaces make for easier user journeys'. The heuristic that included 'effortless' in regard to user input was changed to 'M11: Facilitate easier input', while the heuristic that called on evaluators to ensure that a mobile application made good use of sensors was softened to include the content 'Consider the use of the camera, microphone and sensors to lessen the users' workload'. The full set of modified mobile application usability heuristics follows:

- *M1: Provide immediate notification of application status.* Ensure the mobile application user is informed of the application status immediately and as long as is necessary. Where appropriate do this non-intrusively, such as displaying notifications within the status bar.
- *M2: Use a theme and consistent terms, as well as conventions and standards familiar to the user.* Use a theme for the mobile application to ensure different screens are consistent. Also create a style guide from which words, phrases and concepts familiar to the user will be applied consistently throughout the interface, using a natural and logical order. Use platform conventions and standards that users have come to expect in a mobile application such as the same effects when gestures are used.
- *M3: Prevent problems where possible; Assist users should a problem occur.* Ensure the mobile application is error-proofed as much as is possible. Should a problem occur, let the user know what the problem is in a way they will understand, and offer advice in how they might fix the issue or otherwise proceed. This includes problems with the mobile network connection, whereby the application might work offline until the network connection has been re-established.

- *M4: Display an overlay pointing out the main features when appropriate or requested.* An overlay pointing out the main features and how to interact with the application allows first-time users to get up-and-running quickly, after which they can explore the mobile application at their leisure. This overlay or a form of help system should also be displayed when requested.
- *M5: Each interface should focus on one task.* Being focusing on one task ensures that mobile interfaces are less cluttered and simple to the point of only having the absolute necessary elements onscreen to complete that task. This also allows the interface to be glanceable to users that are interrupted frequently.
- *M6: Design a visually pleasing interface.* Mobile interfaces that are attractive are far more memorable and are therefore used more often. Users are also more forgiving of attractive interfaces.
- *M7: Intuitive interfaces make for easier user journeys.* Mobile interfaces should be easy to learn, whereby next steps are obvious. This allows users to more easily complete their tasks.
- *M8: Design a clear navigable path to task completion.* Users should be able to see right away how they can interact with the application and navigate their way to task completion.
- *M9: Allow configuration options and shortcuts.* Depending on the target user, the mobile application might allow configuration options and shortcuts to the most important information and frequent tasks, including the ability to configure according to contextual needs.
- *M10: Cater for diverse mobile environments.* Diverse environments consist of different types of context of use such as poor lighting conditions

and high ambient noise are common issues that mobile users have to face every day. While the operating system should allow the user to change the interface brightness and sound settings, developers can assist users even more for example by allowing them to display larger buttons and allowing multimodal input and output options.

- *M11: Facilitate easier input.* Mobile devices are difficult to use from a content input perspective. Ensure users can input content more easily and accurately by, for instance displaying keyboard buttons that are as large as possible, as well as allowing multimodal input and by keeping form fields to a minimum.
- *M12: Use the camera, microphone and sensors when appropriate to lessen the user's workload.* Consider the use of the camera, microphone and sensors to lessen the users' workload. For instance, by using GPS so the user knows where they are and how to get where they need to go, or by using optical character recognition (OCR) and the camera to digitally capture the information the user needs to input, or by allowing use of the microphone to input content.

8.1.2.2 SRQ2-2: How Might the Set of Mobile Application Usability Heuristics be Compared to Other Heuristic Sets?

Continuing with the fifth step of the protocol defined by Rusu et al. (2011), namely to conduct a heuristic evaluation using the mobile application heuristics and a set of traditional heuristics, then to compare results. To increase the validity of findings, two steps were added. The first step was to add an alternative set of mobile application heuristics to the heuristic evaluation. The second step was to include an evaluation of heuristics immediately following the heuristic evaluation to gather information on how each heuristic set compared in terms of ease of use, ease of learning, ease of understanding, and appropriateness for the usability evaluation of mobile applications.

Six participants knowledgeable in HCI took part in the study. This was close to the recommended number of evaluators for a heuristic evaluation. The heuristic sets were randomized as to which set was used first. Further, to reduce bias each set of heuristics was referred to by a letter, not the researchers' name(s) nor the year of publication. Participants attempted travel-related tasks within the same Tripadvisor mobile application on the same Android platform. The three sets of heuristics uncovered a total of 145 usability issues, with the mobile application heuristics defined within this programme of research finding the highest number of usability issues in absolute terms, specifically fifty-six issues versus thirty-eight for Bertini et al. (2006) and fifty-one for Nielsen (1994). In addition, the set of heuristics defined within this programme of research uncovered more critical issues and more relevant issues than the heuristics from Bertini et al. (2006) and Nielsen (1994).

Based on the evaluation of heuristics with the same participants, it was clear that the set of mobile application heuristics defined in this programme of research were deemed as easy to use, learn and understand as the other two sets of heuristics. Additionally, participants rated the mobile application heuristics defined in this programme of research as the set they would be most confident in using within a professional context. Participants were also confident that the mobile application heuristics defined in this programme of research would find all of the problems in the mobile application used for the study, were the most applicable in terms of application to a range of mobile devices and screen resolutions, would find most problems in any mobile application, and would be able to capture recent developments in mobile applications. Thus, of the three sets of heuristics used in the study, the mobile application heuristics defined in this programme of research were ranked as the most useful.

That is not to say that no changes were needed. For instance, some participants felt that the number of heuristics should be reduced to ten, yet offer the same level of coverage, which was supported by Sharpe et al. (2007). Further, several heuristics were not effective, including heuristics 7, 8 and 10, the latter which focused on a consideration of context of use. The resulting output from the study was an updated set of ten mobile application usability heuristics and a decision to explore a framework that considers the impact of context of use on the usability of mobile applications, which can be found in section 5.2. Yet, a heuristic regarding context of use, namely 'M2 Micro-usage: Ensure that the mobile application is designed for micro-usage as the user might be frequently distracted' remained as this highlighted the topic of context of use and could still be evaluated even when the heuristics were being used to evaluate a mobile application in a laboratory. Furthermore, a heuristic that focused on the wider technology ecosystem within which mobile devices are part of was added, namely 'M10 Continuity: Where it makes sense to do so, ensure that the mobile application is part of a continuous experience across other device types'. The full set of modified mobile application usability heuristics follows:

- *M1 Interaction:* Ensure that tasks can be completed quickly and easily on mobile apps by focusing on specified user goals, minimizing data input, using device capabilities, smart defaults, appropriately sized tap targets, and offering clear affordances.
- *M2 Micro-usage:* Ensure that the mobile application is designed for micro-usage as the user might be frequently distracted.

- *M3 Readability:* Ensure that all elements, including graphics and text, on small mobile application screens are big enough to be readable in portrait and/or landscape modes.
- *M4 Simplicity:* Ensure that the elements, including graphics and text, on each mobile application screen are only those required to achieve a specified goal.
- *M5 Consistency:* Ensure that elements, including graphics and text, used on each mobile application screen are consistent across the application, while conforming to platform and industry conventions familiar to the user.
- *M6 Errors:* Ensure that the mobile application is tolerant of errors, allowing undo and redo, and using real-time validation of form field input.
- *M7 Responsive:* Ensure that interactions with the mobile application are acknowledged instantly, even if an operation may take time to complete.
- *M8 Learnability:* Ensure that the mobile application is easy to learn, using tutorials or on-boarding if necessary. Tutorials should only focus on critical areas and have a quick way to exit.
- *M9 Personalization:* Ensure that the mobile application can be personalized, customized, and/or configured to suit the user.
- *M10 Continuity:* Where it makes sense to do so, ensure that the mobile application is part of a continuous experience across other device types.

8.1.3 *RQ3: How Might Context of Use be Considered When Evaluating Mobile Application Usability?*

Having defined a set of mobile application usability heuristics that might be used within a laboratory, the next two research questions focused on the impact of context of use. The first research question contemplated how context of use might be considered when evaluating mobile application usability. This research question was also divided into two sub-research questions, the first of which investigated how context of use was considered when evaluating mobile application usability. The second sub-research question focused on the development of a satisfactory protocol that allows for the consideration of context of use when evaluating mobile application usability from the perspective of extending the mobile application heuristics.

8.1.3.1 SRQ3-1: How is Context of Use Currently Considered When Evaluating Mobile Application Usability?

Heuristic evaluation tends to be conducted in usability laboratories, not in real-world conditions (Po et al., 2004). As such, one of the heuristics defined within this programme of research, which focused on context of use, proved to be ineffective. Despite a heuristic regarding context of use remaining in the heuristic set, it was clear that more needed to be done in order to consider the impact of context of use on the usability of mobile applications.

The approach taken to better understand how context of use was considered when evaluating mobile application usability was to conduct a questionnaire with both open-ended and closed questions. A total of 156 participants responded to the questionnaire, of which 149 were deemed complete and valid. Most participants were HCI designers, followed by HCI researchers. The majority of participants believed that the environment and users' activities had an impact on the perception of mobile application usability. Based on 213 responses from many of the 149 participants, other contextual elements that may impact the perception of usability were also gathered, after which they were thematically analysed and categorized. These included cognitive load, cultural differences, regulations and policies, users' level of training, distractions, time of day/night, and mobile applications interacting with other devices and applications. Clearly, there was more to consider than just the environment and user activity alone.

Most participants believed that they always or often considered context of use when designing or evaluating the usability of mobile applications. Yet, the majority of the tools and methods that participants listed, which can be found in section 6.1., were not designed for the effective consideration of the impact context of use when evaluating mobile application usability, such as interviews, focus groups and questionnaires. Many of the approaches were best utilised during the research stage of the SDLC, and not at a final stage when the potential impact of context of use might be better understood. Similar work from Eshet & Bouwman (2014) supported this argument. Consequently, it was clear that a satisfactory protocol allowing for the consideration of context of use when evaluating mobile application usability was required.

8.1.3.2 SRQ3-2: What is a Satisfactory Protocol that Allows for the Consideration of Context of Use When Evaluating Mobile Application Usability?

Understanding the impact of naturally occurring contexts of use in relation to application usability is an inherently complex topic. Yet, some methods embraced by those knowledgeable in HCI have been reduced in complexity, which assist in quickly finding insights within fast-paced Agile SDLC's. In addition to background research and other studies within this programme of research, this led to the belief that a satisfactory protocol that allows for the consideration of context of use when evaluating mobile application usability should reduce complexity, be flexible, make better use of scarce resources, be relatively fast to conduct, and have the ability to offer directional insights, be those practically or statistically significant.

Based on these criteria, a contextual usability evaluation protocol was defined. The protocol extended the mobile application usability heuristics defined as part of this programme of research. The foundation of the protocol was the popular Agile user story format, which was modified to consider three parts of the contextual equation: When <attempting a task>, does <an element of context of use> have an impact on <mobile application usability>? Yet, each part of a contextual usability evaluation user story can be complex. Therefore, three tables with elements of each were defined to guide HCI researchers and practitioners. From these tables, user stories can be created such as: When changing music stations on Spotify, does walking on a busy street have an impact on completing the task quickly and easily?

Furthermore, the protocol included a measurement scale that measured the impact of context of use on mobile application usability by utilising a Likert scale from 0 (no impact) to 7 (severe impact). Finally, the following procedure was defined:

- Step 1 - Create contextual usability evaluation user stories
- Step 2 - Recruit participants
- Step 3 – Conduct contextual usability evaluation
- Step 4 - Create visualization (Dot plot was deemed appropriate)
- Step 5 - Calculate statistics (optional) (a Mann-Whitney U test or a Friedman test was deemed appropriate depending on the number of contexts)

With the contextual usability evaluation protocol defined, it was important to evaluate the approach, which was the focus of the next research question.

8.1.4 *RQ4: How Might a Protocol that Considers the Impact of Context of Use on Mobile Application Usability be Evaluated?*

There were a number of approaches that might have been used to evaluate the contextual usability evaluation protocol. One approach might have been to conduct a study using the protocol. Such an approach would offer a visualization and a statistical result, which would be useful for HCI researchers and practitioners. Yet, the robustness of the protocol in terms of the consideration of real-world contexts of use that mobile application users find themselves in or the concerns of HCI researchers and practitioners, if any, to the adoption of the protocol would be unknown. To that end, a more effective evaluation was three-fold. First of all, an effective evaluation was one that learned more about the actual contexts of use that mobile applications were used in to ensure that the protocol would consider such scenarios. Secondly, an evaluation needed to investigate the attitudes of those knowledgeable in HCI towards the protocol and to resolve concerns, if any. Lastly, it was important to understand how those knowledgeable in HCI might determine the scope of a typical contextual usability evaluation using the proposed protocol.

8.1.4.1 SRQ4-1: How Well Does the Protocol Consider the Actual Contexts of Use that Mobile Applications are Used in?

So far, all of the studies within this programme of research had been conducted with those knowledgeable in HCI. For this study, it was more important to hear from those across a wider mobile application user base, not necessarily those knowledgeable in HCI. That was because the aim of the study was to learn more about the actual contexts of use that mobile applications are used in to ensure that the contextual usability evaluation protocol was robust enough to consider such contexts.

To that end, thirty participants across a wide range of ages that were not knowledgeable in HCI were recruited for semi-structured interviews. Twenty-four participants used iPhones, while six used Android. From these interviews, it became clear which mobile applications that participants used the most, the types of tasks that participants often attempted using the mobile applications, as well as the contexts of use within which participants often found themselves in when using their mobile device. Participants spoke primarily about six distinct contextual areas of focus, namely when they were at home before work or during days off work, traveling to/from work, at work, arriving back home from work, outside locally and on holiday, all of which can be found in section 7.1. Based on this information, thirty-four sample contextual usability evaluation user stories were defined. While diverse in nature, all of the sample user stories created from the interviews fit well into the contextual usability evaluation protocol, such as:

- When writing an iMessage, does walking on a busy street, have an impact on the perception that the mobile application is tolerant of errors?
- When reading messages quickly on WhatsApp, does being with friends or family at a social event, have an impact on the perception that the mobile application feels responsive?
- When writing an email and attaching a photo, does walking from one meeting to another at work, have an impact on the task being completed quickly and easily?

Having learned more about the actual contexts of use that mobile applications were used in to ensure that the contextual usability evaluation protocol could consider such diverse real-world contexts, the next step was to investigate the attitudes of those knowledgeable in HCI towards the protocol.

8.1.4.2 SRQ4-2: What is the Attitude of those Knowledgeable in HCI to the Contextual Usability Evaluation Method?

Twenty participants knowledgeable in HCI were recruited for the study (ten designers and ten researchers). The methods chosen for the study were semi-structured interviews, a demonstration of the contextual usability evaluation protocol, followed by an in-person questionnaire. The findings from the semi-structured interviews supported the results of the questionnaire from SRQ3-1 in that participants believed that there were currently few effective approaches that might consider the impact of context of use when evaluating mobile applications. Participants either made assumptions about the potential impact of context of use, overlooked the issue, or used inappropriate methods to understand the phenomenon. For instance, during one mobile application project, one participant attempted to solve the issue of the impact of context of use on usability by using GPS sensor data. That approach was not successful, so the project they were working on was subsequently cancelled.

Having demonstrated the contextual usability evaluation using a Microsoft PowerPoint presentation, participants answered in-person questionnaire. From this, it was clear that participants considered the contextual usability evaluation protocol relatively easy to use, easy to understand, and easy to learn. Furthermore, participants felt that the protocol should be able to uncover differences in the perception of usability of mobile applications within different contexts and be easily modified to account for a broad range of contexts. That is not to say that participants did not have any concerns or feedback before they might implement the protocol. These concerns included recruiting thirty participants in order to conduct a statistical test, which could take some time. There was also a general lack of understanding of the visualization of results on an example dot plot. Several participants also wished to see additional components, such as a context screener that allowed for user story prioritization. The last part of a contextual usability evaluation user story was also deemed more difficult than the first two parts—HCI researchers and practitioners would need to transcribe heuristics into user story endings, which required some thought. In addition, participants were concerned that those that took part in a contextual usability evaluation might not be able to write detailed notes under some mobile conditions.

The feedback gathered from this study led to several modifications of the contextual usability evaluation protocol, which can be found in section 7.2. For instance, the number of participants required for a statistical test was reduced to eight by using Cochran's Q and post-hoc McNemar tests. In addition to these changes, which allow for fewer participants to be recruited, results could be visualized using green for task success and red for task failure on a bar chart and/or patterns can be used should colour-blindness be a concern. To ensure that the creation of contextual usability evaluation user stories was more straightforward, story endings were added to the end of each heuristic within square brackets (this is the final set of mobile application heuristics, which can also be found in Appendix L):

- *M1 Interaction:* Ensure that tasks can be completed quickly and easily on mobile apps by focusing on specified user goals, minimizing data input, using device capabilities, smart defaults, appropriately sized tap targets, and offering clear affordances [...have an impact on the task being completed quickly and easily?]
- *M2 Micro-usage:* Ensure that the mobile application is designed for micro-usage as the user might be frequently distracted [...have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?]
- *M3 Readability:* Ensure that all elements, including graphics and text, on small mobile application screens are big enough to be readable in portrait and/or landscape modes [...have an impact on the readability of all screen elements in landscape and/or portrait modes?]
- *M4 Simplicity:* Ensure that the elements, including graphics and text, on each mobile application screen are only those required to achieve a specified goal [...have an impact on the perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?]

- *M5 Consistency:* Ensure that elements, including graphics and text, used on each mobile application screen are consistent across the application, while conforming to platform and industry conventions familiar to the user [...have an impact on the perception that all elements, including graphics & text, are consistent and familiar throughout the mobile application?]
- *M6 Errors:* Ensure that the mobile application is tolerant of errors, allowing undo and redo, and using real-time validation of form field input [...have an impact on the perception that the mobile application is tolerant of errors?]
- *M7 Responsive:* Ensure that interactions with the mobile application are acknowledged instantly, even if an operation may take time to complete [...have an impact on the perception that the mobile application feels responsive?]
- *M8 Learnability:* Ensure that the mobile application is easy to learn, using tutorials or onboarding if necessary. Tutorials should only focus on critical areas and have a quick way to exit [...have an impact on the perception that the mobile application is easy to learn?]
- *M9 Personalization:* Ensure that the mobile application can be personalized, customized, and/or configured to suit the user [...have an impact on the perception that the mobile application can be modified, such as personalized, customized or configured, to suit the user?]
- *M10 Continuity:* Where it makes sense to do so, ensure that the mobile application is part of a continuous experience across other device types, such as desktop and TV, thereby allowing the user to continue a task on the mobile application that was started on another device, or continue a task

on another device that was started on the mobile application [...have an impact on the ability to continue a task on the mobile application that was started on another device, or continue a task on another device that was started on the mobile app?]

To enable those that took part in a contextual usability evaluation to write detailed notes, the snippet technique (Brandt et al., 2007) was recommended. This technique was designed specifically to allow study participants capture just enough information under mobile conditions that do not often allow for fuller written text entries. The approach requests that study participants capture words and other bits of text, audio if they are comfortable doing so, and photos, all of which can be done in a few seconds. As soon as study participants are in a safe position to do so, they would revisit the saved snippets to add more information to the study notes. Some changes based on the feedback from the study were not made. For example, a context screener to allow for user story prioritization was deemed as out of scope for a contextual usability evaluation. This context of use prioritisation is conducted at earlier stages in a project, preferably using Maguire's (2001a) context of use framework.

Based on the feedback received, in addition to the statistics and visualization being modified, the procedure for a contextual usability evaluation study was revised as a seven-step process as follows (this is part of the final contextual usability evaluation protocol, which can also be found in Appendix L):

1) *Prioritize tasks and contexts of use:* Create a research plan that prioritizes key tasks and contexts of use from research conducted earlier in the mobile application software development life cycle;
2) *Define user stories:* Within the research plan, define contextual usability evaluation user stories that should be evaluated. Contexts of use can be unidimensional or multidimensional;

3) *Recruit participants:* Recruit at least eight representative participants;

4) *Conduct evaluation:* Request that participants attempt specific tasks within a specified context, whereby the order of contexts of use are randomised;

5) *Task Outcome:* Following an attempt at each task, each participant marks the task as either task success (1) or task failure (0). Notes are written or snippets are created, which may include reasons for any minor or moderate delays or frustrations even if the task is considered successful;

6) *Next context:* Participants move onto the next context of use and attempt the same specific tasks. This cycle continues until all tasks are attempted within all specified contexts of use;

7) *Display output:* Following a statistical analysis using Cochran's Q and follow-up McNemar tests, differences in the impact of context of use on mobile application usability between contexts of use are determined. Results are displayed in statistical format and on a bar chart using green for task success and red for task failure.

8.1.4.3 SRQ4-3: How might those knowledgeable in HCI determine the scope of a typical contextual usability evaluation?

To conclude the protocol evaluation, the scope of a typical contextual usability evaluation was determined to understand how to apply the method in practice. Eight participants knowledgeable in HCI worked through a practical example of contextual usability evaluation planning in a one-hour workshop held remotely over Zoom, where the number of users, number of stories, elements of context, and other general practicalities were considered. Several days prior to the workshop, the principal researcher sent a one-page sheet to participants demonstrating the contextual usability evaluation method, which helped to maximise the limited time available for the workshop. The main points from the sheet were highlighted at the start of the workshop. Participants were given the same task: "Imagine your 3-person research team is tasked with conducting a contextual usability evaluation study on the eBay mobile (iOS or Android) app widget. How would you determine the scope of the study, including number of participants, number of contextual stories, elements of context, and general practicalities? Is there anything else that would need to be determined before conducting this research project?". Participants were given twenty-two minutes to tackle this task within three breakout rooms. Once all participants returned to the main room, a spokesperson from each group took up to ten minutes to readout how their group would determine the scope of a contextual usability evaluation. A discussion regarding the number of contextual usability evaluation stories was also conducted.

Based on the analysis, past research was an important element that had to be considered prior to conducting a contextual usability evaluation as this would inform the study. The types of tasks, contexts of use, and the elements of usability in the form of mobile applications heuristics also needed to be decided upon. Following this, a number of areas had to be contemplated as the scope of a contextual usability evaluation was determined. These areas included the time available for the evaluation, the types of participants, the time needed to recruit participants, the number of participants, if the study would be conducted remotely or in-person, how to pay participants, how to protect business secrets, and how to measure business metrics, not just user-centric metrics. In regard to the number of contextual usability evaluation stories, two types of users were considered, namely C2C sellers and buyers, both of whom use the eBay mobile application. For buyers, four contextual usability evaluation stories were defined, while eight contextual usability evaluation stories were created for C2C sellers, which were simplified to four.

8.2 LIMITATIONS

As with any research effort, there were limitations to how the research questions were addressed. These were largely the result of trade-offs due to time and budget constraints. There were four primary types of limitations within this programme of research, namely location, literature, sample, and scope:

- *Location:* While questionnaires gathered responses from those across the world, the majority of the work conducted during this programme of research was in Massachusetts in the United States. With unlimited time, resources in the form of other researchers, and an unlimited budget, addressing the research questions with a focus on more domestic research across the United States and international research would have been possible;
- *Literature:* Several studies relied on the literature, such as understanding which factors should be included in a set of mobile application usability heuristics. In many cases, however, once an approach has been newly defined or modified from previous research, researchers often conduct a study to show that their newly defined approach finds a statistically significant difference. Yet, this result is not always achieved. To that end, while the research is often invaluable, it might not be published (Amrhein et al., 2019; McShane et al., 2019; Mogil & Macleod, 2017). The limitation in this case is publication bias in that a review of mobile application usability characteristics may have missed research that had not been published;
- *Sample:* The limitation in terms of sample needs to be considered from two perspectives, types of participants and sample size:

- Regarding types of participants, in all cases participants from each sample were recruited from the correct population, were recruited across a range of ages, and where appropriate were recruited across a range of HCI and mobile HCI experience levels. As such, all samples were representative. This does not mean that there were no limitations, especially in terms of selection bias. For instance, during the study that better understood how well the contextual usability evaluation protocol truly considered the actual contexts of use that mobile applications are used in, all participants were white collar professionals who lived in the Boston area that commuted to work, with the majority of participants between 25-34 years old and female. It would be impossible to sample every single type of mobile application user across the world, nor was this within the scope of the study. Yet, this approach did not consider those that are unemployed, were blue collar workers, those that were disadvantaged, those that were older and potentially less technical, and those that lived in other parts of the United States and the world and other factors;
- Regarding sample size, 287 participants took part in seven studies conducted within this programme of research. However, not all sample sizes across all studies were equal, some of which were at or exceeded recommended sample sizes. For instance, while only six participants took part in the heuristic evaluation and evaluation of heuristics, this is close to the recommended number of evaluators for a heuristic evaluation. Furthermore, sixty-three semi-structured interviews were conducted during the programme of research, thirty

with those not knowledgeable in HCI and thirty-three with those that are knowledgeable in HCI. This sample size goes beyond that required within the literature, allowing for a solid understanding of the topics at hand. Conversely, the quantitative studies had lower than recommended sample sizes. While, there is an argument to be made for practical significance of the findings from 60 and 149 valid responses within the two questionnaires, if I had more time to attempt to gather more responses, the number of those responses would ideally be several hundred more in order to apply statistical significance. Regardless of additional attempts, this might continue to prove difficult to obtain. Throughout the programme of research, HCI researchers and practitioners that I spoke with stated how busy they were. In effect, it might be too difficult to achieve statistical significance in the number of sample responses from a difficult to reach population;

- *Scope:* The primary limitation in terms of scope was that the heuristic evaluation and evaluation of heuristics included a limited set of travel-related tasks were conducted on one task-based mobile application (Tripadvisor) on a single type of mobile device (LG G2) running one type of operation system (Android 4.4.2). To that end, the heuristics defined within this program of research are unlikely to be applicable to non-task-based mobile applications, such as mobile games. That said, the heuristics had been defined using general mobile applications characteristics and received feedback from those knowledgeable in HCI, and as such should be generally applicable to most task-based mobile applications across platforms.

8.3 Contributions to Knowledge

There are two main contributions to knowledge from this programme of research, the novelty of which should be viewed as a whole—that is, a combination of both contributions and the approach taken. The two main contributions to knowledge are:

1) Adapting traditional usability heuristics for mobile applications;
2) Extending mobile application heuristics to consider the impact of context of use.

8.3.1 *Adapting Traditional Usability Heuristics for Mobile Applications*

The first contribution of this programme of research is the adaptation of traditional heuristics, originally designed for desktop applications, for the usability evaluation of mobile applications. At face value, it might be argued that adapting heuristics for mobile applications is not new. Indeed, per the literature review, traditional heuristics have been adapted for mobile applications multiple times. Yet, the approach taken within this programme of research differs in important ways. Firstly, aspects of evaluation theory were considered whereas this is not always the case in previous research. Secondly, one of the strengths of this work is the engagement with HCI researchers and practitioners, which allowed for a better understanding of why gaps in knowledge continue to exist, if there are any other issues to consider, and just as importantly, to question if HCI researchers and practitioners would actually use a new set of heuristics. The issues that arose may impact the adoption of heuristic evaluation for mobile applications. For instance, participants believed that good mobile design and their own accumulated HCI knowledge negated the need to use heuristics. Yet, heuristic evaluation has been demonstrated to bring several benefits that complement these approaches. As such, this points to a change of mindset by HCI researchers and practitioners, more so than a change in the heuristics themselves. Such a change in mindset might be brought about through the literature, blog posts, and conference talks aimed at HCI researchers and practitioners.

Having used a six-step approach defined by Rusu et al. (2011) as a base methodology, and having included additional steps to increase validity, the result was a new set of mobile application usability heuristics. During a follow-up evaluation, the mobile application heuristics defined within this programme of research compared at least as well or better than other sets of heuristics in a number of key areas, including ease of use, ease of understanding, easy of learning, and in surfacing usability issues within a mobile application. Just as importantly, participants were more confident that they would use the newly defined mobile application heuristics to evaluate mobile applications within a professional context, that the heuristics were the most applicable in terms of application to a range of mobile devices and screen resolutions, that the heuristics would find most of the problems in any mobile application, and that the heuristics would be able to capture recent developments in mobile applications.

8.3.2 *Extending Mobile Application Heuristics to Consider the Impact of Context of Use*

The second contribution of this programme of research was to extend the mobile application usability heuristics into a contextual usability evaluation protocol that considers the impact of context of use on mobile application usability. To define this protocol, important theoretical underpinnings were considered, such as Stufflebeam's (2000b) CIPP model, which has not been fully considered within previous work. Further, as per the approach taken to define the usability heuristics for mobile applications, those knowledgeable in HCI were asked for their input. It became clear that the vast majority of tools and methods used by participants to consider context of use when designing and evaluating the usability of mobile applications were ineffective for the purpose, such as interviews, questionnaires, focus groups, and market research. Thus, HCI researchers and practitioners working on a mobile application may know much about the context of use that their users might find themselves in. However, the methods and tools in use today are less suitable to better understating how context of use may impact the perception of mobile application usability. Adding a heuristic focused on considering contexts of use when defining the set of mobile application heuristics was not effective, therefore a better approach was needed.

While the protocol was well-received, and considered to be easy to use, easy to understand, and easy to learn, as well as perceived as being able to uncover differences in the perception of usability of mobile applications within different contexts, there were also problems highlighted that needed to be addressed, such as the difficulty in recruiting thirty participants for a study and confusion about the proposed visualization types. All of these issues were addressed in the updated version of the contextual usability evaluation, bringing the protocol one step further to being adopted by HCI researchers and practitioners.

8.4 IMPLICATIONS

There are implications for both theory and practice based on the work conducted during this programme of research, the results of the studies, and the contributions to knowledge. This will be of most value to HCI researchers and practitioners, and potentially to HCI students and educators, as well as product managers, software engineers and others tasked with the creation of a mobile application that do not have resources with knowledgeable in HCI to call upon. The primary importance of this information for those knowledgeable in HCI is best used when planning the formative and summative usability evaluations of task-based mobile applications. Such evaluations may occur multiple times over the course of the design and development of a mobile application. The implications that follow for both theory and practice are suggestions, not imperatives.

8.4.1 *Implications for Theory*

From the perspective of theory, Stufflebeam's (2000a) CDC might be updated to consider some of the aspects of defining a set of usability heuristics, potentially for any domain, not only for mobile applications. In step 1 of the CDC, Stufflebeam (2000a) stated that an evaluator must "Clarify and justify the criteria to be met by the checklist". To aid the evaluator, Stufflebeam (2000a) offered ideas, such as applicability to the full range of intended uses. Yet, for step 9 when evaluating the checklist, Stufflebeam (2000a) stated that an evaluator must assess if the checklist met the requirements after which the original ideas are once again listed, but more so inferred as mandatory items, not ideas. My suggestion is to consider these items as ideas only, and to consider which checklist evaluation items are best suited to the creation of a set of heuristics. For example, an evaluation of heuristics asked participants to rate their belief that the heuristic set defined as part of this work was able to locate all the problems in the mobile application used, to locate most of the problems likely to occur in any mobile application, and to capture problems related to the most recent developments in mobile applications. Additionally, Stufflebeam (2000a) inferred in step 3 of the CDC that categories must be created. Yet, as has been suggested by those knowledgeable in HCI outside of this programme of research and according to some of the participants within this programme of research, a set of heuristics should be confined to ten. Once again, Stufflebeam (2000a) might consider the creation of categories as a suggestion, not a mandatory step within the CDC.

Suggestions can also be made to the same author regarding other work. For instance, within the list of twenty-two types of approaches to evaluation, Stufflebeam (2001) inferred that 'Approach 17: Consumer-oriented studies', which is most applicable to the creation of heuristics, did not need to involve consumers. While this might be true of heuristic evaluation, it is not true of the extension of heuristic evaluation defined within this programme of research, that is contextual usability evaluation. Including end-users within contextual usability evaluation studies allows those knowledgeable in HCI to understand if different contexts of use have an impact on mobile application usability. Given the resource constraints sometimes faced HCI teams, this could be difficult to achieve at best, and impossible at worst, if such approaches did not include end-users. On a related note, Scriven (1996) insisted that the purpose of a consumer-oriented study was a final summative judgement of an evaluand's merit and worth, not a formative judgement leading to continuous improvement as the evaluand is being developed. I would argue that this is not necessarily the case. A heuristic evaluation can be classified under Stufflebeam's (2001) 'Approach 17: Consumer-oriented studies', yet this method, as well as the extended heuristics within the contextual usability evaluation, can both be used during a final summative judgement of an evaluand's merit and worth, and as formative judgements leading to continuous improvement as an evaluand is being developed.

On the topic of an evaluand's merit and worth, Mertens & Wilson (2012) argued that both are not dependent on each other. I would argue that the opposite might also be true. When considering the context of use within which a mobile application is being evaluated, both merit and worth need to be considered. Should a mobile application be found to be easy to use in a laboratory environment, then only merit is taken into account as the worth of the mobile application, that is the merit of the application within specified contexts of use, is not considered. In this case, the outcome of the evaluation, that is the merit, is less accurate. As argued by Maguire (2001a. p. 453) "It is incorrect to describe a product as ergonomic or usable, without also describing the context in which the product will be used". Thus, merit might not depend on worth when conducting a laboratory-based heuristic evaluation. However, when conducting a contextual usability evaluation, merit is critically dependent on worth.

8.4.2 *Implications for Practice*

From the perspective of practice, Rogers (2004) maintained that researchers need to be cautious when applying theories from established fields to the messy world of HCI. I would argue that such applications of theory are critical. The discipline of HCI is too recent to have many established theories of its own, and HCI researchers and practitioners will learn much from the application of such theories from planning a research effort to the analysis of findings. Yet, HCI researchers and practitioners need to be open and flexible in order to account for the so-called 'messiness of the real world' outside of a laboratory. In turn, it is also important that HCI researchers and practitioners suggest changes to established theoretical foundations to account for this real-world messiness, which will assist others within HCI and beyond. Thus, one of my suggestions is that the HCI researchers and practitioners use theoretical foundations when considering the creation of new sets of usability heuristics or addition to existing sets of heuristics. A consideration of Scriven's (2005) comlist requirements, for instance, would have enabled sets of heuristics to be more concise, clear and not overlap. Equally as important is Stufflebeam's (2000b) CIPP model, whereby an evaluator should consider the approach best suited to the evaluation, and the environment within which an evaluand is being assessed.

Another suggestion is related to the first contribution of this programme of research, namely the adaptation of traditional usability heuristics for mobile applications—that is to use the most appropriate set of heuristics for the job at hand, not the most convenient nor the most well-known. Additionally, it is important to note that even though the maturity of mobile application design has come a lot way in the past decade, that does not negate the need to conduct heuristic evaluation, which is still likely to find issues that would otherwise impact the usability of the mobile application. Further, usability testing is much less expensive and easier to recruit for in 2020 than it was in 1990. Again, this does not mean that heuristic evaluation is not needed. Heuristic evaluation and usability testing will likely find different types of issues, complementing each other in creating a great mobile application user experience. I am also confident that other researchers will continue to define additional sets of heuristics across diverse domains. My suggestion is to follow defined methodologies, such as those from Rusu et al. (2011) and Quiñones & Rusu (2017). Following the steps laid out within these methodologies will help to ensure that the resulting heuristics are the most applicable for the job at hand.

Related to the second contribution, namely extending the mobile application heuristics to consider the impact of context of use. My suggestion to HCI researchers and practitioners is to continue evolving the work on this important topic. That includes the work started within this programme of research, as well as adapting and creating new, effective approaches that consider context of use when evaluating mobile applications and other digital products. Such suggestions are listed within the 'Recommendations for future research' section. On a related note, no matter what types of approaches that other researchers define, it is important to consider the constraints of the business world within which HCI practitioners work. Within this programme of research, HCI researchers and practitioners had several opportunities to critique the work, resulting in a much-improved set of mobile application heuristics and contextual usability evaluation protocol, the usage of which is described in the next section.

8.5 USING THE SET OF MOBILE APPLICATION HEURISTICS AND CONTEXTUAL USABILITY EVALUATION PROTOCOL

The set of mobile application heuristics and contextual usability evaluation protocol defined within this programme of research will be of use to HCI researchers and practitioners and can be applied in the following ways:

8.5.1 *Evaluating the Usability of Mobile Applications in a Laboratory*

When evaluating the usability of mobile applications in a laboratory or similar environment where the impact of context of use is more difficult to ascertain, the final set of heuristics defined within this programme of research (Appendix L) can be utilised. The procedure follows that of a standard heuristic evaluation, whereby one or more screens of a mobile application is evaluated against heuristics within the set. It is not required that all heuristics be used, only the heuristics that are deemed most applicable. The content within the square brackets following each heuristic is not required at this stage. Usability issues that are found can be evaluated on any number of severity scales, many of which are depicted by Sauro (2013).

It is recommended that at least three to five HCI researchers and/or practitioners conduct the evaluation. Fewer than the recommended number of evaluators can use the heuristics to evaluate the usability of mobile applications, but the evaluation may not be as effective. Should more than one evaluator conduct the evaluation, all evaluators should meet to discuss the usability issues found, including the frequency of issue occurrence, to remove duplicates, and to agree on severity scores for remaining issues. The mobile application heuristics can be used during formative and summative evaluations.

8.5.2 *Evaluating the Usability of Mobile Applications in the Field*

When evaluating the usability of mobile applications in the field where the impact of context of use is easier to ascertain, the final contextual usability evaluation protocol defined within this programme of research (Appendix L) can be utilised. To do so, HCI researchers and practitioners would initially prioritise tasks that users would attempt on a mobile application. The contexts of use within which these tasks will be attempted also need to be prioritised. Ideally, this information should come from prior research conducted earlier in a product development life cycle.

With the tasks and prioritised contexts of use determined, HCI researchers and practitioners can establish which elements of usability are most important to evaluate. The final set of heuristics defined within this programme of research (Appendix L) can be used for this purpose. It is not required that all heuristics be used, only the heuristics that are deemed most applicable. Once it is decided which heuristics to use, the content within the square brackets after each mobile application heuristic can be used as a guide, along with the tasks and contexts of use, to form one or more contextual usability evaluation user stories—the format of a user story is: When <attempting a task>, does <an element of context of use> have an impact on <mobile application usability>? For example, HCI researchers and practitioners may wish to evaluate the user story: When reading a book on the Kindle mobile app, does being shuffled around on a bus have an impact on readability in portrait mode?

Once the contextual usability evaluation user stories have been defined, participants for a study are recruited. Using participants and not HCI researchers and practitioners has two benefits. Firstly, it may be faster to conduct a contextual usability evaluation as it may be possible to recruit participants that regularly find themselves in particular contexts of use within which an HCI researcher or practitioner wishes to evaluate a mobile application. Secondly, using participants frees up potentially scarce HCI resources for other projects. Should a statistical analysis of the contextual usability evaluation results be of interest to an HCI researcher or practitioner conducting the evaluation, at least eight participants need to be recruited.

The next step is to conduct the contextual usability evaluation, whereby participants are given the contextual usability evaluation user stories. From these, participants know which task to attempt on which mobile application, the context of use the task needs to be attempted within, and which element(s) of usability that are being evaluated. Contexts of use can have more than one element, such as walking on a bright sunny day. The order of contexts of use should be randomised across participants. Having attempted a specific task based on a specific element of usability, each participant marks the task as either task success or failure. The participant is encouraged to write notes, or snippets if needed and notes later, to offer more details on why the task succeeded or failed. This cycle continues until all tasks have been attempted by all participants within all contexts of use.

The HCI researcher or practitioner analyses the data and writes up the final report. A statistical analysis can be conducted using Cochran's Q, which will determine if there is a difference in the perception of usability between contexts of use. For example, participants may have been more successful in attempting a task in a usability laboratory and less successful in a real-world context(s). Follow-up McNemar tests will allow the HCI researcher or practitioner to establish where the differences lie. Finally, results are displayed in statistical format and on a bar chart using green for task success and red for task failure or patterns should colour-blindness be a concern.

8.6 RECOMMENDATIONS FOR FUTURE RESEARCH

A substantial amount of work has been conducted during this programme of research and the research questions have been addressed. Yet, this work may never be done. Further research on this topic will be continued by myself and potentially by other HCI researchers and practitioners. Recommendations for further research include, but are not limited to:

- Using the mobile heuristics to evaluate the usability of other types of mobile applications to better understand if any changes need to be made, such as other task-based mobile applications and non-task-based mobile applications, such as games;
- Critically evaluating the set of mobile application heuristics defined within this work with other sets of usability heuristics, including those that currently exist and those that exist in the future;
- Using the mobile heuristics and contextual usability evaluation within business environments to ensure that both approaches are truly suitable, including within fast-paced software development environments;
- Considering the effectiveness and gaps in knowledge when evaluating the usability of other forms of mobile computing using the mobile heuristics, such as wearables and Augmented Reality/Virtual Reality (AR/VR);
- Considering the potential for combining the contextual usability evaluation protocol with usability testing;
- Considering the usage of the contextual usability evaluation protocol within other environments that change, such as the usability evaluation of

aircraft cockpits, nurses' stations in hospitals, user interfaces on devices used by construction workers etc.;

- Continuing the effort to research the impact of changing contexts of use on usability, not only for mobile applications, but also on other products across other diverse domains within and outside computing. That effort may consist of refining the contextual usability evaluation protocol or to introduce new, even more effective ideas that would work within ever faster Agile software development environments.

8.7 FINAL REMARKS

During the course of working on this programme of research, I have learned much about the topics of philosophy, HCI theory and practice. In addition to furthering my collective knowledge on the usability evaluation of mobile applications, the largest impact has been on my growth, strength and confidence as an HCI researcher/practitioner. That growth, strength and confidence has not only come from the increase in my personal knowledge, it has also come from learning what works well and what might be changed if I had an opportunity to do this work all over again.

While I felt that all the research questions were adequately addressed despite the limitations, one of the potential changes I would make would be to use closed answer choices in the questionnaire that investigated current methods used to consider context of use within section 6.1. The use of open-ended questions allowed participants to list different types of methods that might not have been part of a closed answer list, yet this might have been approached from the angle of using a relatively short set of methods along with an 'Other-write in' answer type. The approach that was taken did address the research questions, however the approach substantially elongated the time needed for data analysis. Another change that might have been made was the use of a focus group in section 7.1 to better understand how robust the contextual usability evaluation protocol was in considering actual contexts of use within which mobile applications are used. One focus group with twelve participants might have taken two hours and potentially have gathered the same insights as thirty interviews that were conducted over two months. Nonetheless, I am quite satisfied with the work that was conducted, with the way that each research question was addressed, and with the contributions and implications on an important topic. Thank you for reading.

REFERENCES

14th Annual State of Agile Development Survey. (2020). https://stateofagile.com/#ufh-i-615706098-14th-annual-state-of-agile-report/7027494

About Sands. (n.d.). Retrieved July 4, 2020, from https://www.sands.org.uk/

About UserTesting.com. (n.d.). https://www.usertesting.com/faq

Abowd, G. D., Dey, a K., Brown, P. J., Davies, N., Smith, M. E., Steggles, P., & Session, P. (1999). Towards a better understanding of context and context-awareness. *CHI 2000 Workshop on the What Who Where When and How of Contextawareness*, *4*(What, Who, Where, When and How of Context-Awareness), 1–6. https://doi.org/10.1007/3-540-48157-5

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, *11*, 325–338. https://doi.org/10.1023/A:1025869312943

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-Centered Design. In W. Bainbridge (Ed.), *Encyclopedia of Human-Computer Interaction*. Sage Publications.

Adobe company profile. (n.d.). https://www.adobe.com/about-adobe.html

Al-Zewairi, M., Biltawi, M., Etaiwi, W., & Shaout, A. (2017). Agile software development methodologies: survey of surveys. *Journal of Computer and Communications*, *5*(05), 74.

Alkhafaji, A., Fallahkhair, S., & Cocea, M. (2019). Design challenges for mobile and wearable systems to support learning on-the-move at outdoor cultural heritage sites. *IFIP Conference on Human-Computer Interaction*, 185–207.

Allen, M., Currie, L. M., Bakken, S., Patel, V. L., & Cimino, J. J. (2006). Heuristic evaluation of paper-based Web pages: A simplified inspection usability methodology. *Journal of Biomedical Informatics*, *39*, 412–423. https://doi.org/10.1016/j.jbi.2005.10.004

Alonso-Ríos, D., Mosqueira-Rey, E., & Moret-Bonillo, V. (2018). A systematic and generalizable approach to the heuristic evaluation of user interfaces. *International*

Journal of Human–Computer Interaction, *34*(12), 1169–1182.

Amelung, H., Ohl, C., Schade, G., & Wagner, S. (2010). Mobile Usability. In S. A. Ahson & M. Ilyas (Eds.), *Mobile Web 2.0: Developing and Delivering Services to Mobile Phones.* (pp. 60–65). CRC Press.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*. https://www.nature.com/articles/d41586-019-00857-9

Anney, V. N. (2014). Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. *Journal of Emerging Trends in Educational Research and Policy Studies*, *5*(2), 272–281.

Apple Company Profile. (n.d.). Markets Insider. https://markets.businessinsider.com/stocks/aapl/company-profile

Arif, A., Iltisberger, B., & Stuerzlinger, W. (2011). Extending mobile user ambient awareness for nomadic text entry. *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, 21–30. https://doi.org/10.1145/2071536.2071539

Auerbach, C., & Silverstein, L. (2003). *Qualitative Data. An Introduction to Coding and Analysis*. NYU Press.

Baber, C. (2009). Evaluating Mobile Human-Computer Interaction. In D. Taniar (Ed.), *Mobile Computing: Concepts, Methodologies, Tools, and Applications* (pp. 225–239). Information Science Reference. https://doi.org/doi:10.4018/978-1-60566-054-7.ch020

Baharuddin, R., Singh, D., & Razali, R. (2013). Usability dimensions for mobile applications-a review. *Research Journal of Applied Sciences, Engineering and Technology*, *5*(6), 2225–2231. http://maxwellsci.com/print/rjaset/v5-2225-2231.pdf

Bannon, L. J. (2006). Forgetting as a feature, not a bug: the duality of memory and implications for ubiquitous computing. *CoDesign*, *2*(1), 3–15.

Barrett, B. (2017). *Adobe Finally Kills Flash Dead: No, for real this time*. https://www.wired.com/story/adobe-finally-kills-flash-dead/

Beasley, M. (2013). *Practical Web Analytics for User Experience: How Analytics Can Help*

You Understand Your Users. Morgan Kaufmann.

Beck, E., Christiansen, M., Kjeldskov, J., Kolbe, N., & Stage, J. (2003). Experimental evaluation of techniques for usability testing of mobile systems in a laboratory setting. *Proc. of OzChi*, 106–115. http://people.cs.aau.dk/~jesper/pdf/papers/OzCHI03-final.pdf

Bedford, A. (2014). *Instructional Overlays and Coach Marks for Mobile Apps*. Nielsen Norman Group. http://www.nngroup.com/articles/mobile-instructional-overlay/

Benedek, J., & Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. *Proceedings of Usability Professionals Association*, 8–12.

Bergvall-Kåreborn, B. (2011). Mobile applications development on Apple and Google platforms. *Commun. Assoc. Inform. Syst*, *29*, 565–580.

Bernard, H. R. (2000). Handbook of Methods in Cultural Anthropology (2nd edition). In *Sociological Research Online* (Vol. 5). https://doi.org/10.1525/aa.2000.102.1.183

Bernhaupt, R., Mihalic, K., & Obrist, M. (2008). Usability Evaluation Methods for Mobile Applications. In J. Lumsden (Ed.), *Handbook of Research on User Interface Design and Evaluation for Mobile TechnologyJ. Lumsden* (pp. 745–758).

Bertini, E., Sapienza, L., Gabrielli, S., Sapienza, L., & Kimani, S. (2006). Appropriating and Assessing Heuristics for Mobile Computing. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 119–126.

Bevan, N., Carter, J., Earthy, J., Geis, T., & Harker, S. (2016). New ISO Standards for Usability , Usability Reports and Usability Measures. *International Conference on Human-Computer Interaction*, *July*, 268–278. https://doi.org/10.1007/978-3-319-39510-4

Bingham, W., & Moore, B. (1959). *How to Interview*. Harper International.

BlackBerry company profile. (n.d.). https://www.blackberry.com/us/en/company/overview

Blackmon, M. H. (2004). Cognitive Walkthrough. In W. S. Bainbridge (Ed.), *Encyclopedia of Human-Computer Interaction* (pp. 104–107). Berkshire Publishing Group.

Blaxter, L. (2010). *How to Research* (4th Ed.). McGraw-Hill Education.

Blom, B., & Morén, S. (2011). Analysis of Generative Mechanisms. *Journal of Critical Realism*, *10*(1), 60–79.

Bloomberg. (2002). *Usability is next to profitability*. http://www.bloomberg.com/bw/stories/2002-12-03/usability-is-next-to-profitability

Bobek, S., Nalepa, G. J., & Ślażyński, M. (2019). HEARTDROID—Rule engine for mobile and context-aware expert systems. *Expert Systems*, *36*(1).

Boehm, B. W. (1978). Characteristics of software quality. *North-Holland*, *1*.

Boren, M. T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*(3), 261–278. https://doi.org/10.1109/47.867942

Borisova, L. (2019). *How startups should evaluate their website's UX?*

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27.

Brandt, J., Weiss, N., & Klemmer, S. R. (2007). Txt 4 L8R: Lowering the Burden for Diary Studies Under Mobile Conditions. *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, 2303–2308. https://doi.org/10.1145/1240866.1240998

Brinck, T., Gergle, D., & Wood, S. (2002). *Designing Web sites that work: Usability for the Web*. Morgan Kaufmann.

Busk, N. K. (2011). *Smartphone Etiquette - Do we need a smartphone etiquette?*

Carroll, J. M. (2003). *HCI Models, Theories, and Frameworks*. Morgan Kaufmann.

Cerbone, D. R. (2014). *Understanding phenomenology*. Routledge.

Chittaro, L. (2011). Designing visual user interfaces for mobile applications. *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems - EICS '11*, 331–332. https://doi.org/10.1145/1996461.1996550

Choi, B. C., & Pak, A. W. (2005). Peer reviewed: a catalog of biases in questionnaires. *Preventing Chronic Disease*, *2*(1).

Christensen, L. B., Johnson, B., Turner, L. A., & Christensen, L. B. (2011). *Research methods, design, and analysis*.

Clark, J. (2010). *Tapworthy: Designing Great iPhone Apps* (1st ed.). O'Reilly Media.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3/4), 256–266.

Cockton, G., & Woolrych, A. (2002). Sale must end: should discount methods be cleared off HCI's shelves? *Interactions*, *9*(october), 13–18. https://doi.org/10.1145/566981.566990

Collins, H. (2010). *Creative Research: The Theory and Practice of Research for the Creative Industries*. AVA Publishing.

Combined global Apple App Store and Google Play app downloads from 1st quarter 2015 to 1st quarter 2020. (2020). https://www.statista.com/statistics/604343/number-of-apple-app-store-and-google-play-app-downloads-worldwide/

Cottingham, H., & Snyder, M. (2011). Get your mobile app out the door. *Communications in Computer and Information Science*, *173 CCIS*, 13–17. https://doi.org/10.1007/978-3-642-22098-2_3

Coursaris, C., & Kim, D. (2006). A Qualitative Review of Empirical Mobile Usability Studies. *Americas Conference in Information Systems (AMCIS)*, *6*(3), 117–171. https://doi.org/10.1038/nrc839

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research* (Third Edit). SAGE Publications.

Cruz, L., Abreu, R., & Lo, D. (2019). To the attention of mobile software developers: guess what, test your app! *Empirical Software Engineering*, *24*(4), 2438–2468.

Cypress, B. S. (2017). Rigor or reliability and validity in qualitative research: Perspectives, strategies, reconceptualization, and recommendations. *Dimensions of Critical Care Nursing*, *36*(4), 253–263.

Da Costa, R. P., Canedo, E. D., De Sousa, R. T., Albuquerque, R. D. O., & Villalba, L. J. G. (2019). Set of usability heuristics for quality assessment of mobile applications on smartphones. *IEEE Access*, *7*, 116145–116161.

Denscombe, M. (2010). *Good Research Guide : For small-scale social research projects* (4th Ed.). McGraw-Hill Education.

Deraman, A. B., & Salman, F. A. (2019). Managing usability evaluation practices in agile development environments. *International Journal of Electrical & Computer Engineering*, *9*(2), 2088–8708.

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing Journal*, *1*(5), 4–7.

Dicks, R. S. (2002). Mis-usability : On the uses and misuses of usability testing. *SIGDOC '02 Proceedings of the 20th Annual International Conference on Computer Documentation*, 26–30. https://doi.org/10.1145/584955.584960

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2003). *HUMAN-COMPUTER INTERACTION* (3rd ed.). Pearson Education Limited.

Dourish, P. (2004). What we talk about when we talk about context. *Ubiquit Computing*, *8*, 19–30. https://doi.org/10.1007/s00779-003-0253-8

Duh, H., Tan, G., & Chen, V. (2006). *Usability evaluation for mobile device*. 181. https://doi.org/10.1145/1152215.1152254

Dunlop, M., & Brewster, S. (2002). The challenges of mobile devices for human computer interaction (editorial for special edition). *Personal and Ubiquitous Computing*, 3–5. http://strathprints.strath.ac.uk/2574/

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928.

Echessa, J. (2014). *Improving Your App's Onboarding UX.*

http://www.sitepoint.com/improving-apps-onboarding-ux/

Elkoutbi, M., Khriss, I., & Keller, R. K. (1999). Generating user interface prototypes from scenarios. *Proceedings IEEE International Symposium on Requirements Engineering (Cat. No.PR00188)*, 150–158. https://doi.org/10.1109/ISRE.1999.777995

Eshet, E. (2016). *Examining human-centered design practice in the mobile apps era* (Issue 212). https://doria32-kk.lib.helsinki.fi/handle/10024/123313

Eshet, E., & Bouwman, H. (2017). Context of Use: The Final Frontier in the Practice of User-Centered Design? *Interacting with Computers*, *29*(3), 368–390. https://doi.org/10.1093/iwc/iww030

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fernandez, A., Abrahão, S., & Insfran, E. (2012). A systematic review on the effectiveness of Web usability evaluation methods. *16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012)*, *2012*, 52–56. https://doi.org/10.1049/ic.2012.0007

Ferreira, D., Goncalves, J., Kostakos, V., Barkhuus, L., & Dey, A. K. (2014). *Contextual experience sampling of mobile application micro-usage*. *SEPTEMBER*, 91–100. https://doi.org/10.1145/2628363.2628367

Fetterman, D. M. (2019). *Ethnography: Step-by-step*. SAGE Publications.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Fiotakis, G., Raptis, D., & Avouris, N. (2009). Considering cost in usability evaluation of mobile applications: who, where and when. *Interact*, 231–234.

Fouka, G., & Mantzorou, M. (2011). What are the major ethical issues in conducting research? Is there a conflict between the research ethics and the nature of nursing? *Health Science Journal*, *5*(1).

Franklin, C., & Ballan, M. (2001). Reliability and validity in qualitative research. *The*

Handbook of Social Work Research Methods, *4*(273–292).

Frauenberger, C. (2016). Critical realist HCI. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 341–351.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, *11*(1), 86–92.

Frost, N. (2011). *Qualitative Research Methods in Psychology*. McGraw-Hill Education.

Fu, B., Lin, J., Liy, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013). Why people hate your App - Making sense of user feedback in a mobile app store. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1288*, 1276–1284. https://doi.org/10.1145/2487575.2488202

Gadamer, H. G. (1976). *Philosophical hermeneutics* (D. E. T. Linge (Ed.)). University of California Press.

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, *13*(1), 117. https://doi.org/10.1186/1471-2288-13-117

Galitz, W. (2007). *The essential guide to user interface design: An introduction to GUI design principles and techniques*. John Wiley & Sons.

García-Horta, J., & Guerra-Ramos, M. (2009). The use of CAQDAS in educational research: some advantages, limitations and potential risks. *International Journal of Research & Method in Education*, *32*(2), 151–165.

Garner, M., Wagner, C., & Kawulich, B. (2009). *Teaching Research Methods in the Social Sciences*. Ashgate Publishing Group.

Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. Metropolitan Books.

Georgsson, M., & Staggers, N. (2016). An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. *Journal of Biomedical*

Informatics, *59*, 115–129.

Gerhardt-Powals, J. (1996). Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction*, *8*(Issue 2), 189–211. https://doi.org/10.1080/10447319609526147

Ginsburg, S. (2010). *Designing the iPhone User Experience: A User-centered Approach to Sketching and Prototyping iPhone Apps*. Addison-Wesley Professional.

Giorgi, A., Giorgi, B., & Morley, J. (2017). The Descriptive Phenomenological Psychological Method. In C. Willig & W. S. Rogers (Eds.), *The SAGE Handbook of Qualitative Research in Psychology*. SAGE Publications.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, *8*(4), 597–607. http://www.nova.edu/ssss/QR/QR8-4/golafshani.pdf

Gómez, R. Y., Caballero, D. C., Sevillano, J., Yáñez Gómez, R., Cascado Caballero, D., & Sevillano, J. (2014). Heuristic evaluation on mobile interfaces: a new checklist. *TheScientificWorldJournal*, *2014*, 434326. https://doi.org/10.1155/2014/434326

Gong, J., & Tarasewich, P. (2004). Guidelines for handheld mobile device interface design. *Proceedings of DSI 2004 Annual Meeting*, 3751–3756. https://doi.org/10.1.1.87.5230

Google company profile. (n.d.). https://about.google/

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, *28*(3), 300–311. https://doi.org/10.1145/3166.3170

Grayling, T. (1998). Fear and Loathing of the Help Menu : A Usability Test of Online Help. *Technical Communication*, *November 1997*, 166–177.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*, *29*(2), 75.

Guest, G. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82.

https://doi.org/10.1177/1525822X05279903

Gulliksson, H. (2012). *Human-information-thing interaction: technology and design*.

Gunthert, K. C., & Wenze, S. J. (2012). *Daily diary methods*.

Hagen, P., Robertson, T., Kan, M., & Sadler, K. (2005). Emerging research methods for understanding mobile technology use. *OZCHI '05 Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, 1–10. http://dl.acm.org/citation.cfm?id=1108417

Hall, T., Sharp, H., Beecham, S., Baddoo, N., & Robinson, H. (2008). What do we know about developer motivation? *IEEE Software*, *25*(4), 92–94. https://doi.org/10.1109/MS.2008.105

Han, M., Vinh, L. T., Lee, Y. K., & Lee, S. (2012). Comprehensive context recognizer based on multimodal sensors in a smartphone. *Sensors (Switzerland)*, *12*(12), 12588–12605. https://doi.org/10.3390/s120912588

Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science*, *1*(1), 1–16. https://doi.org/10.1186/2194-0827-1-1

Heidegger, M. (1971). *Building, dwelling, thinking* (A. T. Hofstadter (Ed.)). Springer.

Hermawati, S., & Lawson, G. (2016). Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus? *Applied Ergonomics*, *56*, 34–51. https://doi.org/10.1016/j.apergo.2015.11.016

Hesse-Biber, S. N. (2010). *Mixed Methods Research : Merging Theory with Practice*. The Guilford Press.

Hiller, J. (2016). Epistemological foundations of objectivist and interpretivist research. *Music Therapy Research*, 99–117.

Hintze, J. (1998). *NCSS statistical software*.

Høegh, R., Kjeldskov, J., Skov, M., & Stage, J. (2008). A field laboratory for evaluating in situ. *Handbook of Research on User Interface Design and Evaluation for Mobile*

Technology, IGI Global, 982–996.

Høegh, R. T., Kjeldskov, J., Skov, M. B., & Stage, J. (2008). A field laboratory for evaluating in situ. In *Handbook of Research on User Interface Design and Evaluation for Mobile TechnologyIGI Global* (pp. 982–996). IGI Global.

Hoekman, R. (2010). *Designing the Obvious: A Common Sense Approach to Web and Mobile Application Design* (Second Edi). Pearson Education.

Hollingsed, T., & Novick, D. G. (2007). Usability inspection methods after 15 years of research and practice after 15 Years of Research and Practice. *Proceedings of the 25th Annual ACM International Conference on Design of Communication*, 249–255.

Hopkins, J., & Turner, J. (2012). Go Mobile: Location-Based Marketing, Apps, Mobile Optimized Ad Campaigns, 2D Codes and Other Mobile Strategies to Grow Your Business. In *What's the difference between a mobile website and an app?* John Wiley & Sons.

Humayoun, S. R., Chotala, P. H., Bashir, M. S., & Ebert, A. (2017). Heuristics for evaluating multi-touch gestures in mobile applications. *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*, 89.

Humayoun, S. R., Dubinsky, Y., & Catarci, T. (2014). User Evaluation Support through Development Environment for Agile Software Teams. In *Smart Organizations and Smart Artifacts* (pp. 183–191).

Hussain, A., & Mkpojiogu, E. O. (2016). Usability evaluation techniques in mobile commerce applications: a systematic review. *AIP Conference Proceedings*.

Husserl, E. (1931). *Ideas: General introduction to pure phenomenology* (W. R. B. T. Gibson (Ed.)). Springer.

Hvannberg, E. T., Law, E. L. C., & Lárusdóttir, M. K. (2007). Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers*, *19*, 225–240. https://doi.org/10.1016/j.intcom.2006.10.001

Inbar, O., Lavie, T., & Meyer, J. (2009). Acceptable intrusiveness of online help in mobile devices. *Proceedings of the 11th International Conference on Human-Computer*

Interaction with Mobile Devices and Services - MobileHCI '09, 1. https://doi.org/10.1145/1613858.1613892

Inostroza, R., Rusu, C., Roncagliolo, S., Jiménez, C., & Rusu, V. (2012). Usability heuristics for touchscreen-based mobile devices. *Proceedings of the 9th International Conference on Information Technology, ITNG 2012*, 662–667. https://doi.org/10.1109/ITNG.2012.134

Inostroza, R., Rusu, C., Roncagliolo, S., Rusu, V., & Collazos, C. A. (2016). Developing SMASH: A set of SMArtphone's uSability Heuristics. *Computer Standards & Interfaces*, *43*, 40–52. https://doi.org/10.1016/j.csi.2015.08.007

ISO. (n.d.). https://www.iso.org/home.html

ISO 9241-11:1998. (1998). Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) — Part 11: Guidance on Usability. https://www.iso.org/standard/16883.html

ISO 9241-210. (2010). *BSI Standards Publication Ergonomics of human-system interaction Part 210 : Human-centred design for interactive systems*.

ISO 9241-210. (2019). *Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems.* https://www.iso.org/standard/77520.html

Jacko, J. A., & Sears, A. (2003). *Handbook of Research on Ubiquitous Computing Technology for Real Time Enterprises*.

Jackson, R. L., Drummond, D. K., & Camara, S. (2007). What is qualitative research? *Qualitative Research Reports in Communication*, *8*(1), 21–28.

Jaferian, P., Hawkey, K., Sotirakopoulos, A., & Beznosov, K. (2011). Heuristics for Evaluating IT Security Management Tools (Extended Abstract). *29th International Conference Extended Abstracts on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1080/07370024.2013.819198

Jantvongso, S., & Nuansomsri, C. (2020). Heuristic Evaluation of "PaPaYa PokPok": Case Study of a Mobile Game. *KKU Research Journal (Graduate Studies)*, *20*(2), 21–32.

Jaspers, M. W. M. (2009). A comparison of usability methods for testing interactive health technologies: methodological aspects. Advances in Human-Computer Interaction and empirical evidence. *International Journal of Medical Informatics*, *78*(5), 40–353.

Jeddi, F. R., Nabovati, E., Bigham, R., & Farrahi, R. (2020). Usability evaluation of a comprehensive national health information system: A heuristic evaluation. *Informatics in Medicine Unlocked*, *100332*.

Jeffries, R., & Desurvire, H. W. (1992). Usability Testing vs. Heuristic Evaluation: Was There a Contest? *ACM SIGCHI Bulletin*, *24*(4), 39–41.

Jeffries, R., Miller, J., Wharton, C., & Uyeda, K. (1991). User interface evaluation in the real world: a comparison of four techniques. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Reaching through Technology - CHI '91*, 119–124. https://doi.org/10.1145/108844.108862

Jeng, J. (2005). Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability. *Libri*, *55*(2–3), 96–121.

Ji, Y. G., Park, J. H., Lee, C., & Yun, M. H. (2006). A Usability Checklist for the Usability Evaluation of Mobile Phone User Interface. *International Journal of Human-Computer Interaction*, *20*(3), 207–231. https://doi.org/10.1207/s15327590ijhc2003_3

Jiang, K., Ling, F., Feng, Z., Ma, C., Kumfer, W., Shao, C., & Wang, K. (2018). Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study. *Accident Analysis & Prevention*, *115*, 170–177.

Jokela, T., Ojala, J., & Olsson, T. (2015). A Diary Study on Combining Multiple Information Devices in Everyday Activities and Tasks. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3903–3912. https://doi.org/10.1145/2702123.2702211

Joyce, G. (2013). *An investigation of heuristics for the evaluation of smartphone-deployed mobile applications [Unpublished Master's thesis]*. University of Hertfordshire.

Joyce, G. (2014). Adaption of Usability Evaluation Methods for Native Smartphone

Applications. *The 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, 409–410.

Joyce, G., & Lilley, M. (2014). Towards the Development of Usability Heuristics for Native Smartphone Mobile Applications. *International Conference of Design, User Experience, and Usability*, 465–474.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2014). Adapting Heuristics for the Mobile Panorama. *Proceedings of ACM Interacción '14, the XV International Conference on Human Computer Interaction*, 1–2.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2016a). Evaluating the impact of changing contexts on mobile application usability within agile environments. *2016 Future Technologies Conference (FTC)*, 476–480.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2017a). From healthcare to Human-Computer interaction: Using framework analysis within qualitative inquiry. *International Conference on Applied Human Factors and Ergonomics*, 93–100.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2018). Heuristic evaluation for mobile applications: Extending a map of the literature. *International Conference on Applied Human Factors and Ergonomics*, 15–26.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2016b). Mobile application tutorials: Perception of usefulness from an HCI expert perspective. *International Conference on Human-Computer Interaction*, 302–308.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2016c). Mobile application usability: Heuristic evaluation and evaluation of heuristics. *Advances in Human Factors, Software, and Systems Engineering*, 77–86.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2017b). Mobile application usability heuristics: Decoupling context-of-use. *International Conference of Design, User Experience, and Usability*, 410–423.

Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2015). Smartphone Application Usability Evaluation: The Applicability of Traditional Heuristics. *International Conference of*

Design, User Experience, and Usability, 541–550.

Jumisko-Pyykkö, S., & Vainio, T. (2010). Framing the Context of Use for Mobile HCI. *International Journal of Mobile Human Computer Interaction*, *2*(4), 1–28. https://doi.org/10.4018/jmhci.2010100101

Kaikkonen, A., Kekalainen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies*, *1*(1), 4–16. https://doi.org/10.1234/12345678

Kendall, M. G., & Gibbons, J. D. (1990). *Rank correlation methods*. Oxford University Press.

Ketola, P., & Röykkee, M. (2001). The Three Facets of Usability In Mobile Handsets. *CHI 2001 Workshop: Mobile Communications: Understanding Users, Adoption & Design*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.4302&rep=rep1&type=pdf

King, N., Horrocks, C., & Brooks, J. (2018). *Interviews in Qualitative Research* (Second edi). SAGE Publications.

Kirmani, S., & Rajasekaran, S. (2007). Heuristic Evaluation Quality Score (HEQS): a measure of heuristic evaluation skills. *Journal of Usability Studies*, *2*(2), 61–75. https://doi.org/10.1.1.111.6049

Kjeldskov, J., & Skov, M. B. (2014). Was it Worth the Hassle ? Ten Years of Mobile HCI Research Discussions on Lab and Field Evaluations. *Acm*, 43–52. https://doi.org/10.1145/2628363.2628398

Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004). Is It Worth the Hassle ? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. *Mobile Human-Computer Interaction Proceedings of the 6th International Symposium*, 61–73. https://doi.org/10.1007/978-3-540-28637-0_6

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies*, *60*(5–6), 599–620. https://doi.org/10.1016/j.ijhcs.2003.11.001

Korhonen, H., & Koivisto, E. M. (2006). Playability heuristics for mobile games.

Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, 9–16.

Krosnick, J. (2018). Questionnaire design. In *The Palgrave handbook of survey research* (pp. 439–455). Palgrave Macmillan, Cham.

Kumar, B. A., & Goundar, M. S. (2019). Usability heuristics for mobile learning applications. *Education and Information Technologies*, *24*(2), 1819–1833.

Kuparinen, L., Silvennoinen, J., & Isomäki, H. (2013). Introducing Usability Heuristics for Mobile Map Applications. *Proceedings of the 26th International Cartographic Conference (ICC 2013).*

Kuzel, A. (1992). Sampling in qualitative inquiry. In B. Crabtree & W. Miller (Eds.), *Doing qualitative research* (pp. 31–44). SAGE Publications.

Kwasnicka, D., Dombrowski, S. U., White, M., & Sniehotta, F. F. (2014). Qualitative Study Assessing Technology Usability in People who Have Participated in Multimodal Weight Loss Study. *European Health Psychologist*, *16*(S), 807.

Lacey, A., & Luff, D. (2001). Trent Focus for Research and Development in Primary Health Care: An Introduction to Qualitative Research. *Research and Development Group of NHS Executive Trent*, 1–27. https://doi.org/10.1136/ebnurs.2011.100352

Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2015). Test development process. In *Handbook of test development* (pp. 19–34). Routledge.

Law, E. L.-C., & Hvannberg, E. T. (2004). Analysis of strategies for improving and estimating the effectiveness of heuristic evaluation. *Proceedings of the Third Nordic Conference on Human-Computer Interaction - NordiCHI ’04*, 241–250. https://doi.org/10.1145/1028014.1028051

Lee, N. (2010). *The 411: Feature phones vs. smartphones*. https://www.cnet.com/news/the-411-feature-phones-vs-smartphones/#:~:text=A smartphone%2C as CNET defines,a third-party operating system.&text=Feature phones%2C on the other,phones support third-party software.

Lee, V., Schneider, H., & Schell, R. (2004). *Mobile Applications: Architecture, Design, and*

Development. Prentice Hall.

Letherby, G., Scott, J., & Williams, M. (2012). *Objectivity and subjectivity in social research*. SAGE Publications.

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care*, *4*(3), 324.

Lincoln, Y. S., & Guba, E. G. (1980). The distinction between merit and worth in evaluation. *Educational Evaluation and Policy Analysis*, *2*(4), 61–71.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Longoria, R. (2004). Designing software for the mobile context : a practitioner's guide. In *Computer communications and networks,*. Springer.

Loranger, H. (2014). *Infinite Scrolling Is Not for Every Website*. https://www.nngroup.com/articles/infinite-scrolling/

Love, S. (2005). *Understanding Mobile Human-Computer Interaction*. Elsevier.

Lubis, M., Sutoyo, E., Azuddin, M., & Handayani, D. (2019). User Experience in Mobile Application Design: Utility Defined Context of Use. *Journal of Physics*, *1361*(1), 012043.

Lyzara, R., Purwandari, B., Zulfikar, M. F., Santoso, H. B., & Solichah, I. (2019). E-government usability evaluation: Insights from a systematic literature review. *2nd International Conference on Software Engineering and Information Management*, 249–253.

MacDonald, C. M. (2019). User Experience (UX) Capacity-Building: A Conceptual Model and Research Agenda. *2019 on Designing Interactive Systems Conference*, 187–200.

Machado Neto, O., & Pimentel, M. D. G. (2013). Heuristics for the assessment of interfaces of mobile devices. *Proceedings of the 19th Brazilian Symposium on Multimedia and the Web - WebMedia '13*, 93–96. https://doi.org/10.1145/2526188.2526237

Mack, N. (2005). *Qualitative research methods: A data collector's field guide*.

Maguire, M. (2001a). Context of Use within usability activities. *International Journal of Human-Computer Studies*, *55*(4), 453–483. https://doi.org/10.1006/ijhc.2001.0486

Maguire, M. (2001b). Methods to support human-centred design. *International Journal of Human-Computer Studies*, *55*(4), 587–634. https://doi.org/10.1006/ijhc.2001.0503

Maguire, M. (2019). Development of a Heuristic Evaluation Tool for Voice User Interfaces. *International Conference on Human-Computer Interaction*, 212–225.

Malan, D. J. (2011). *Computer Science E-76 - Building Mobile Applications Lecture*. http://cs76.tv/2012/spring/

Male, T. (2016). Analysing Qualitative Data. In I. Palaiologou, D. Needham, & T. Male (Eds.), *Doing Research in Education: Theory and Practice* (pp. 177–191). SAGE.

Marinacci, J. (2012). *Building Mobile Applications with Java: Using the Google Web Toolkit and PhoneGap*. O'Reilly Media.

Martin, K., & Shilton, K. (2016). Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society*, *32*(3), 200–216.

Maruyama, G., & Ryan, C. S. (2014). *Research methods in social relations* (Eight Ed.). John Wiley & Sons.

Mason, W., Mirza, N., & Webb, C. (2018). *Using the framework method to analyze mixed-methods case studies*. SAGE Publications.

Mathison, S. (2005). *Encyclopedia of evaluation*. SAGE Publications.

Mays, N., & Pope, C. (1995). Rigour and qualitative research. *BMJ: British Medical Journal*, *311*(6997), 109.

McAleese, S., & Kilty, J. M. (2019). Stories matter: Reaffirming the value of qualitative research. *The Qualitative Report*, *24*(4), 822–845.

McCambridge, J., de Bruin, M., & Witton, J. (2012). The effects of demand characteristics on research participant behaviours in non-laboratory settings: A systematic review. *PloS One*, *7*(6). https://doi.org/10.1371/journal.pone.0039116

McEvoy, P., & Richards, D. (2006). A critical realist rationale for using a combination of quantitative and qualitative methods. *Journal of Research in Nursing*, *11*(1), 66–78.

McGregor, M., Brown, B., McMillan, D., & Chi, W. (2014). 100 Days of iPhone Use : Mobile Recording in the Wild. *Chi'14*, 2335–2340. https://doi.org/10.1145/2559206.2581296

McIlroy, S., Ali, N., & Hassan, A. E. (2016). Fresh apps: an empirical study of frequently-updated mobile apps in the Google play store. *Empirical Software Engineering*, *21*(3), 1346–1370. https://doi.org/10.1007/s10664-015-9388-2

McMillan, D., McGregor, M., & Brown, B. (2015). From in the wild to in vivo. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*, 494–503. https://doi.org/10.1145/2785830.2785883

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(Sup1), 235–245.

McWherter, J., & Gowell, S. (2012). *Professional Mobile Application Development*. John Wiley & Sons.

Mertens, D. M., & Wilson, A. T. (2012). *Program Evaluation Theory and Practice : A Comprehensive Guide*. Guilford Publications.

Micallef, A. (2015). *UX Matters: Agile Manifesto for Product Management*. http://www.uxmatters.com/mt/archives/2015/06/agile-manifesto-for-product-management-part-1.php

Microsoft company profile. (n.d.). https://www.microsoft.com/en-us/about/company

Millen, D. R. (2000). Rapid Ethnography: Time Deepening Strategies for HCI Field Research. *3rd Conf. Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 280–286.

Miller, R. B. (1971). *Human ease of use criteria and their tradeoffs*.

Mogil, J. S., & Macleod, M. R. (2017). No publication without confirmation. *Nature News*, *542*(7642), 409.

Mohajan, H. K. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University. Economic Series*, *17*(4), 59–82.

Moran, D. (2000). *Introduction to Phenomenology*. Routledge.

Morgan, A. (2011). *Investigating Our Experience in the World : A Primer on Qualitative Inquiry*. University of Tennessee Press.

Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, *20*(8), 1045–1053.

Murphy, D. (2012). *Apple's App Store Hits 25 Billion Downloads: How Many Per iPhone?* PC Magazine. http://www.pcmag.com/article2/0,2817,2401122,00.asp

Myszewski, J. M. (2012). Management responsibility for human errors. In *The TQM Journal* (Vol. 24, pp. 326–337). https://doi.org/10.1108/17542731211247355

Nacheva, R. (2020). Standardization Issues of Mobile Usability. *LearnTechLib*, 149–157.

Nalepa, G. J., Kutt, K., & Bobek, S. (2019). Mobile platform for affective context-aware systems. *Future Generation Computer Systems*, *92*, 490–503.

Nardi, B. A. (1992). The use of scenarios in design. *ACM SIGCHI Bulletin*, *24*(October), 13–14. https://doi.org/10.1145/142167.142171

Nassiuma, D. K. (2001). *Survey sampling: Theory and methods*.

Negulescu, M., Ruiz, J., Li, Y., & Lank, E. (2012). Tap, swipe, or move- attentional demands for distracted smartphone input. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, 173–180. https://doi.org/10.1145/2254556.2254589

Neil, T. (2014). *Mobile Design Pattern Gallery: UI Patterns for Mobile Applications* (Second Edi). O'Reilly Media.

Newton, C. (2012). *Google Play store hits 25 billion downloads, launches discounts*. CNET. http://news.cnet.com/8301-1023_3-57520324-93/google-play-store-hits-25-billion-downloads-launches-discounts/

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings of the SIGCHI Conference on Human*, 373–380. https://doi.org/10.1145/142750.142834

Nielsen, J. (1993). *Usability engineering*. Academic Press.

Nielsen, J. (1994). *10 Usability Heuristics for User Interface Design*. https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J. (2008). *Usability ROI Declining, But Still Strong*. http://www.nngroup.com/articles/usability-roi-declining-but-still-strong/

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*, 206–213. https://doi.org/10.1145/169059.169166

Nielsen, J., & Molich, R. (1990). Heuristic Evaluation of user interfaces. *Chi'90 - SIGCHI Conference on Human Factors in Computing Systems*, *April*, 249–256. https://doi.org/10.1145/97243.97281

Nilsson, E. G. (2009). Design patterns for user interface for mobile applications. *Advances in Engineering Software*, *40*(12), 1318–1328. https://doi.org/10.1016/j.advengsoft.2009.01.017

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence-Based Nursing*, *18*(2), 34–35. https://doi.org/10.1136/eb-2015-102054

Nørgaard, M., & Hornbæk, K. (2006). What do usability evaluators do in Practice? An explorative study of think aloud testing. *DIS*, 209–219.

Norman, D. A., & Draper, S. W. (1986). *User Centered System Design*. Lawrence Erlbaum Associates.

Novick, D. G., & Ward, K. (2006). Why Don't People Read the Manual ? *Proceedings of the 24th Annual Conference on Design of Communication - SIGDOC '06*, 11.

https://doi.org/10.1145/1166324.1166329

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Nvivo. (n.d.). https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Obrist, M., & Roto, V. (2009). User Experience Evaluation – Do You Know Which Method to Use? *CHI 2009 - Conference on Human Factors in Computing Systems*, 2763–2766. https://doi.org/10.1145/1520340.1520401

Ortiz, C. A. (2012). *The Psychology of Lean Improvements: Why Organizations Must Overcome Resistance and Change the Culture*. {CRC} Press.

Paz, F., & Pow-Sang, J. A. (2016). Usability Evaluation Methods for Software Development: A Systematic Mapping Review. *Proceedings - 8th International Conference on Advanced Software Engineering and Its Applications, ASEA 2015*, *March*, 1–4. https://doi.org/10.1109/ASEA.2015.8

Pérez-Edgar, K., MacNeill, L. A., & Fu, X. (2020). Navigating Through the Experienced Environment: Insights From Mobile Eye Tracking. *Current Directions in Psychological Science*, *0963721420*.

Pernecky, T., & Jamal, T. (2010). (Hermeneutic) Phenomenology in tourism studies. *Annals of Tourism Research*, *37*(4), 1055–1075. https://doi.org/10.1016/j.annals.2010.04.002

Petrillo, F., Spritzer, A. S., Freitas, C. D. S., & Pimenta, M. (2011). Interactive analysis of Likert scale data using a multichart visualization tool. *Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*, 358–365.

Picoto, W. N., Duarte, R., & Pinto, I. (2019). Uncovering top-ranking factors for mobile apps through a multimethod approach. *Journal of Business Research*, *101*, 668–674.

Pierre, R. da S. de Q. (2015). Heuristics in Design: A Literature Review. *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences*, *3*(AHFE), 6571–6578. https://doi.org/10.1016/j.promfg.2015.07.961

Po, S., Howard, S., Vetere, F., & Skov, M. B. (2004). Heuristic evaluation and mobile usability: Bridging the realism gap. *Mobile Human-Computer Interaction–MobileHCI 2004*, *3160*, 591–592. https://doi.org/10.1007/b100594

Polkinghorne, D. E. (1983). *Methodology for the human sciences*. State University of New York Press.

Privitera, M. B., Kossack, M., & Cox, K. (2019). Human factors validation (summative usability) testing including residual risk analysis. *Applied Human Factors in Medical Device Design*, 249–269.

Quinn, C. N. (1996). Pragmatic evaluation: lessons from usability. *Proceedings of the 13th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE'96)*.

Quiñones, D., & Rusu, C. (2017). How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, *53*, 89–122. https://doi.org/10.1016/j.csi.2017.03.009

Quiñones, D., Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C. A., & Usabilidad, A. (2016). Developing Usability Heuristics : A Formal or Informal Process ? *IEEE LATIN AMERICA TRANSACTIONS*, *14*(7), 3400–3409.

Rahi, S., Alnaser, F. M., & Abd Ghani, M. (2019). Designing survey research: recommendation for questionnaire development, calculating sample size and selecting research paradigms. *Economic and Social Development: Book of Proceedings*, 1157–1169.

Rahmati, A., & Zhong, L. (2013). Studying smartphone usage: Lessons from a four-month field study. *IEEE Transactions on Mobile Computing*, *12*(7), 1417–1427. https://doi.org/10.1109/TMC.2012.127

Raison, C., & Schmidt, S. (2013). Keeping user centred design (UCD) alive and well in your organisation: Taking an agile approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8012 LNCS*(PART 1), 573–582. https://doi.org/10.1007/978-3-642-39229-0_61

Rakotonirainy, A., Loke, S. W., & Fitzpatrick, G. (2000). Context-awareness for the mobile environment. *CHI2000 Workshop #11 Proposal*.

Reichmuth, R., & Möller, S. (2014). Classification of the Context of Use for Smart Phones. In C. Stephanidis (Ed.), *HCI International 2014 - Posters' Extended Abstracts SE - 110* (Vol. 435, pp. 638–642). Springer International Publishing. https://doi.org/10.1007/978-3-319-07854-0_110

Reiss, E. (2012). *Usable Usability: Simple Steps for Making Stuff Better (Google eBook)*. Wiley. http://books.google.com/books?id=5Wzj9IWcAGIC&pgis=1

Resnik, D. B. (2011). What is ethics in research & why is it important. *National Institute of Environmental Health Sciences*, *1*(10), 49–70.

Ritchie, J., Lewis, J., McNaughton, N., & Ormston, R. (2013). *Qualitative research practice: A guide for social science students and researchers*. SAGE Publications.

Robbins, N. B., Heiberger, R. M., Court, C., & Hall, S. (2011). Plotting Likert and Other Rating Scales. *Joint Statistical Meetings*, *October 2005*, 1058–1066. http://www.amstat.org/sections/srms/proceedings/y2011/Files/300784_64164.pdf

Roethlisberger, F., & Dickson, W. (1939). *Management and the Worker*. Harvard University Press.

Rogers, Y. (2004). New theoretical approaches for HCI. *Annual Review of Information Science and Technology*, *38*, 1–43. https://doi.org/10.1002/aris.1440380103

Rogers, Y., Connelly, K., Tedesco, L., Hazlewood, W., Kurtz, A., Hall, R. E., Hursey, J., & Toscos, T. (2007). Why It's Worth the Hassle - The Value of In-Situ Studies When Designing Ubicomp. *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp'07)*, *4717*, 336–353. https://doi.org/10.1007/978-3-540-74853-3

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: beyond human-computer interaction*. John Wiley & Sons.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*, 313–338.

Rosenbaum, S., Cockton, G., Coyne, K., Muller, M., & Rauch, T. (2002). Focus groups in HCI: wealth of information or waste of resources? *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, 702–703.

Rosoff, M. (2012). *For The First Time Ever, Smartphones Outsold PCs Last Quarter*. Business Insider. https://www.businessinsider.com/for-the-first-time-ever-smartphones-are-outselling-pcs-2012-2

Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley.

Rusu, C., Roncagliolo, S., Rusu, V., & Collazos, C. (2011). A Methodology to Establish Usability Heuristics. *Proceedings ACHI 2011*, 59–62.

Rusu, V., Rusu, C., Quiñones, D., Botella, F., Roncagliolo, S., & Rusu, V. Z. (2019). On-Line Travel Agencies' Usability: Evaluator eXperience. *International Conference on Human-Computer Interaction*, 452–463.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. SAGE Publications.

Salgado, A., & Freire, A. P. (2014). Heuristic Evaluation of Mobile Usability: A Mapping Study. *International Conference on Human-Computer Interaction*, 178–188. https://doi.org/10.1007/978-3-642-39262-7

Salgado, A., Rodrigues, S., & Fortes, R. (2016). Evolving Heuristic Evaluation for multiple contexts and audiences. *34th ACM International Conference on the Design of Communication*, 19.

Santhipriya, S., Sastry, B. V. S. S. R. S., & Akshitha, K. (2011). Securing smartphone apps in an online environment. *International Journal of Mathematical Archive*, *2*(11-Nov), 2462–2469.

Satia, G. (2014). *Mobile Onboarding: A Beginner's Guide*. http://www.smashingmagazine.com/2014/08/11/mobile-onboarding-beginners-guide/

Satyanarayanan, M., Gao, W., & Lucia, B. (2019). The computing landscape of the 21st century. *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*, 45–50.

Sauro, J., & Lewis, J. (2020). *Should all scale points be labelled?* https://measuringu.com/scale-points-labeled/

Sauro, Jeff. (2013). *Rating The Severity Of Usability Problems*. http://www.measuringu.com/blog/rating-severity.php

Savio, N., & Braiterman, J. (2007). Design Sketch : The Context of Mobile Interaction. *Mobile HCI 2007*, 5–7.

Schutt, R. (2014). *Investigating the Social World: The Process and Practice of Research*. SAGE Publications.

Schwaber, K., & Beedle, M. (2002). *Agile software development with Scrum* (First Volu). Prentice Hall.

Schwarz, N., Knäuper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, *55*(4), 570–582.

Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, *17*(2), 151–161.

Scriven, M. (2003). Evaluation theory and metatheory. *International Handbook of Educational Evaluation*, 15–30.

Scriven, M. (2005). *The logic and methodology of checklists* (Issue October). Western Michigan University.

Sedgwick, P. (2012). Pearson's correlation coefficient. *BMJ*, 345, e4483.

Sharpe, H., Rogers, Y., & Preece, J. (2007). *Interaction design: beyond human-computer interaction* (2nd ed.). John Wiley & Sons.

Shneiderman, B. (1983). *Eight Golden Rules of Interface Design*.

Silva, C. M., Macedo, V., Lemos, R., & Okimoto, M. L. L. R. (2014). Evaluating Quality and Usability of the User Interface : A Practical Study on Comparing Methods with and without Users. In A. Marcus (Ed.), *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience* (pp. 318–328). Springer

International Publishing.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software (No. MTR-10090)*.

Sohaib, O., & Khan, K. (2011). Incorporating discount usability in extreme programming. *International Journal of Software Engineering and Its Applications*, *5*(1), 51–62.

Sonnenberg, C. (2020). Mobile Media Usability: Evaluation of Methods for Adaptation and User Engagement. *Journal of Media Management and Entrepreneurship (JMME)*, *2*(1), 86–107.

Speicher, M. (2015). *What is Usability? A characterization based on ISO 9241-11 and ISO/IEC 25010* (Issue October).

Spradley, J. P. (2016). *Participant observation*. Waveland Press.

Stewart, D. W., Shamdasani, P. N., & Rook, D. W. (2009). Group depth interviews: Focus group research. In *The SAGE handbook of applied social research methods* (pp. 589–616). SAGE Publications.

Stone, D., Jarrett, C., Woodroffe, M., & Minocha, S. (2005). *User Interface Design and Evaluation*. Morgan Kaufmann.

Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. SAGE Publications.

Stringer, E. T. (2013). *Action research*. SAGE Publications.

Stufflebeam, D. L. (2000a). *Guidelines for developing evaluation checklists: The Checkllists Development Checklist (CDC)* (Issue July).

Stufflebeam, D. L. (2000b). The CIPP model for evaluation. *Evaluation Models*, 279–317.

Stufflebeam, D. L. (2000c). The Ten Commandments , Constitutional Amendments , and Other Evaluation Checklists by. *Evaluation 2000 Annual Meeting of the American Evaluation Association*.

Stufflebeam, D. L. (2001). *Evaluation Models* (Issue 89).

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation Theory, Models, and Applications*. John Wiley & Sons.

Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. In *How to Choose a Sampling Technique for Research*.

Tan, C., Silva, A., Lee, R., Wang, K., & Nah, F. F. (2016). HCI Testing in Laboratory or Field Settings. *International Conference on HCI in Business, Government and Organizations*, 110–116. https://doi.org/10.1007/978-3-319-39399-5

Tan, Z., & Lindberg, B. (2010). Speech Recognition on Mobile Devices. *Mobile Multimedia Processing*, 221–237. http://www.springerlink.com/index/1765187424833114.pdf%5Cnpapers2://publication/uuid/9F4325B6-F001-49C1-A8A3-5E1165B6B692

Taylor, K., & Silver, L. (2019). *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally In emerging economies, technology use still much more common among young people and the well-educated*. Pew Research. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

The 2017 U.S. Mobile App Report. (2017). https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/The-2017-US-Mobile-App-Report

Thitichaimongkhol, K., & Senivongse, T. (2016). Enhancing usability heuristics for android applications on mobile devices. *Proceedings of the World Congress on Engineering and Computer Science*, 19–21.

Tognazzini, B. (2003). AskTog: First Principles of Interaction Design. In *AskTog* (pp. 1–6). https://asktog.com/atc/principles-of-interaction-design/

Tokárová, L., & Weideman, M. (2013). Understanding the process of learning touch-screen mobile applications. *Proceedings of the 31st ACM International Conference on Design of Communication - SIGDOC '13*, 157–164. https://doi.org/10.1145/2507065.2507066

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, *16*(10), 837–851.

Tripadvisor Android App. (n.d.). https://play.google.com/store/apps/details?id=com.tripadvisor.tripadvisor&hl=en_US

Trivedi, M. C. (2012). Role Of Context In Usability Evaluations: A Review. *Advanced Computing: An International Journal*, *3*(2), 69–78. https://doi.org/10.5121/acij.2012.3208

Trochim, W. (1998). An evaluation of Michael Scriven's "Minimalist theory: The least theory that practice requires." *American Journal of Evaluation*, *19*(2), 243–249.

Tsiaousis, A. S. (2015). Ad hoc context of use in Mobile Usability Evaluation. *International Journal of Computer Systems*, *02*(06), 253–256.

Use gestures to navigate your iPhone X and later. (n.d.). Use gestures to navigate your iPhone X and later

User Experience Professionals Association. (n.d.). http://uxpaboston.org/

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*(4).

Varsaluoma, J. (2009). Scenarios in the heuristic evaluation of mobile devices: Emphasizing the context of use. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5619 LNCS*, 332–341. https://doi.org/10.1007/978-3-642-02806-9_38

Ventresca, M. J., & Mohr, J. W. (2017). Archival research methods. In *The Blackwell companion to organizations* (pp. 805–828).

Vermeeren, A., Law, E., & Roto, V. (2010). User experience evaluation methods: current state and development needs. *Proceedings: NordiCHI 2010*, 521–530. https://doi.org/10.1145/1868914.1868973

Wale-kolade, A. Y., & Nielsen, P. A. (2014). Integrating Usability Practices into Agile Development : A Case Study. *23RD INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS DEVELOPMENT (ISD2014*, 337–347.

Waqar, W., Chen, Y., & Vardy, A. (2011). Exploiting Smartphone Sensors for Indoor Positioning : A Survey. *Proceedings of the Newfoundland Conference on Electrical and Computer Engineering.*

Weichbroth, P. (2020). Usability of mobile applications: a systematic literature study. *IEEE Access*, *8*, 55563–55577.

Weiss, S. (2003). *Handheld usability*. John Wiley and Sons. https://doi.org/10.1145/778712.778775

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). *The Cognitive Walkthrough Method: A Practitioner's Guide*.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. *Breakthroughs in Statistics*, 196–202.

Williams, C. (2007). Research methods. *Journal of Business & Economics Research (JBER)*, *5*(3).

Williamson, J., & Brewster, S. (2012). A performative perspective on UX. *Communications in Mobile Computing*, *1*(3).

Wilson, C. (2013). *User Interface Inspection Methods: A User-Centered Design Method* (Vol. 15). Newnes. https://books.google.com/books?id=LQ0wAAAAQBAJ&pgis=1

Winter, M. (2011). *Scan Me - Everybody's Guide to the Magical World of Qr Codes*. Westsong Publishing.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth ed.). SAGE Publications.

Yon, K., Nettleton, S., Walters, K., Lamahewa, K., & Buszewicz, M. (2015). Junior doctors' experiences of managing patients with medically unexplained symptoms: a qualitative study. *BMJ Open*, *5*(12).

Yu, J., Sheng, Q. Z., Younas, M., & Shakshuki, E. (2013). Advances in context-aware mobile services. *Personal and Ubiquitous Computing*, *18*(5), 1027–1028. https://doi.org/10.1007/s00779-013-0716-5

Zhang, D., & Adipat, B. (2005). Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. *International Journal of Human-Computer Interaction*, *18*(3), 293–308. https://doi.org/10.1207/s15327590ijhc1803_3

Zhang, D., Boonlit, A., & Yaser, M. (2009). User-Centered Context-Aware Mobile Applications—The Next Generation of Personal Mobile Computing. *Communications of the Association for Information Systems*, *24*(3). https://doi.org/10.17705/1CAIS.02403

APPENDIX A: GLOSSARY

This glossary defines how different terms are used in the context of this work:

Accelerometer: Used to measure acceleration.

Agile software development life cycle: An iterative software development approach that aims to deliver software faster to customer.

Amazon, eBay, Macy's, Target, CVS, Wayfair, Myntra: Companies focused on eCommerce that offer mobile applications.

Android: A mobile device operating system from Google.

Apple iOS: Mobile device operating system from Apple Inc.

Bumble and Hing: Companies focused on dating that offer mobile applications.

Context-aware applications: Mobile applications that utilise sensors and are thus aware of their immediate contexts, such as if the user is moving, if the mobile device is being used in bright or dark conditions, if the ambient noise is high etc.

Customer Effort Score (CES): A questionnaire used to discover a customer's impression of working with an organization.

DuckDuckGo: A web browser that is deemed to protect privacy more than Google Chrome or Apple Safari and other similar web browsers.

Eye tracking: Devices with the ability to track eye movements and record where a user is looking and how long their gaze stays fixed on any particular object or part of a screen.

Feature phone: A type of mobile phone that allows users to browse the internet and play music, but without the advanced capabilities of smartphones.

Foursquare, Pinterest, Instagram, Twitter, Facebook, Medium, Reddit, LinkedIn: Popular social media websites that are focused on user-generated content (UGC).

Global positioning system (GPS): A worldwide system of satellites that can determine the latitude and longitude of an object, commonly used to receive directions when driving to an unfamiliar location.

Goal-Oriented Context-Aware Measurement and Evaluation (GOCAME): A system for software quality assurance.

GoKitchen, Hello Fresh, NY Times Cooking: Companies focused on food preparation that offer mobile applications.

GrubHub and Seemless: Companies focused on food delivery that offer mobile applications.

International Organization for Standardisation (ISO): An international organisation operating in 164 countries that focuses on defining International standards.

Net Promoter Score (NPS): A single question method that aims to measure a customer's loyalty to an organisation.

Netflix, Xfinity, IMDB, ESPN, and Amazon Prime Video: Companies focused on entertainment that offer mobile applications.

Nvivo: Software used for the analysis of qualitative data.

Pop-socket: A device that can be attached to the back of a mobile device that offers more comfort and stability than holding a device normally.

Qualtrics: A company based in Salt Lake City in Utah in the United States focused on the creation of powerful questionnaires that offers a mobile application.

Quick Response (QR) code: A code in a square format allowing users to quickly open a webpage, copy an email address for late use, download an application and so on.

Runkeeper, Nike run club, Rappa, FitBit, MapMyRun: Companies focused on health that offer mobile applications.

Single Ease Question (SEQ): A single question used to determine the difficult of a task during usability testing.

Slack and HipChat: Companies focused on communication that offer mobile applications.

Smart phone: An advanced mobile device allowing users to browse the internet, play music and download mobile applications.

Spotify and Pandora: Companies focused on music that offer mobile applications.

System Usability Scale (SUS): A method made up of ten questions used to measure the usability of a system.

Tripadvisor, United Airlines, Kayak, and Yelp: Companies focused on travel that offer mobile applications. Yelp also focuses on finding nearby businesses and restaurants.

Uber and Lyft: Companies focused on food preparation that offer mobile applications.

User Experience Professionals Association (UXPA): A non-profit organization focused on delivering content and organising talks and conferences for the User Experience (UX) community.

User interface (UI): The screen of a computer or mobile device.

Venmo, Chase, Bank of America, Venmo, Saffire, Mint, Citizens Bank, DCU, MyBluebird, Avidia, Capital One, Splitwise, Apple Stocks application, Yahoo finance: Companies focused on finance that offer mobile applications.

Waze: A company acquired by Google focused on driving directions that offers a mobile application.

Zoom: A cloud-based enterprise video conferencing platform.

APPENDIX B: ETHICS APPROVALS

UNIVERSITY OF HERTFORDSHIRE
SCIENCE & TECHNOLOGY

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Marianna Lilly
FROM	Dr Simon Trainis, Science and Technology ECDA Chairman
DATE	16/4/15

Protocol number: COM/PG/UH/00089

Title of study: Evaluation of smartphone application heuristics (Phase 2)

Your application for ethical approval has been accepted and approved by the ECDA for your school.

This approval is valid:

From: 15/5/15

To: 15/5/15

Please note:

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.

Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.

Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Students must include this Approval Notification with their submission.

UNIVERSITY OF HERTFORDSHIRE
SCIENCE & TECHNOLOGY

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Dr Mariana Lilley
FROM	Dr Simon Trainis, Science and Technology ECDA Chairman
DATE	17/8/15

Protocol number: COM/PG/UH/00107

Title of study: Evaluation of smartphone application heuristics (phase 3)

Your application for ethics approval has been accepted and approved by the ECDA for your school.

This approval is valid:

From: 1/9/15

To: 31/1/16

Please note:

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.

Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.

Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Students must include this Approval Notification with their submission.

University of Hertfordshire UH

UNIVERSITY OF HERTFORDSHIRE
SCIENCE AND TECHNOLOGY

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Dr Mariana Lilley
FROM	Dr Simon Trainis, Science and Technology ECDA Chairman
DATE	29/01/2016

Protocol number: **aCOM/PG/UH/00107(1)**

Title of study: Evaluation of smartphone application heuristics (Phase 3)

Your application to extend the existing protocol COM/PG/UH/00107 as detailed below has been accepted and approved by the ECDA for your School.

This approval is valid:

From: 31/01/2016

To: 31/03/2016

Please note:

Any conditions relating to the original protocol approval remain and must be complied with.

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1 or as detailed in the EC2 request. Should you amend any further aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit a further EC2 request. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.

Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.

Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Students must include this Approval Notification with their submission.

UNIVERSITY OF HERTFORDSHIRE
SCIENCE & TECHNOLOGY

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Marianna Lilley
FROM	Dr Simon Trainis, Science and Technology ECDA Chairman
DATE	30/01/15

Protocol number: COM/PG/UH/00084

Title of study: Evaluation of smartphone application heuristics (Phase 1)

Your application for ethical approval has been accepted and approved by the ECDA for your school.

This approval is valid:

From: 01/02/15

To: 01/08/15

Please note:

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.

Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.

Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Students must include this Approval Notification with their submission.

UNIVERSITY OF HERTFORDSHIRE
FACULTY OF SCIENCE, TECHNOLOGY AND CREATIVE ARTS

M E M O R A N D U M

TO	**Ger Joyce**
CC	**Mariana Lilley**
FROM	**Dr Simon Trainis – Chair, Faculty Ethics Committee**
DATE	**2 August 2012**

Your Ethics application for your project entitled:

An investigation of heuristics for the evaluation of mobile applications

has been granted approval and assigned the following Protocol Number:

1112/299

This approval is valid:

From 15 October 2012

Until 15 February 2013

If it is possible that the project may continue after the end of this period, you will need to resubmit an application in time to allow the case to be considered.

University of Hertfordshire UH

HEALTH SCIENCE ENGINEERING & TECHNOLOGY ECDA

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Dr Mariana Lilley
FROM	Dr Simon Trainis, Health, Sciences, Engineering & Technology ECDA Chair
DATE	13th December 2017

Protocol number: COM/PGR/UH/03021

Title of study: *Contextual Usability Evaluation*

Your application for ethics approval has been accepted and approved by the ECDA for your School and includes work undertaken for this study by the named additional workers below:

This approval is valid:

From: 13/12/2017

To: 15/05/2018

Additional workers: No additional workers named

Please note:

If your research involves invasive procedures you are required to complete and submit an EC7 Protocol Monitoring Form, and your completed consent paperwork to this ECDA once your study is complete. You are also required to complete and submit an EC7 Protocol Monitoring Form if you are a member of staff. This form is available via the Ethics Approval StudyNet Site via the 'Application Forms' page http://www.studynet1.herts.ac.uk/ptl/common/ethics.nsf/Teaching+Documents?Openview&count=9999&restricttocategory=Application+Forms

Any necessary permissions for the use of premises/location and accessing participants for your study must be obtained in writing prior to any data collection commencing. Failure to obtain adequate permissions may be considered a breach of this protocol.

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1A. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval (if you are a student) and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1A may need to be completed prior to the study being undertaken.

University of Hertfordshire UH

HEALTH SCIENCE ENGINEERING & TECHNOLOGY ECDA

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Mariana Lilley
FROM	Dr Simon Trainis, Health, Science, Engineering & Technology ECDA Chair.
DATE	18/05/2018

Protocol number: **aCOM/PGR/UH/03021(1)**

Title of study: Contextual Usability Evaluation

Your application to modify and extend the existing protocol as detailed below has been accepted and approved by the ECDA for your School and includes work undertaken for this study by the named additional workers below:

Modification: as stated in the EC2 Form.

This approval is valid:

From: 18/05/2018

To: 15/08/2018

Additional workers: Mariana Lilley

Please note:

If your research involves invasive procedures you are required to complete and submit an EC7 Protocol Monitoring Form, and your completed consent paperwork to this ECDA once your study is complete. You are also required to complete and submit an EC7 Protocol Monitoring Form if you are a member of staff. This form is available via the Ethics Approval StudyNet Site via the 'Application Forms' page http://www.studynet1.herts.ac.uk/ptl/common/ethics.nsf/Teaching+Documents?Openview&count=9999&restricttocategory=Application+Forms

Any conditions relating to the original protocol approval remain and must be complied with.

Any necessary permissions for the use of premises/location and accessing participants for your study must be obtained in writing prior to any data collection commencing. Failure to obtain adequate permissions may be considered a breach of this protocol.

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1/EC1A or as detailed in the EC2 request. Should you amend any further aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval (if you are a student) and must complete and

University of Hertfordshire UH

HEALTH, SCIENCE, ENGINEERING AND TECHNOLOGY ECDA

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Dr Mariana Lilley
FROM	Dr Simon Trains, Health, Science, Engineering 7 Technology ECDA Chair.
DATE	12/08/2019

Protocol number:	**ECS/PGR/UH/03864**
Title of study:	Evaluating the Contextual Usability Evaluation protocol.

Your application for ethics approval has been accepted and approved with the following conditions by the ECDA for your School and includes work undertaken for this study by the named additional workers below:

no additional workers named.

General conditions of approval:

Ethics approval has been granted subject to the standard conditions below:

Permissions: Any necessary permissions for the use of premises/location and accessing participants for your study must be obtained in writing prior to any data collection commencing. Failure to obtain adequate permissions may be considered a breach of this protocol.

External communications: Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Invasive procedures: If your research involves invasive procedures you are required to complete and submit an EC7 Protocol Monitoring Form, and copies of your completed consent paperwork to this ECDA once your study is complete.

Submission: Students must include this Approval Notification with their submission.

Validity:

This approval is valid:

From: 12/08/2019

To: 24/12/2019

HEALTH, SCIENCE, ENGINEERING AND TECHNOLOGY ECDA

ETHICS APPROVAL NOTIFICATION

TO	Ger Joyce
CC	Dr Mariana Lilley
FROM	Dr Simon Trainis, Health, Science, Engineering & Technology ECDA Chair
DATE	01/12/2020

Protocol number: **SPECS/PGR/UH/04311**

Title of study: Contextual usability evaluation scope determination

Your application for ethics approval has been accepted and approved with the following conditions by the ECDA for your School and includes work undertaken for this study by the named additional workers below:

no additional workers named

General conditions of approval:

Ethics approval has been granted subject to the standard conditions below:

Permissions: Any necessary permissions for the use of premises/location and accessing participants for your study must be obtained in writing prior to any data collection commencing. Failure to obtain adequate permissions may be considered a breach of this protocol.

External communications: Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Invasive procedures: If your research involves invasive procedures you are required to complete and submit an EC7 Protocol Monitoring Form, and copies of your completed consent paperwork to this ECDA once your study is complete.

Submission: Students must include this Approval Notification with their submission.

Validity:

This approval is valid:

From: 01/12/2020

To: 29/01/2021

APPENDIX C: MOBILE CHARACTERISTICS

The summarised list of mobile characteristics from the second step of Rusu et al.'s (2011) approach were:

- Users' awareness of the current mobile application status
- Ensuring that mobile platform consistency, conventions, and standards meet the users' real-world mental model
- Graceful management of mobile application input errors
- Introducing users to a new mobile application
- Having a user interface that has a clean and simple presentation, while being focused and aesthetically pleasing
- A mobile application that is easy to understand, easy to learn, and easy to use
- A mobile application with clear indications of what to do next in order to successfully complete a task
- Clear, well-written content that is easy to read
- The ability to use the mobile application or parts of the mobile application offline, where it makes sense to do so
- The ability for the mobile application to be used as part of an ecosystem with other devices, such as laptops, TV's, and wearables
- The ability to define shortcuts, configure, customize, and personalize a mobile application
- Consideration of the users', potentially complex, context of use, such as the environment they are in when using the mobile application, including

inevitable interruptions and the ability to glean important information at a glance

- The ability to quickly and easily enter data using a variety of methods
- Data entry techniques that pass the burden to the system, not to the user, via a full utilization of the camera and sensors on the mobile device as needed
- With so many mobile applications being installed, a memorable application icon would allow a user to find the mobile application quickly and easily
- The need to consider varying screen sizes, generally from 4” to 7”, along with varying degrees of display resolution
- Consideration of slow, variable, and potentially non-existent network connectivity
- Limited storage capacity and processing power, especially on older mobile devices
- Text and other user interface elements that are the right size in portrait and landscape modes
- Tap targets that are also the right size and not too close together
- Content with whitespace that is not wider than the screen
- A correctly set viewport
- Ensure that the mobile application rarely crashes or causes problems
- Trusting the mobile application with personal information
- Convincing the user about the value proposition of the mobile application
- Ensuring that only the features needed are part of the mobile application
- Gating the mobile application by requiring an email address, social login or too much information upfront may spark disinterest unless there is a value

add to the user, such as personalizing the experience or accessing additional features

- Ads that relevant and not too frequent nor intrusive
- Ensuring that mobile applications do not cause excessive battery drain
- Offering the mobile application user that a mobile app's creators value privacy
- Keeping push notifications only to those that are relevant and important
- Understanding the importance of localization into other languages, currencies, date formats etc.
- Information that it passed to other screens to ensure that the memory load of the mobile applications user is reduced
- While only indirectly related to mobile applications, ensuring that the mobile device is built good ergonomics, and a minimalist design, and also that it is easily found if lost

APPENDIX D: QUESTIONNAIRE TOOL

The questionnaire tool created to assist in gathering the attitudes of those knowledgeable in HCI toward the mobile application heuristics is detailed below:

Key characteristics of users: The key characteristics of the website users are:

- Knowledge of Human-Computer Interaction, such as HCI practitioners, researchers or educators, not necessarily in the mobile domain given the relative recency of the domain;
- Experienced in the area of usability evaluations.

Functional Requirements: The Web 2.0 Application needed to:

- Greet participants and offer them a brief introduction to the project;
- Allow participants to give details about their role, how long they are in that role and the country they are based in;
- Knowing how busy participants might be, display only a summarized version of research results to explain how the initial set of mobile application heuristics were defined;
- Allow participants to rate the usefulness of each heuristic on a 5-point Likert scale denoting 1 as 'Not Useful' and 5 as 'Very Useful';
- Allow participants the opportunity to leave comments, which may inform their thought process on why they rated heuristics the way they did and/or offer ideas on heuristics that were missed.

Non-Functional Requirements: The Web 2.0 Application needed to:

- Have clear and unambiguous instructions in how to use the application;

- Be easy to use, enabling participants to rate heuristics quickly;
- Follow both traditional heuristics and Web Content Accessibility Guidelines as much as feasible;
- Be aesthetically pleasing.

Application Design: Microsoft Visual Web Developer 2010 Express with ASP.NET, XHTML, CSS and SQL are used for the front end, while Microsoft Access was chosen for the back-end database. Website pages were created with a wizard-style approach, which in effect walked participants through a series of defined steps:

1) Quickly read a brief introduction;
2) Offer information about their role, how long they have been in that role, and the country they were based in;
3) Read more detailed information about the project to give a sense of perspective as to why the research was required and how the initial set of heuristics was defined;
4) Rate each mobile application heuristic on a Liberty scale from Not Useful to Very Useful;
5) Optionally add comments about the heuristics while the heuristics were visible;
6) Offer any extra comments they might have had, not necessarily related to the heuristics, such as they thoughts on the survey, their experience etc.;
7) Thank the participants and offer each the chance to enter a prize draw for one of two Amazon vouchers in appreciation for their time. The participants could elect to exit the survey without being included in the prize draw.

Database Design

To store data from participants as they participated in the questionnaire, the following relational database schema was designed:

- **Evaluators** (<u>EvaluatorID: *Integer*</u>, EvaluatorRole: *String*, EvaluatorOtherRole: *String*, EvaluatorRoleLength: *String*, EvaluatorCountry: *String*, EvaluatorName: *String*, EvaluatorEmail: *String*)

- **Feedback** (<u>FeedbackID: *Integer*</u>, EvaluatorID: *Integer**, HeuristicRating1: *Integer*, HeuristicRating2: *Integer*, HeuristicRating3: *Integer*, HeuristicRating4: *Integer*, HeuristicRating5: *Integer*, HeuristicRating6: *Integer*, HeuristicRating7: *Integer*, HeuristicRating8: *Integer*, HeuristicRating9: *Integer*, HeuristicRating10: *Integer*, HeuristicRating11: *Integer*, HeuristicComments: *Memo*, OtherComments: *Memo*, ReviewCreated: *Date/Time*)

APPENDIX E: LIST OF COUNTRIES

Sixty participants (n=60) from eighteen countries reviewed the initial set of mobile application heuristics:

Table E.1 Countries where participants were based (n=60)

Country	Number of participants
United States	10
United Kingdom	8
Spain	7
Germany	5
Finland	5
Australia	4
Italy	3
Austria	3
Switzerland	2
Portugal	2
Canada	2
Mexico	2
Denmark	2
China	1
Korea (Rep. of)	1
Norway	1
Taiwan	1
Jordan	1

APPENDIX F: LIST OF MOBILE DEVICE AND NON-DEVICE VARIABLES

The mobile device and non-device variables for the study were:

- **Mobile device:** LG G2 mobile device running Android 4.4.2
- **Mobile application:** TripAdvisor, version 9.6.1 build 90060034. Updates on the mobile device were turned off to ensure that all participants used the same version of the TripAdvisor mobile application despite the study being conducted over several weeks.
- **Environment:** Specify in the instructions that evaluating the usability of a native mobile application should be done under good lighting and quiet conditions. The data collection form need not ask if these conditions were met. In your instructions all non-device variables should be controlled as much as possible.
- **Experience:** Ask for HCI and Mobile HCI experience within the demographics section of the survey. Consider comparing only HCI practitioners, educators and researchers that have experience against other sets of heuristics, then compare those with and without experience with the SMART Heuristics only.
- **Fatigue:** Participants will be asked if they wished to take a break between using the different sets of heuristics.
- **Ambiguity:** The exact same instructions will be given to all participants prior to the evaluation. These will specify the scenario(s), tasks and questions to be used during the evaluation.
- **Learning Effect:** Different combinations of heuristics will be defined, and participants will be randomly assigned. All possible combinations will be covered and randomly assigned such that there are no gaps.

APPENDIX G: LIST OF HEURISTICS

Set A: Nielsen's (1994) 10 Heuristics for User Interface Design

Visibility of system status
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

Match between system and the real world
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

User control and freedom
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Consistency and standards
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

Error prevention
Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

Recognition rather than recall
Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Flexibility and efficiency of use
Accelerators — unseen by the novice user — may often speed up the interaction for

the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

Aesthetic and minimalist design
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

Help users recognize, diagnose, and recover from errors
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Help and documentation
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Set B: Bertini et al. (2006) - Heuristic evaluation for mobile computing

Heuristic 1 - Visibility of system status and losability/findability of the mobile device: Through the mobile device, the system should always keep users informed about what is going on. Moreover, the system should prioritize messages regarding critical and con-textual information such as battery status, network status, environmental conditions, etc. Since mobile devices often get lost, adequate measures such as encryption of the data should be taken to minimize loss. If the device is misplaced, the device, system or application should make it easy to find it back.

Heuristic 2 - Match between system and the real world: Enable the mobile user to interpret correctly the information provided, by making it appear in a natural and logical order; whenever possible, the system should have the capability to sense its environment and adapt the presentation of information accordingly.

Heuristic 3 - Consistency and mapping: The user's conceptual model of the possible function/interaction with the mobile device or system should be consistent with the context. It is especially crucial that there be a consistent mapping between user actions/interactions (on the device buttons and controls) and the corresponding real tasks (e.g. navigation in the real world).

Heuristic 4 - Good ergonomics and minimalist design: Mobile devices should be easy and comfortable to hold/ carry along as well as robust to damage (from environmental agents). Also, since screen real estate is a scarce resource, use it with parsimony. Dialogues should not contain information which is irrelevant or rarely needed.

Heuristic 5 - Ease of input, screen readability and glancability: Mobile systems should provide easy ways to input data, possibly reducing or avoiding the need for the user to use both hands. Screen content should be easy to read and navigate through not withstanding different light conditions. Ideally, the mobile user should be able to quickly get the crucial information from the system by glancing at it.

Heuristic 6 - Flexibility, efficiency of use and personalization: Allow mobile users to tailor/personalize frequent actions, as well as to dynamically configure the system according to contextual needs. Whenever possible, the system should support and suggest system-based customization if such would be crucial or beneficial.

Heuristic 7 - Aesthetic, privacy and social conventions: Take aesthetic and emotional aspects of the mobile device and system use into account. Make sure that user's data are kept private and safe. Mobile inter-action with the system should be comfortable and respectful of social conventions.

Heuristic 8 - Realistic error management: Shield mobile users from errors. When an error occurs, help users to recognize, to diagnose, if possible to recover from the error. Mobile computing error messages should be plain and precise. Constructively suggest a solution (which could also include hints, appropriate FAQs, etc). If there is no solution to the error or if the error would have negligible effect, enable the user to grace-fully cope with the error.

Set C: Joyce & Lilley SMART Heuristics for mobile applications

Provide immediate notification of application status. Ensure the mobile application user is informed of the application status immediately and as long as is necessary. Where appropriate do this non-intrusively, such as displaying notifications within the status bar.

Use a theme and consistent terms, as well as conventions and standards familiar to the user. Use a theme for the mobile application to ensure different screens are consistent. Also create a style guide from which words, phrases and concepts familiar to the user will be applied consistently throughout the interface, using a natural and logical order. Use platform conventions and standards that users have come to expect in a mobile application such as the same effects when gestures are used.

Prevent problems where possible; Assist users should a problem occur. Ensure the mobile application is error-proofed as much as is possible. Should a problem occur, let the user know what the problem is in a way they will understand, and offer advice in how they might fix the issue or otherwise proceed. This includes problems with the mobile network connection, whereby the application might work offline until the network connection has been re-established.

Display an overlay pointing out the main features when appropriate or requested. An overlay pointing out the main features and how to interact with the application allows first-time users to get up-and-running quickly, after which they can explore the mobile application at their leisure. This overlay or a form of help system should also be displayed when requested.

Each interface should focus on one task. Being focusing on one task ensures that mobile interfaces are less cluttered and simple to the point of only having the absolute necessary elements onscreen to complete that task. This also allows the interface to be glanceable to users that are interrupted frequently.

Design a visually pleasing interface. Mobile interfaces that are attractive are far more memorable and are therefore used more often. Users are also more forgiving of attractive interfaces.

Intuitive interfaces make for easier user journeys. Mobile interfaces should be easy to learn whereby next steps are obvious. This allows users to more easily complete their tasks.

Design a clear navigable path to task completion. Users should be able to see right away how they can interact with the application and navigate their way to task completion.

Allow configuration options and shortcuts. Depending on the target user, the mobile application might allow configuration options and shortcuts to the most important information and frequent tasks, including the ability to configure according to contextual needs.

Cater for diverse mobile environments. Diverse environments consist of different types of context of use such as poor lighting conditions and high ambient noise are common issues that mobile users have to face every day. While the operating system should allow the user to change the interface brightness and sound settings, developers can assist users even more for example by allowing them to display larger buttons and allowing multimodal input and output options.

Facilitate easier input. Mobile devices are difficult to use from a content input perspective. Ensure users can input content more easily and accurately by, for instance displaying keyboard buttons that are as large as possible, as well as allowing multimodal input and by keeping form fields to a minimum.

Use the camera, microphone and sensors when appropriate to lessen the user's workload. Consider the use of the camera, microphone and sensors to lessen the users' workload. For instance, by using GPS so the user knows where they are and how to get where they need to go, or by using OCR and the camera to digitally capture the information the user needs to input, or by allowing use of the microphone to input content.

APPENDIX H: APPROACHES USED TO CONSIDER CONTEXT OF USE

Types of tools and methods used to consider context of use:

Table H.1. Tools and methods used to consider context of use when designing and evaluating the usability of mobile applications

Tool/Method	Number of instances within responses
Ethnographic research	51
Usability testing (total responses/responses that mentioned in-context usability testing)	39/9
Interviews	33
Sketching/Prototyping	18
User stories/Use cases	16
Questionnaires	14
Diary studies	13
Focus groups	9
Market research	8
Personas	8
Context-aware-sensors/GOCAME	7
Making assumptions	6
Heuristic evaluation	6
Photos/Videos/Screen capture	6
User feedback/Reviews	4
Guerrilla testing	3
Guidelines/Standards	2
Co-design	1
Design thinking	1
Emotional Response Testing	1
Eye tracking	1
Literature review	1
Mental models	1
Social media	1
AEIOU	1

APPENDIX I: LIST OF CONTEXTUAL USABILITY EVALUATION USER STORIES

A full list of thirty-four contextual usability evaluation user stories that originated from the interviews:

1) When writing an iMessage, does walking on a busy street, have an impact on the perception that the mobile application is tolerant of errors?
2) When reading messages quickly on WhatsApp, does being with friends or family at a social event, have an impact on the perception that the mobile application feels responsive?
3) When writing an email and attaching a photo, does walking from one meeting to another at work, have an impact on the task being completed quickly & easily?
4) When sending a Slack message with a video attachment, does keeping an eye out for your shuttle bus, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?
5) When reviewing upcoming events on Fantastical Calendar, does walking from one meeting to another at work, have an impact on readability of all screen elements in portrait mode?
6) When checking the weather for the day on Yahoo weather, does being sleepy having just woke up, have an impact on perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?
7) When quickly reading the latest posts on LinkedIn, does attempting to meet an upcoming deadline at work, have an impact on readability?

8) When looking for inspiring things to do nearby on Instagram, does being in an unfamiliar location while on holiday, have an impact on the perception that the mobile application is personalized to suit the user?
9) When posting to Twitter, does standing on a busy underground train while holding an overhead rail, have an impact on the perception that the mobile application is tolerant of errors?
10) When requesting an Uber ride, does keeping an eye on your surroundings while on a dark street in the city, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?
11) When finding directions to a location using Waze, does having the phone display dimmed due to a low battery, have an impact on readability of all screen elements in landscape mode?
12) When navigating to a destination using Google Maps, does using voice commands while driving, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?
13) When finding and purchasing a flight with Kayak, does having the TV on in the background, have an impact on the perception that the mobile application is easy to learn?
14) When searching for open, inexpensive nearby Italian restaurants on Yelp, does walking in an unfamiliar neighbourhood, have an impact on the task being completed quickly & easily?
15) When finding and purchasing a product on Amazon, does moving to a laptop from a mobile device, have an impact on the ability to continue a task on another device that was started on the mobile app?
16) When finding directions to Experiences in your saved list on Tripadvisor, does not having Wi-Fi or an International data plan on holiday, have an

impact on the perception that all elements, including graphics & text, are consistent & familiar throughout the mobile app?

17) When starting to track a run with Runkeeper, does having touch-sensitive gloves on in cold weather, have an impact on the task being completed quickly & easily?

18) When casting a movie from Netflix to a smart TV, does making dinner, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?

19) When changing playlists on Spotify, does walking while holding a bag of groceries, have an impact on the task being completed quickly & easily?

20) When looking for books about Leadership on Audible, does sitting on a constantly moving/stopping shuttle bus in heavy traffic, have an impact on the perception that the mobile application is easy to learn?

21) When sending money using Venmo to someone that just paid for dinner at a restaurant, does being surrounded by people in a noisy environment, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?

22) When checking a stock price with Yahoo finance, does walking in the rain, have an impact on perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?

23) When checking your DCU credit union money market balance, does holding the phone at an angle due to the need for privacy while sitting next to strangers on a bus, have an impact on readability of all screen elements in portrait mode?

24) When ordering food via GrubHub, does having a significant other changing their minds about what they would like to eat, have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?
25) When ordering a latte from Starbucks while redeeming stars, does walking quickly to catch a train to work, have an impact on perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?
26) When reserving a table for 6 at a local restaurant using OpenTable, does getting ready to go out, have an impact on the task being completed quickly & easily?
27) When scrolling through the latest news stories on BBC, does walking in cold weather without gloves, have an impact on readability of all screen elements in portrait mode?
28) When writing a thought on Google Keep, does sitting in an Uber while chatting with the driver, have an impact on the task being completed quickly & easily?
29) When reading a recipe on GoKitchen, does having hands covered in flour, have an impact on perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?
30) When looking for a date on Bumble, does lying in bed while feeling sleepy, have an impact on the perception that the mobile application can be modified, such as personalized, customized or configured, to suit the user?

31) When reading a book on the Kindle mobile app, does being shuffled around on a fast-moving shuttle bus, have an impact on readability of all screen elements in landscape and/or portrait modes?

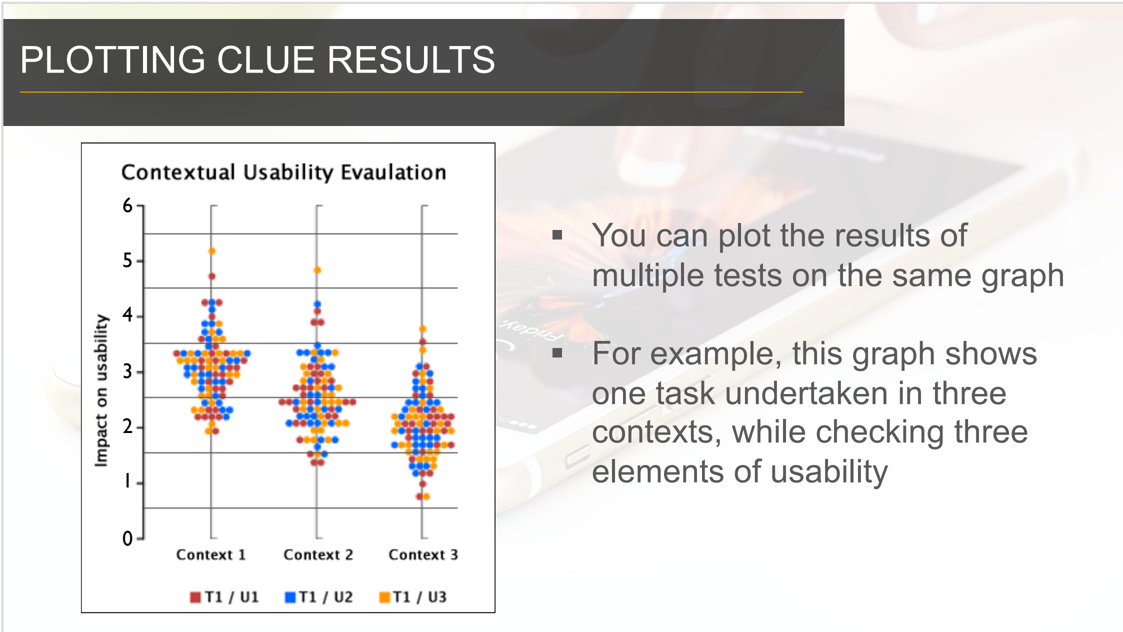
32) When writing a review for a hotel on Tripadvisor, does having a throttled, slow connection due to being over your daily International data limit, have an impact on the perception that the mobile application feels responsive?

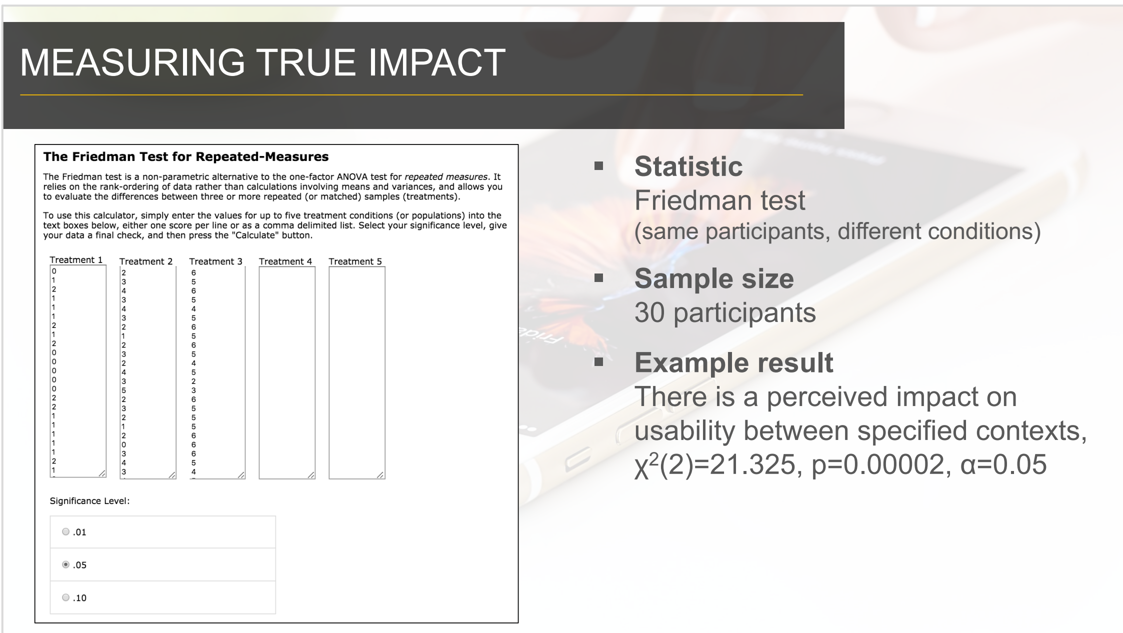
33) When looking for the latest gate information on the United Airlines app, does walking quickly through an airport, have an impact on the task being completed quickly & easily?

34) When looking for a good podcast to listen to, does sitting on a beach in bright sunlight, have an impact on readability of all screen elements in landscape and/or portrait modes?

APPENDIX J: DEMONSTRATION

The statistical analysis using a Friedman test and a visualization of a dot plot, which was demonstrated to HCI practitioners (section 7.2). This evaluation was **hypothetical** and **had not been conducted**, it was included only to offer a practical demonstration of how the protocol might be used by HCI practitioners:





APPENDIX K: CONTRIBUTING PUBLICATIONS

1) *Joyce, G., & Lilley, M. (2014). Towards the development of usability heuristics for native smartphone mobile applications. International Conference of Design, User Experience, and Usability, 465-474, Springer.*

Abstract: This paper reports on initial work in the identification of heuristics that may be most usefully applied in the heuristic evaluation of native smartphone applications. Given the prevalence of such applications, this work seems pertinent, particularly as it also seems under-represented in the literature. Once defined, the heuristics were developed further based on the quantitative and qualitative feedback received from sixty Human-Computer Interaction experts in eighteen countries. The resulting heuristics could be beneficial to HCI researchers and educators and could also potentially expedite and cut the cost of smartphone application usability evaluations for HCI practitioners.

2) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2014). Adapting heuristics for the mobile panorama. Proceedings of Interacción '14, the XV International Conference on Human Computer Interaction, 1-2, ACM.*

Abstract: Expert-based usability inspection methods are well established; the heuristic method in particular is widely known for being fast, relatively inexpensive and easy to learn. However, traditional heuristics are not easily applied to the mobile panorama. This paper is concerned with preliminary work in the design of a set of heuristics that are tailored to the evaluation of native smartphone applications. In this work, Nielsen's original set of heuristics as well as research in the field of usability evaluation for mobile computing were analysed and used to derive a set of heuristics for the evaluation of native smartphone applications.

3) *Joyce, G. (2014). Adaption of usability evaluation methods for native smartphone applications. The 16th International conference on Human-computer interaction with mobile devices & services, 409-410, ACM.*

Abstract: Research has shown that traditional usability evaluation methods cannot be readily applied to the evaluation of native smartphone applications. This research investigates this issue by adapting two usability evaluation methods, applying each at different stages of the design life cycle. Both adapted methods, when combined as a framework, may help in the design of more usable native smartphone applications.

4) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2015). Smartphone application usability evaluation: the applicability of traditional heuristics. International Conference of Design, User Experience, and Usability, 541-550, Springer.*

Abstract: The Heuristic Evaluation method has been popular with HCI experts for over 20 years. Yet, we believe that the set of heuristics defined by Nielsen in 1994 needs to be modified prior to the usability evaluation of smartphone applications. In this paper, we investigate the applicability of each of Nielsen's traditional heuristics to the usability evaluation of smartphone applications following an analysis of 105 peer-reviewed papers. It is anticipated that this work might benefit HCI practitioners, educators and researchers as they attempt to define usability heuristics for smartphone applications. This set of heuristics, once defined, could enable the discovery of usability issues early in the smartphone application development life cycle, while continuing to be a discount usability engineering method as originally defined by Nielsen.

5) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2016). Mobile application usability: heuristic evaluation and evaluation of heuristics. Advances in Human Factors, Software, and Systems Engineering, 77-86, Springer.*

Abstract: Many traditional usability evaluation methods do not consider mobile-specific issues. This can result in mobile applications that abound in usability issues. We empirically evaluate three sets of usability heuristics for use with mobile applications, including a set defined by the authors. While the set of heuristics defined by the authors find more usability issues in a mobile application than other sets of heuristics, improvements to the set can be made.

6) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2016). Evaluating the impact of changing contexts on mobile application usability within agile environments. 2016 Future Technologies Conference (FTC), 476-480, IEEE.*

Abstract: Mobile applications tend to be used in contexts that change over time. These varying contexts may impact the usability, and potentially the overall user experience, of mobile applications. However, the impact of context from a temporal perspective is not fully considered within usability evaluations. Consequently, this work focuses on a conceptual method that attempts to address this limitation. The proposed Contextual Usability Evaluation method promises to allow Human- Computer Interaction experts to evaluate the perceived impact of varying contexts over time on the usability of mobile applications. Despite the focus on context over time, the method is well suited to fast-paced Agile environments.

7) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2016). Mobile application tutorials: Perception of usefulness from an HCI expert perspective.*

International Conference on Human-Computer Interaction, 302–308, Springer.

Abstract: Mobile application tutorials are an opportunity to educate users about a mobile application. Should a mobile application tutorial not be used, the number of frustrated users and uninstalled applications could increase, resulting in a substantial loss in revenue for mobile application developers. Yet, the historical ineffectiveness of printed documentation and online help may have a negative influence on the perception of usefulness of mobile application tutorials for more experienced HCI experts. This in turn may influence their design decisions, whereby they may choose to not design a mobile application tutorial when it may have been better for the user. Our research suggests that while there is a split in the perception of usefulness of mobile application tutorials within the HCI community, the length of time in an HCI role did not have a statistically significant effect on this perception.

8) *Salgado, A., do Amaral, L. A., de Mattos Fortes, R. P., Chagas, M. H. N., & Joyce, G. (2017). Addressing mobile usability and elderly users: Validating contextualized heuristics. International Conference of Design, User Experience, and Usability, 379-394, Springer.*

Abstract: Diverse heuristic sets were proposed in order to evolve Heuristic Evaluation for new contexts, as contexts related to the elderly and mobile devices. However, heuristics for evaluation of mobile usability regarding elderly users still need additional validations. For this reason, our study aimed to enhance the validation of a heuristic set proposed by Al-Razgan et al. for evaluation of mobile usability regarding elderly users. Results showed that the major part of heuristics proposed by Al-Razgan et al. matches with traditional heuristics of Nielsen, while a few remain valuable for evaluations in this context. Also, after validations, we found evidences that the heuristics of Al-Razgan et al. have a great coverage of usability problems of mobile applications used by the elderly, as detected from test with users.

9) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2017). Mobile application usability heuristics: Decoupling context of use. International Conference of Design, User Experience, and Usability, 410-423, Springer.*

Abstract: Context of use is a vital consideration when evaluating the usability of mobile applications. Thus, when defining sets of heuristics for the usability evaluation of mobile applications, a common practice has been to include one or more heuristics that consider context of use. Yet, most evaluations are conducted within usability labs. Consequently, the aim of this research is to question the utility of attempting to include inherently complex areas of context of use within limited sets of mobile application usability heuristics. To address this, a mapping study uncovered six sets of heuristics that can be applied to mobile application usability evaluations. A within-subjects empirical test with six Human-Computer Interaction practitioners evaluated a well-known travel mobile application using three sets of the mapped heuristics. The study found that the common practice of including context of use within mobile application usability heuristics is an ineffective approach.

10) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2017). From healthcare to Human-Computer interaction: Using framework analysis within qualitative inquiry. International Conference on Applied Human Factors and Ergonomics, 93-100, Springer.*

Abstract: One of the primary methods used by healthcare researchers following qualitative inquiry is framework analysis. The method lends itself to revealing patterns within a matrix, which can be easier than attempting to surface patterns across pages of coded text. However, despite the reliance of framework analysis by healthcare researchers, few Human-Computer Interaction researchers have used the method. It is not clear why this is the case. Within this paper, the authors demonstrate a step-by-step empirical example of framework analysis within a Human-Computer Interaction project consisting of 16 interviews; thereafter discussing the benefits of the method.

11) *Joyce, G., Lilley, M., Barker, T., & Jefferies, A. (2018). Heuristic evaluation for mobile applications: extending a map of the literature. International Conference on Applied Human Factors and Ergonomics, 15-26, Springer.*

Abstract: Ensuring that mobile applications are as usable as possible is an important area of Human-Computer Interaction research. Part of that research effort is to consider how traditional, tried-and-tested usability evaluation approaches might be applied to newer technologies, including mobile applications. The contribution of this work is to further the work of other researchers by discovering if heuristic evaluation is commonly applied to mobile applications by Human-Computer Interaction practitioners. Additionally, the authors empirically test the suggestion that Nielsen's heuristics may be generic enough for the usability evaluation of mobile applications.

APPENDIX L: FINAL SET OF MOBILE APPLICATION USABILITY HEURISTICS AND CONTEXTUAL USABILITY EVALUATION PROTOCOL

The final set of mobile application usability heuristics are:

- *M1 Interaction:* Ensure that tasks can be completed quickly and easily on mobile apps by focusing on specified user goals, minimizing data input, using device capabilities, smart defaults, appropriately sized tap targets, and offering clear affordances [...have an impact on the task being completed quickly and easily?]
- *M2 Micro-usage:* Ensure that the mobile application is designed for micro-usage as the user might be frequently distracted [...have an impact on the perception that the mobile application is tolerant of a user that is frequently distracted?]
- *M3 Readability:* Ensure that all elements, including graphics and text, on small mobile application screens are big enough to be readable in portrait and/or landscape modes [...have an impact on the readability of all screen elements in landscape and/or portrait modes?]
- *M4 Simplicity:* Ensure that the elements, including graphics and text, on each mobile application screen are only those required to achieve a specified goal [...have an impact on the perception that only the elements, including graphics and text, required to achieve a specified goal are on the mobile application screen(s)?]
- *M5 Consistency:* Ensure that elements, including graphics and text, used on each mobile application screen are consistent across the application, while conforming to platform and industry conventions familiar to the user

[...have an impact on the perception that all elements, including graphics & text, are consistent and familiar throughout the mobile application?]

- *M6 Errors:* Ensure that the mobile application is tolerant of errors, allowing undo and redo, and using real-time validation of form field input [...have an impact on the perception that the mobile application is tolerant of errors?]
- *M7 Responsive:* Ensure that interactions with the mobile application are acknowledged instantly, even if an operation may take time to complete [...have an impact on the perception that the mobile application feels responsive?]
- *M8 Learnability:* Ensure that the mobile application is easy to learn, using tutorials or onboarding if necessary. Tutorials should only focus on critical areas and have a quick way to exit [...have an impact on the perception that the mobile application is easy to learn?]
- *M9 Personalization:* Ensure that the mobile application can be personalized, customized, and/or configured to suit the user [...have an impact on the perception that the mobile application can be modified, such as personalized, customized or configured, to suit the user?]
- *M10 Continuity:* Where it makes sense to do so, ensure that the mobile application is part of a continuous experience across other device types, such as desktop and TV, thereby allowing the user to continue a task on the mobile application that was started on another device, or continue a task on another device that was started on the mobile application [...have an impact on the ability to continue a task on the mobile application that was

started on another device, or continue a task on another device that was started on the mobile app?]

The final contextual usability evaluation protocol is:

1) *Prioritize tasks and contexts of use:* Create a research plan that prioritizes key tasks and contexts of use from research conducted earlier in the mobile application software development life cycle;
2) *Define user stories:* Within the research plan, define contextual usability evaluation user stories that should be evaluated. Contexts of use can be unidimensional or multidimensional;
3) *Recruit participants:* Recruit at least eight representative participants;
4) *Conduct evaluation:* Request that participants attempt specific tasks within a specified context, whereby the order of contexts of use are randomised;
5) *Task Outcome:* Following an attempt at each task, each participant marks the task as either task success (1) or task failure (0). Notes are written or snippets are created, which may include reasons for any minor or moderate delays or frustrations even if the task is considered successful;
6) *Next context:* Participants move onto the next context of use and attempt the same specific tasks. This cycle continues until all tasks are attempted within all specified contexts of use;
7) *Display output:* Following a statistical analysis using Cochran's Q and follow-up McNemar tests, differences in the impact of context of use on mobile application usability between contexts of use are determined. Results are displayed in statistical format and on a bar chart using green for task success and red for task failure.