A Process-Based Approach to Test Usability of Multi-platform Mobile Applications

Ingrid do Nascimento Mendes and Arilo Claudio Dias-Neto(✉)

Institute of Computing (IComp),
Federal University of Amazonas (UFAM), Manaus, Brazil
{inm,arilo}@icomp.ufam.edu.br

Abstract. In order to make a mobile application available to a wider audience, developers need to implement different versions on various mobile platforms. As consequence, the application needs to work properly in these platforms considering all quality characteristics. Usability is one of these characteristics that has direct impact on the success or failure of a mobile application. There are many studies that address mobile application tests, however they often do not handle multiplatform tests and also only cover few number of usability attributes used in mobile applications. In this paper, we propose a process-based approach that guides the activities related to usability testing in multiplatform mobile applications. To do that, specific guidelines and templates were proposed to support usability testing in a multiplatform environment. In addition, our approach comprises all seven usability attributes described in the PACMAD (People At the Centre of Mobile Application Development) model. In order to evaluate the proposed approach, a case study using the three current main mobile platforms was performed taking the Flipboard app as the sample mobile application.

Keywords: Usability · Mobile application · Compatibility · Testing

1 Introduction

Since their appearance, applications for mobile platforms, or simply mobile apps, have been constantly evolving in terms of technology and popularity. They are being used in many areas such as retail, media, travel, education, health, finance and others [1]. This large and growing number of mobile apps has challenged software engineers to develop applications with a high level of quality in order to make them more attractive and competitive in this new market. In addition, this platform has introduced some aspects and challenges to be considered during the software development, such as: the need to adapt to the mobile context of use, limited connectivity, small screen size, different video resolutions, limited processing capacity, diversity of data access methods and diversity of platforms [2, 3].

In this context, one of the main challenges to mobile app developers is to keep compatibility of their applications with different mobile platforms. In general, each mobile platform has a particular programming language with its own characteristics and constraints. In order to reach a wider audience, developers need to implement versions of a mobile app on various platforms. These versions must be compatible,

A. Marcus (Ed.): DUXU 2016, Part I, LNCS 9746, pp. 456–468, 2016.
DOI: 10.1007/978-3-319-40409-7_43

considering functional and non-functional requirements despite the specific characteristics of each platform. Therefore, a mobile app needs to work properly and similarly on these platforms, considering all quality characteristics, aiming to facilitate its use by different users in case of an exchange platform, as well as minimizing development effort, such as user interface design and tests.

Among the software quality characteristics present on ISO/IEC 25000 [4], Duh et al. [5] claim that usability has a direct impact on the success or failure of a mobile app. A good usability design contributes to increase the operation capacity of mobile device users and, thus, to improve the product quality as a whole. Users tend to choose mobile apps that are easily learned and appear to be more friendly to them [2, 6]. In addition, the compatibility of usability among versions of a multi-platform mobile app is an important requirement to attend users in different platforms.

In this scenario, usability evaluation for these applications becomes an important task to ensure not only their efficiency and user satisfaction, but also the compatibility between the different versions of the same mobile app on different platforms. Usability evaluation in mobile apps is an emerging research field. In recent years, usability analysis and measurement approaches have been proposed and/or evolved in the technical literature [7]. However, they do not support usability testing in multi-platform mobile apps, providing equivalence analysis or reusing of tests among the platforms. Thus, testers have a high effort to design and execute tests in this scenario.

In this paper, we propose an approach that comprises a process, guidelines, artifact templates, and metrics that aims to support usability testing of multi-platform mobile apps. This approach was designed using the PACMAD (*People At the Centre of Mobile Application Development*) usability evaluation model for mobile apps [8]. We applied the proposed approach in a case study that tested a mobile app developed for the three main mobile platforms (Android, iOS and Windows Phone). As a result, we obtained quantitative and qualitative data that provide initial evidences of the feasibility/applicability of the proposed approach for evaluating the usability of a multi-platform mobile app from the seven usability attributes cited previously.

The paper is organized as follows: Sect. 2 describes concepts of usability testing in mobile apps and related works. Section 3 presents the proposed approach to support usability testing in multi-platform mobile apps, component-by-component. Section 4 reports a case study aiming to describe the feasibility of the proposed approach for usability testing in a multi-platform mobile app. Finally, in Sect. 5, the conclusions of this research and future works are described.

2 Usability Testing in Mobile Applications

Usability testing in mobile apps must take into account the variety of devices with different screen sizes and resolution. Depending on the device platform, its visual elements should follow a guide with the correct patterning of interface elements positions, such as buttons, windows, colors, to not compromise system's functional aspects [3]. Mobile platforms have specificities that affect the development of apps. In mobile apps development, the interface navigation flow design must be provided to support the tester in the software validation according to the specified requirements.

There are several and specific techniques for usability evaluation that can be applied to mobile apps. They can be classified as Empirical (involve the use of participants representing users of the application under test, such as field observations, interviews or focus group) or Non-Empirical (do not use participants and they may be necessary in situations where there may be trouble in finding the target audience or even when there are confidentiality issues, such as expert review, cognitive walkthroughs, heuristic evaluation). These techniques must be combined with approaches that aim to evaluate and measure the usability in mobile apps.

Some works tried to describe experiences on reporting usability testing in multi-platform mobile apps. In [2], the influence of phone complexity and user expertise on performance, ease of use and learnability is analyzed of different mobile phones from different platforms (Nokia, Siemens, Motorola). Nokia's device had the lowest and Motorola, the highest complexity, having Siemens' device ranging between them. The second independent variable was user expertise: 30 novices and 30 experts solved six tasks. Differences among the mobile phones regarding effectiveness, efficiency and learnability were found: Nokia users showed the best performance.

In [3, 6, 9] the authors present studies that evaluated and characterized approaches to support the usability evaluation in mobile apps. From 2004 until 2013, there was a considerable increase in approaches to support usability evaluation in mobile apps, and these assessments cover three types of mobile apps: web, native and hybrid [10]. In those studies, the authors did not map the approaches that deal with usability testing for multi-platform mobile apps.

In this work, we propose an approach that aims to test the compatibility of usability in multi-platform mobile apps comprised of process, guidelines, artifacts templates, and metrics, to be described in Sect. 3.

3 The Process-Based Approach for Compatibility Usability Testing in Multi-platform Mobile Applications

The proposed approach has the purpose of supporting testers of mobile apps in usability testing considering different versions of the same application for different devices/platforms (compatibility testing). It comprises four elements, detailed bellow:

- **Usability Testing Process:** describes four activities (and 16 sub activities), roles, and artifacts to be produced for usability testing in mobile apps. Moreover, these (sub)activities aim to support the analysis and reuse of usability tests considering versions of the same application for multiple mobile platforms. In this paper, we cannot show the complete description of this process due the limited space.
- **Guidelines:** recommend how to perform each activity of the usability testing process in order to test multi-platform mobile apps according to practices extracted and adapted from the technical literature.
- **Artifact Templates:** propose a "skeleton" of the artifacts to be produced when following the proposed process in order to test multi-platform mobile apps.

- **Metrics:** indicates how to measure and evaluate the several usability attributes that need to be analyzed in a mobile app. The attributes were based on the PACMAD model [8] and the metrics were extracted from the technical literature.

3.1 Process to Support Compatibility Usability Testing in Mobile Apps

It is composed of four main activities (Fig. 1), described in the following subsections. This process can be applied to test usability in (multi-platform) mobile apps. However, we introduced sub activities to deal with scenarios in which testers need to evaluate the usability of a mobile app implemented for several mobile platforms/devices.

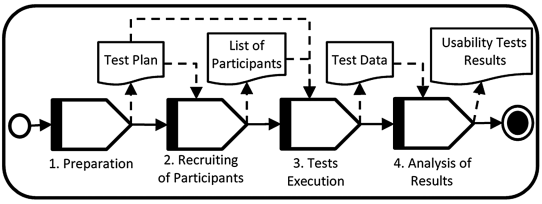

Fig. 1. Graphical representation of the usability testing process

Preparation. To build and prepare all necessary framework for implementing an usability test, starting with an initial study of the mobile app under test to understand its goals/features, target audience identification and specificities for each platform. After this step, we can define the test method, in order to choose the type of evaluation (e.g. field study and/or laboratory experiment), test environment (e.g. device and/or emulators) and mobile platforms in which the mobile app needs to be evaluated. From that, scenarios/tasks and sessions can be structured for the different mobile platforms, we can select attributes and their respective metrics and data collection approaches, the support material is constructed and a pilot test is executed, to try the designed tests. This activity is structured in six sub activities, as presented in Fig. 2.

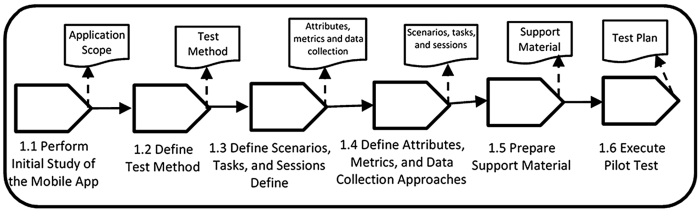

Fig. 2. Graphical representation of the activity *Preparation*

They will be described using guidelines (one of the approach elements) to perform usability testing in multi-platform mobile app.

- *Guideline #1*. The platforms/devices to be tested for a mobile app must be defined at the beginning of the process, since this decision will affect all planning stage.
- *Guideline #2*. In the creation of a task list in a new test instantiated for multiple platforms, we need to identify tasks required for all platforms and tasks not required for one or more platforms. That can be done by an analysis of equivalence among the features provided for all mobile app's versions.
- *Guideline #3*. When you already have a previous test as baseline, we need to identify tasks that will be kept, upgraded or that no longer exist, and if it necessary to create new tasks. Based on these issues, it is possible to reformulate a list of tasks and scenarios instantiated to another platform.
- *Guideline #4*. From this list of tasks and scenarios, we can specify some metrics that can be used to compare results among the selected devices/platforms, such as: possible paths to complete the scenario/task, total of keys or screen to complete the scenario/task, according to what was defined on the previous activity.
- *Guideline #5*. Apply the planned test to participants of all platforms aiming to confirm the good test planning and capture metrics or identify gaps.

Recruiting of Participants. To select participants to perform test on all platforms. Initially, we can apply a survey to candidates to show desirable characteristics of the participants (e.g. if they attend the target audience for the application). After that, we identify the participants that attend to the test's goal. Finally, we need to contact them to schedule the tests. This activity is structured in three sub activities (Fig. 3).

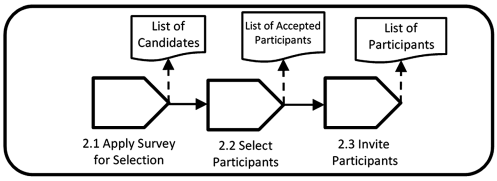

Fig. 3. Graphical representation of the activity *Recruiting of Participants*

The following guidelines describe how this activity is performed in order to support usability testing of multi-platform mobile apps:

- *Guideline #6*. Choose the same amount of participants for each platform under test with similar profile and background in their platforms.
- *Guideline #7*. Balance the distribution of participants in each platform according to their expertise, which influence on the test results because, for example, participants with high and low expertise may tend to act in different ways.

Tests Execution. The test environment must be ready to receive the participants, avoiding any loss of time and data. The tools used to capture data should be initialized and previously checked as well as printing any other support material used during the tests. The test itself must be performed and the data from videos/audios/survey responses should be stored in spreadsheets or other documents. This activity is structured in three sub activities, as presented in Fig. 4.

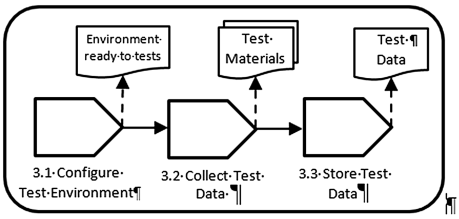

Fig. 4. Graphical representation of the activity *Tests Execution*.

Guidelines to perform this activity in order to support usability testing of multi-platform mobile apps are described below:

- *Guideline #8*. Configure the test environment the same way for each platform, following the test plan, avoiding influence on the results.
- *Guideline #9*. Storing individually the obtained data for each platform under test.

Analysis of Results. To analyze data obtained in tests execution, providing statistics and visual information (graphs). Associating this information to observed usability issues, reporting them to stakeholders. It is structured in three sub activities (Fig. 5).

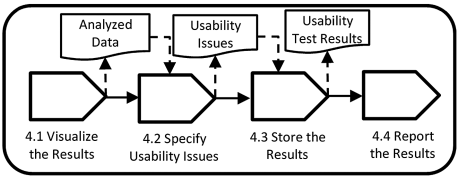

Fig. 5. Graphical representation of the activity *Analysis of Results*.

Some describe guidelines to perform this activity in order to support usability testing of multi-platform mobile apps are:

- *Guideline #10*. Usability issues found in all platforms may mean that this problem is independent of platform structure and it would be more associated with the app.
- *Guideline #11*. In some cases, participants of a mobile platform may have difficult to use a screen or task and the participants of another platform do not have this difficulty. This scenario suggests the usability issue would not be present on both platforms, representing an improvement opportunity in the first platform.

Due to space restriction, it is not possible to show and describe all sub activities that comprise the usability testing process. In the next section, we present the list of metrics, one of the elements that comprise the proposed approach, which can be applied to test usability in mobile apps.

3.2 Metrics for Usability Testing in Multiplatform Mobile Apps

Harrison et al. [8] proposed a usability model called PACMAD (*People At the Centre of Mobile Application Development*). This usability model aims to address some of the shortcomings of existing usability models when applied to mobile apps.

The PACMAD model identifies seven attributes used to define metrics for usability evaluation in mobile apps. Each attribute has an impact on the overall application usability.

- **Effectiveness:** the ability of a user to complete a task in a specified context. To evaluate effectiveness, we suggest the metric *total of (not) completed tasks* [8].
- **Efficiency:** the ability of the user to complete their task with speed and accuracy. It can be evaluated by the *total of attempts*, *keys and time* to complete each task [8].
- **Errors:** how well the user can complete the desired tasks without errors. Errors, as attributed related to effectiveness and efficiency, can be measured by tasks with more problems to be concluded (*attempts*, *time* and/or *total of keys*) [3, 11].
- **Learnability:** the ease with which users can gain proficiency with an application. It can be measured by repeating similar pairs of tasks in each session [2].
- **Memorability:** the ability of a user to retain how to use an application effectively. It can be measured by performing similar tasks into two sections, the second one after a time period of 3 days – default for all participants – in order to verify the performance after a period of inactivity, as described in [2].
- **Cognitive Load:** analyzes the impact that using the mobile device will have on the user's performance. It can be analyzed by several metrics [1]: user experience, level of distractions in the test environment, such as environment with low, medium, and high distraction level, and interface complexity, observed by the screen in which participants reported more problems.
- **Satisfaction:** the perceived level of comfort and pleasantness afforded to the user through the use of the software. It can be measured analyzing the attitudes of participants when using the application. It can be collected by interviews or questionnaires designed during the activity *Preparation*, as suggested by [11].

In the next section, we introduce some new artifacts produced during the usability testing process and propose templates for them, aiming to provide an instrument to support the usability testing process execution. The goal is to minimize the effort to adopt the proposed approach in real software projects.

3.3 New Artifacts to Support Usability Testing in Multi-platform Mobile Apps

The main artifacts to be produced in each activity of the proposed process are presented in Fig. 1. They are traditional artifacts present in usability testing process. Due to space restriction, the template of these artifacts (and others) with an example can be downloaded at https://goo.gl/JDqYcu. In order to support usability testing in multi-platform mobile apps, mainly the Preparation activity, we propose some new artifacts, as described below:

- **Map/Description of Equivalent Elements:** the map shows visually by images the equivalent elements of the application screens on all platforms under test in order to support the identification of key differences. It must be accompanied by a textual detail of the most significant differences present in screens or elements that are part of the tasks already selected. With these artifacts, it is possible to get an overview of the differences among the interface elements in all platforms.
- **Visual Map:** Map describing the flow of application's screens for each platform. It supports the identification of attributes/metrics to be evaluated, such as minimum number of screens or keys for each task.
- **Tasks Flow Graph:** Each screen in the visual map contains an identification. This artifact brings another representation for the flow of screens (described in the visual map), by means of graphs, in which each node is a visual map's screen and each edge indicates a possible transition between screens.

In the next section, we will present the use of the proposed approach (including these new artifacts) to test the usability compatibility of a real mobile app (*Flipboard*) developed for the three main mobile platforms: Android, iOS, and Windows Phone.

4 A Case Study to Test Usability in a Multi-platform Mobile Apps Using the Proposed Approach

We selected *Flibboard* as application to be tested in this case study. It is an aggregator of social networking content with a particular news feed customizable according to the user's interests. There were two reasons for choosing this mobile app: (1) it is easy to find people in the target audience and (2) the existence of the application on three major current mobile platforms (Android, iOS and Windows Phone).

4.1 Preparation and Recruiting of Participants

In Preparation activity, the main concerns are the selection of platforms versions to be tested (in this case, Android, iOS, and Windows Phone – *Guideline #1*) and the definition of sessions, scenarios and tasks for all platforms. We planned two 15 min sessions. The first session aimed to observe the first participant's interaction with the application during the tasks, i.e. the attribute Learnability. The second session was held after a period of inactivity (3 days) in using Flipboard application in order to see if they could perform tasks with varying degrees of proficiency over the first session and, thus, infer a result to the attribute Memorability. Metrics for the seven attributes (effectiveness, efficiency, satisfaction, ease of learning, memorability, errors and cognitive load) that comprise the proposed approach were evaluated (*Guideline #4*).

To support the analysis of equivalence among the platforms a *Map of Equivalent Elements* (partially showed in Fig. 6) has been created. We also created the artifact *Description of Equivalent Elements*, detailing the differences among the platforms, but there is no space to present it in this paper.

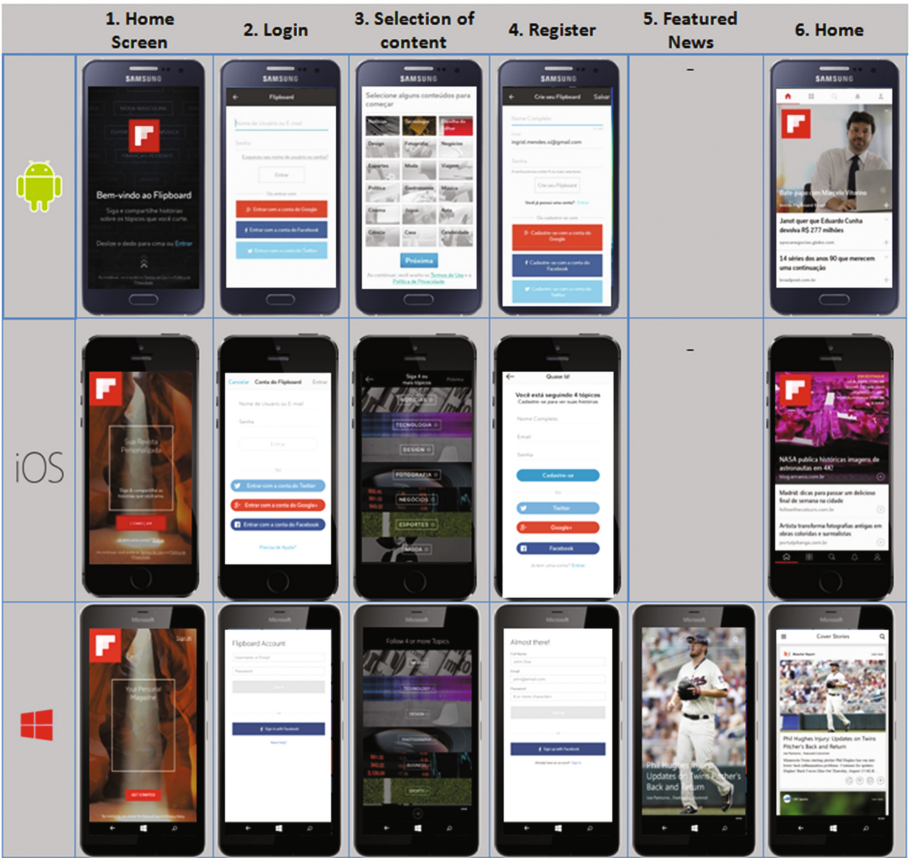

Fig. 6. Equivalence Features Map of (Partial View)

From this analysis, we could easily define five scenarios and eight tasks to be followed by the participants (Table 1 – *Guideline* #2). Task T1 is available only in the 1st session and the task T2 is available only in the 2nd session. Moreover, Task T4 has three different descriptions, one for each platform under test, due to their particularities. Task T6 does not exist in Windows Phone's platform (*Guideline* #3).

For each task, we specified possible paths to be followed to complete it using the artifact *Visual Map* (Fig. 7: left) and *Tasks Flow Graph* (Fig. 7: right) for each platform.

From these paths, we calculated total of keys and screens to define an oracle to be used in the results analysis (Table 2). We can observe some cells contain "-", which means that this platform does not have a particular path or task. This information will be used in the analysis of usability metrics that comprises the proposed approach.

We selected 15 participants (5 per platform – *Guideline* #5 and *Guideline* #6) for this study, totaling ten men and five women with average of 22.3 years old (*Guideline* #7). They were selected from a survey conducted with students in UFAM/Brazil.

Table 1. List of tasks to be performed per scenario

Scenarios	Tasks
S1: Start the use.	T1: Create an account in the application
	T2: Log in using your account
S2: Setting feed of news.	T3: Search and add categories of news in the feed
S3: Definition of news to be read later in a private magazine.	T4: [Android] Select some interesting news and flip in the option *read later*; [iOS] Select some interesting News and flip in the option *read later*; [WP] Select some interesting news and flip in any magazine
S4: Organizing Facebook's content in the format of magazine.	T5: Link your Facebook's account to Flipboard (or check it)
	T6: Navigate in the section "Pages" of Facebook by Flipboard.
S5: Additional features.	T7: Create a new public magazine
	T8: Flip in your magazine posts related to it

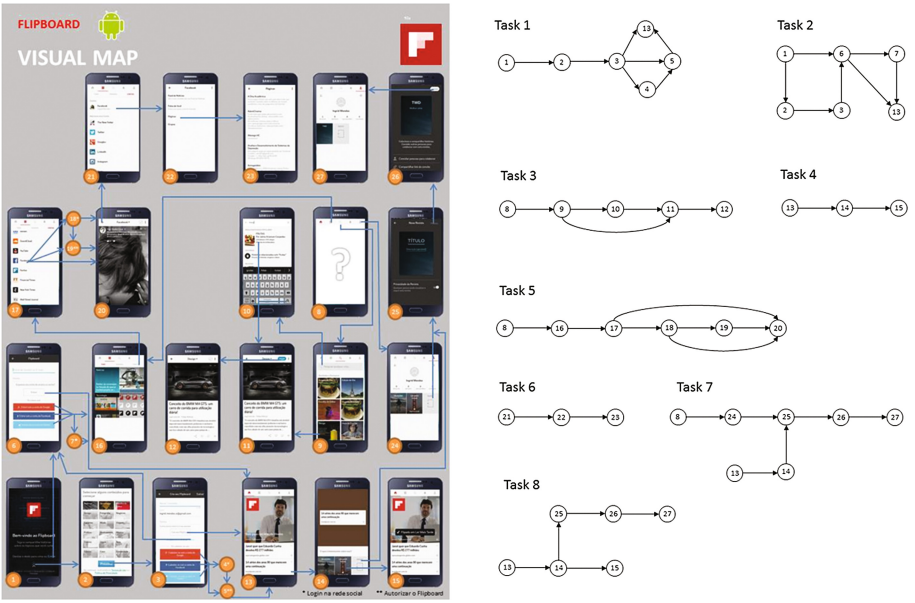

Fig. 7. (left) Visual Map of Tasks and (right) Tasks Flow Graph (Android Platform)

4.2 Tests Executions and Results Analysis

In order to execute tests, a mobile device of each platform under test was used (*Guideline #8*). In Table 3, '-' means that no platform's participant had completed the task in the session/environment, or the task does not exist in the platform/session.

Table 2. Total of keys and screens per path for all Tasks/Scenarios/Platforms

			Android		iOS		WP	
Scenario	Task	Path	Key	Screen	Key	Screen	Key	Screen
S1	T1	1	6	4	10	4	10	4
		2	6	6	11	6	11	6
		3	3	5	8	5	8	5
	T2	1	4	3	4	3	4	3
		2	5	4	5	4	5	4
		3	2	3	2	3	2	3
		4	5	5	10	5	10	5
		5	6	6	11	6	11	6
		6	3	5	8	5	8	5
S2	T3	1	4	5	4	5	4	5
		2	3	4	3	4	3	4
		3	4	5	4	5	4	5
		4	–	–	–	–	5	6
		5	–	–	–	–	4	5
		6	–	–	–	–	5	6
S3	T4	1	2	3	2	3	2	3
S4	T5	1	3	4	3	4	5	6
		2	6	5	6	5	8	7
		3	7	6	7	6	9	8
	T6	1	2	3	2	3	–	–
S5	T7	1	5	5	5	5	5	5
		2	5	5	5	5	5	5
	T8	1	2	6	2	3	2	3
		2	5	5	5	5	5	5

Table 3. Total of keys minus total of expected keys per task for all sessions/platforms

	Session 1			**Session 2 Environment 1**			**Session 2 Environment 2**			**Session 2 Environment 3**		
												
#1	31	8	0	-	-	-	-	-	-	-	-	-
#2	-	-	-	4	2	0	5	0	0	3	0	0
#3	0	0	4	0	2	0	0	0	1	0	0	0
#4	0	3	3	0	2	4	0	0	6	0	0	7
#5	3	0	0	1	0	0	1	0	1	1	0	0
#6	-	0	-	2	0	-	0	0	-	0	0	-
#7	4	3	4	1	3	3	2	1	4	1	1	5
#8	2	3	8	0	0	3	1	0	4	0	0	5

Table 3 shows the data collected during the tests organized per task/session/platform, reporting the difference between the total of keys obtained with all participants and the

total of expected keys. The same analysis was performed with the time and number of attempts required to conclude each task for each session/environment (*Guideline* #9).

The selected metrics could be evaluated based on the data presented in Table 3. From these collected data, all selected metrics could be evaluated in order to analyze the usability quality of the application under test. Due to space restrictions, it is not possible to present the complete tests analysis for each attribute selected during the preparation activity (*Guidelines #10* and *#11*). The goal in this section was to describe how the proposed approach contributed with its process, guidelines, artifacts, and metrics to perform usability testing in multi-platform mobile apps.

5 Conclusions and Future Works

In this work, we described a proposal of approach comprised by process, guidelines, artifacts templates and metrics to support usability compatibility testing in multi-platform mobile apps. This approach has been defined from the instantiation of PACMAD (usability evaluation model for mobile apps).

The application of the process, guidelines, templates and metrics aimed to enhance the results of a usability test instantiated for multiple mobile platforms, with well-defined scope and without reworking in the preparation activity, whereas the proposal aims to support the reuse of tasks and metrics across different platforms.

In order to assess the feasibility of the approach proposed using metrics related to the seven attributes that comprises the PACMAD model, a case study was carried out with the mobile app Flipboard on three mobile platforms (Android, iOS and Windows Phone). Throughout the tests, it was possible to collect information on the number of completed tasks, number of attempts, total of keys and time used for completing tasks. As a result, the approach and its guidelines could be fully applied and it contributed to the identification of problems related to the evaluated attributes.

As future works, we intend to apply and evaluate the approach in the software industry with other mobile apps in order to confirm its feasibility and improve it from the feedback of software engineers with different experience levels in usability testing.

Acknowledgment. The authors thank FAPEAM and INDT for financial support to conduct this research.

References

1. Flood, D., Germanakos, P., Harrison, R., Mc Caffery, F.: Estimating cognitive overload in mobile applications for decision support within the medical domain. In: 14th International Conference on Enterprise Information Systems, Wroclaw, Poland (2012)
2. Ziefle, M.: The influence of user expertise and phone complexity on performance, ease of user and learnability of different mobile phones. Behav. Inf. Technol. **21**, 303–311 (2002)
3. Zhang, D., Adipat, B.: Challenges, methodologies, and issues in the usability testing of mobile applications. Int. J. Hum.-Comput. Interact. **18**(3), 293–308 (2005)

4. International Organization for Standardization, “ISO/IEC 25000 - Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE”, ISO/IEC, March 2011
5. Duh, H., Tan, G., Chen, V.: Usability evaluation for mobile device: a comparison of laboratory and field tests. In: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI). pp. 181–186 (2006)
6. Nayebi, F., Desharnais, J.M., Abran, A.: The state of the art of mobile application usability evaluation. In: 25th IEEE Canadian Conference on Electrical Computer Engineering CCECE, pp. 1–4 (2012)
7. Kjeldskov, J., Stage, J.: New techniques for usability evaluation of mobile systems. Int. J. Hum.-Comput. Stud. **60**(5), 599–620 (2004)
8. Harrison, R., Flood, D., Duce, D.: Usability of mobile applications: literature review and rationale for a new usability model. J. Interact. Sci. **1**(1), 1–16 (2013)
9. Kjeldskov, J., Paay, J.: A longitudinal review of Mobile HCI research methods. In: International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2012), pp. 69–78. doi:10.1145/2371574.2371586
10. Huy, N.P., VanThanh, D.: Evaluation of mobile app paradigms. In: Proceedings of International Conference on Advances in Mobile Computing & Multimedia (MoMM 2012), Ismail Khalil, pp. 25–30 (2012)
11. Masoodian, M., Lane, N.: An empirical study of textual and graphical travel itinerary visualization using mobile phones. In: Proceedings of the Fourth Australian User Interface Conference on User Interfaces vol. 18, pp. 11–18 (2003)