

Towards a prior validation of a model-based approach for mobile usability evaluation



Lassaad Ben Ammar ^{1, 2, *}

¹College of Sciences and Humanities, Pince Sattam bin Abdul-Aziz University, Al-Kharj, Saudi Arabia

²National School of Engineering of Sfax, University of Sfax, Sfax, Tunisia

ARTICLE INFO

Article history:

Received 29 January 2020

Received in revised form

1 July 2020

Accepted 2 July 2020

Keywords:

Early usability evaluation

Mobile applications

Model-driven engineering

Mobile usability

ABSTRACT

Recently, evaluating the usability of mobile applications is gaining a lot of attention. The interest is focused on the user interface design choices that may lead to the rejection of the application. The evaluation is usually performed at the last stage of the development life cycle when the system is fully implemented. At this stage, it is difficult/expensive to go back to the design and make the required changes to solve usability issues. This problem may be alleviated by using the Model-Driven Engineering (MDE) approach, where conceptual models (elaborated at the design phase) undergo a series of transformations to generate the final applications, as automatically as possible. The transformation process establishes a mechanism of traceability between these conceptual models and the final application. Due to this mechanism, the analysis of these models to improve their usability is likely to preserve this usability at the final application or improve it. The present paper presents a model-based approach to evaluate the usability of mobile applications in the design phase. The proposed approach provides a set of usability metrics that are defined based on conceptual primitives that constitute the conceptual models. The objective is to allow the measurement of usability issues from the conceptual models. Prior validation of our proposal is presented in this paper.

© 2020 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Recently, the evaluation of mobile usability has gained a lot of attention (Reis et al., 2015; Nayebe et al., 2012). Several research works are elaborated to measure the usability of mobile applications. They usually perform usability once the system is implemented using traditional methods such as laboratory experiments and field studies. Such methods involve activities that require a lot of resources (e.g., participant users, recording systems, usability lab, etc.). Besides, a lot of rework is usually required to go back to the design and made changes. This is not always trivial considering their cost and complexity.

This problem can be alleviated using the newly proposed approach in the software engineering domain: Model-driven Engineering (MDE). In such

an approach, the interest of the analyst is focused on creating conceptual models that represent the system abstractly, and the final application will be generated (as-automatically as possible) by means of model transformation. The transformation process establishes a mechanism of traceability between the conceptual models prepared during the design phase and the final application. Taken this mechanism into account, changes made in the conceptual models are usually reflected in the final application. The usability literature presents some initiatives that demonstrated such assumptions (Abrahao et al., 2008; Abrahao and Insfran, 2006; Ammar et al., 2016).

These research works have demonstrated that evaluating the usability from the conceptual models is likely to preserve this usability at the final application or to improve it, at least to some extent. The objects of these research works are traditional desktop and web applications. However, the mobile device has some specific features (e.g., small screen size, data entry methods, limited capacity, and power process) that may introduce some new challenges to be considered by mobile application engineers. The most important one is, without a doubt, to introduce

* Corresponding Author.

Email Address: l.benammar@psau.edu.sa

<https://doi.org/10.21833/ijaas.2020.11.009>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0002-4698-3693>

2313-626X/© 2020 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

a usability evaluation method that suits mobile applications and considers their features.

The present paper presents a continuation of previous work that addresses this issue and proposes to incorporate usability engineering as part of the mobile application development process, which follows MDE principles. The aim of this paper is to empirically validate an early usability evaluation process.

The remainder of this paper will be structured as follows. Section 2 presents a brief review of research works related to the context of this paper. Section 3 details our proposal for an early usability evaluation process. The feasibility and the potentialities of the proposed approach are illustrated in section 4. Finally, the conclusion and some perspectives for future works are presented in section 5.

2. Literature review

Several research works exist in mobile usability literature (Moumane et al., 2016; Zhang and Adipat, 2005). They are usually classified into two main categories:

- **Laboratory experiment:** That is usually conducted in a controlled environment (usability labs) where participants are given a pre-defined set of tasks to be accomplished. These tasks are defined before the experiment, and participants are intended to accomplish them without any assistance. During the experiment, data on performance measures are collected and documented. These data are analyzed to highlight usability problems. Based on the evaluation result, the design of the application object of the evaluation and its user interface can be improved. The work presented in Moumane et al. (2016), Barros et al. (2014), and Umuhoza and Brambilla (2016) were, among others, examples of research studies that belong to this category. The main limitation for all research works belonging to this category is related to the differences between the controlled environment and the real world.
- **Field studies:** That involves activities such as observation or interviews to collect data about the user's perception concerning the system usability. Participants are allowed to use the application and usability experts take notes about their behavior while interacting with the system. Their perception concerning system usability can be collected through a questionnaire (Moumane et al., 2016). The analysis of the notes taken by usability experts or the users' perception collected through the questionnaire aims to identify usability problems and suggest recommendations to improve the design choices leading to these problems. The main limitations of the field study techniques are related to the quality of the questionnaire and sufficient control over users during the field study. In addition, data collection in the real world is usually a complex task to be performed.

Note that mobile usability literature presents some initiatives that gather a set of usability metrics into a model. Hussain and Kutar (2009) and Harrison et al. (2013) were examples, among others, of such related works. The main drawback of these initiatives is related to the proposed metrics dealing with usability issues that can be measured only if the application is implemented. Besides, there are no specific details about how the proposed metrics can be measured and how to interpret their scores.

After reviewing the mobile usability literature, we conclude that, even if the actual state of the art presents several initiatives that we consider relevant and important, a lot of problems persist, and the following shortcoming can be identified:

- Usability evaluation is usually performed once the system is implemented.
- There is a lack of precise details about how to measure usability metrics and how to interpret their scores.
- Usability measures are independent of the development cycle. Hence, designers and developers do not know what to change and where to improve the value of these measures and, consequently, the usability of the final product.
- To the best of our knowledge, no proposal is dealing with the usability evaluation of mobile applications generated with an MDE approach. However, the adoption of MDE in the mobile context has experienced exponential growth (Umuhoza and Brambilla, 2016; Balagtas-Fernandez and Hussmann, 2008).

Taken into account these limitations, we concluded that mobile usability evaluation is still an immature area, and more research works are needed. To cover this need, we proposed in previous work an early usability evaluation process based on the conceptual schemas that represent the system abstractly. The goals of such a proposal are the following:

- The evaluation can be applied quickly.
- The evaluation can be applied to any MDE-compliant method.
- The evaluation must be independent of the end-users.

In this paper, a prior validation of the instrument that will be used while empirically validate the proposed evaluation process is conducted. Before that, a brief review of the proposed usability evaluation process is presented.

3. Mobile usability evaluation process: A brief review

As Fig. 1 illustrates, our proposal to measure the internal usability of a mobile application through conceptual models is made up of our steps:

- Selection of usability attributes;

- Selection of usability metrics to quantify previously defined attributes;
- The definition of indicators to interpret the value obtained by each metrics;
- The generation of the usability report which contains all detected usability problems.

Grey boxes represent existing elements that are extracted from our previous work (Ammar, 2019). The others represent new elements introduced through our proposal.

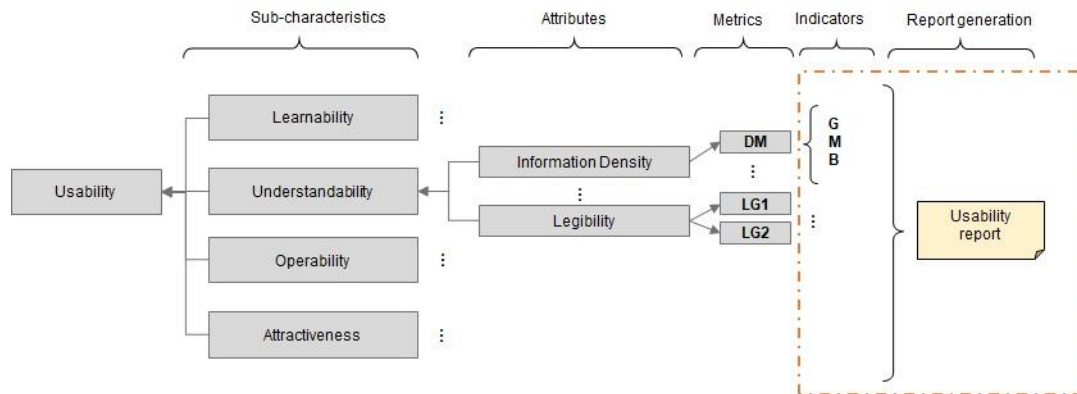


Fig. 1: Summary of the approach to measuring the internal usability from conceptual models

For the first two steps, the aim was to break down the usability sub-characteristics initiated in the ISO/IEC (2001) usability model into measurable attributes and to associate with each one a set of metrics that enable its quantification from the conceptual models.

As a result, for these two steps, a set of usability attributes and their metrics are defined. Note that metrics are defined based on the conceptual primitives that constitute the conceptual model to allow their calculation from this stage in the development process. In addition, they are defined generically, which allows their adoption in other MDE method with similar models and primitives.

Once we calculate the value of each metric, we need to interpret the meaning of its numeric value. This is the aim of the third step, wherein the mechanism of an indicator is used. This later transforms the numeric values obtained using metrics into ordinal values (qualitative values such as Good, Medium, Bad). To do this, a range of values is defined for each qualitative value. These ranges are extracted and adapted from their original work for those metrics which are previously defined in related works. For the others, which are added in our proposal, their ranges are defined by analogy to some similar metrics which are available.

Metrics values that are mapped as a Bad value will be considered as a usability problem and are added in a report. This is the aim of the last step in our proposal. Usability problems are presented in the report using a template allowing their analysis in an efficient way. The result of the analysis will help us to detect the source of the problem in the conceptual models and suggest the required change that is likely to fix such a problem.

3.1. Attributes selection

The first step in the usability evaluation process is to select the most relevant attributes to be

measured. This may depend especially on the domain of the application and the users' profiles. For example, security is more important in a mobile application for the bank. The learnability is more important for novice users, then experts.

This step usually uses an agreed usability model as a catalog from which attributes and metrics are selected. We proposed in previous work (Ammar, 2019) a usability model that can be used to measure the usability of a mobile application. In this model, the usability concept is dividing into four sub-characteristics: Learnability, Understandability, Operability, and Attractiveness. These sub-characteristics are, in turn, breaking down into attributes that can be measured using metrics. Note that metrics are defined based on the conceptual models in a generic way. This allows their application from the early stage of any MDE method with similar conceptual schemas. In addition, they are defined, taking into account the specific features of mobile devices such as small screen size and data entry methods.

Several usability models and guidelines are analyzed to extract and adapt the most relevant usability attributes and metrics by Hussain and Kutar (2009) and Harrison et al. (2013). In addition, user interface guidelines for iOS and Android are analyzed to extract the usability attributes that are considered relevant and can be evaluated from conceptual models. Note that these two operating systems form the main basis for our proposal concerning usability attributes and metrics because they are currently the most prominent operating systems, and they hold more than 98% of the worldwide market share (Jindal and Jain, 2012).

As a result of this step, a set of usability attributes are proposed, and their general description is associated with better clear their meaning. Table 1 summarizes the set of attributes associated with each usability sub-characteristic.

Table 1: Proposed attributes and their description

Sub-characteristic	Attribute	Meaning
Learnability	Prompting	Available means to help users to take specific actions such as data entry
	Predictability	The available means to help users predict his/her future action.
	Feedback	System responses to user actions.
	Information Density	The users' workload from a perceptual and cognitive point of view.
Understandability	Brevity	Available means to reduce the cognitive efforts of the users while interacting with the system.
	Navigability	The ease with which a user can move around in the application.
	Message Quality	The expressiveness of the error message.
	Legibility	The degree to which a reader can easily recognize a text.
Operability	Cancel Support	The degree of control that users have over the treatment of their actions.
	Undo Support	
	Explicit User Action	The relationship between computer processing and the users' actions.
	Error Prevention	Available means to prevent data entry errors.
Attractiveness	Font Style Uniformity	Total number of used font style per user interface
	Color Uniformity	Total number of used color per user interface
	Consistency	The maintaining of the interface design choices in a similar context.
	Balance	The distribution of the optical weight in a user interface.

Note that the list of usability attributes in [Table 1](#) is not intended to be exhaustive. These attributes are a starting point for identifying the usability features of mobile apps from the early stage of the development process. We intend to add the missing attributes to the list when more information becomes available.

3.2. Metric selection

After selecting the usability attributes to be measured during the evaluation process, the next step is to select one or more metrics from those associated with each attribute. As is mentioned before, a set of usability metrics is proposed in ([Ammar, 2019](#)). Recall that metrics are defined generically based on the conceptual primitives that constitute the conceptual models.

For each metric, we proposed a calculation formula allowing us to calculate the value of this metric. In what follows, we present some calculation formula for the newly added metrics that are based on user interface guidelines for mobile applications such as IOs and Android.

Structured Text Entry (STE): This metric is defined as the percentage of structured text entry used when the system requires an exact format for the data to be entered (e.g., phone numbers, credit cards). It was considered contributing to the Prompting attribute. It can be calculated using Eq. 1.

$$STE = \frac{\sum_{i=1}^n \text{Structured_Text_Entry}(i)}{n} \quad (1)$$

where, *Structured_Text_Entry* (i)=1 for an input element with a mask, 0 otherwise; n is the total number of input element requiring and exact format.

Built-in Icons (BI): This metric is defined as the percentage of built-in icons (system icons) used by an active element in the user interface. Built-in icons are largely recommended because they are familiar to users. It can be calculated using Eq. 2.

$$BI = \frac{\sum_{i=1}^n \text{Built_in_Icon}(i)}{n} \quad (2)$$

where, *Built-In_Icons* (i)=1 for action element with system icon, 0 otherwise; n is the total number of the active element in the user interface.

Default Value (DV): This metric is defined as the percentage of input elements with default values. It can be calculated using Eq. 3.

$$DV = \frac{\sum_{i=1}^n a_i}{n} \quad (3)$$

where, $a_i=1$ for an input/select element with a default value, 0 otherwise; n is the total number of input element.

Tapped element Size (TeS): This metric is defined as the percentage of the pointer target element, which has a size greater or equal than 44pt x 44pt. It can be calculated using Eq. 4.

$$TeS = \frac{\sum_{i=1}^n a_i}{n} \quad (4)$$

where, $a_i=1$ for object i with an area greater or equal than 44 x 44 dp, 0 otherwise; n is the total number of pointer target elements in the user interface.

Text Size (TxS): This metric is defined as the percentage of an interface element (text, list items, etc.), which have a size greater or equal than 16pt. It can be calculated using Eq. 5.

$$TxS = \frac{\sum_{i=1}^n \text{FontSize}_i}{n} \quad (5)$$

where, *FontSize_i*=1 for element i with size greater or equal than 16pt, 0 otherwise; n is the number of input text in a user interface.

[Table 2](#) illustrates the total set of metrics that are associated with the usability attributes.

3.3. Indicator definition

To be able to interpret the meaning of the numerical values afforded from the previous metrics, we opted for the mechanism of indicators. This later consists of assigning a qualitative value with each range of numerical values obtained using the calculation formula of a metric. Three indicators are used in this paper: Good (G), Medium (M), and Bad (B). Some indicators are extracted from related works that are empirically validated ([Ammar et al., 2016](#); [Panach et al., 2011](#); [Fernandez et al., 2013](#)). For the newly added elements, the good value is built on guidelines and models presented in the usability

literature and is used to estimate the bad value. After that, the medium value is obtained by distributing

the range (since we have the two extremes) equitably.

Table 2: Proposed attributes and their description

Attribute	Metric Description
Prompting	Percentage of labels with supplementary information.
Predictability	Percentage of structured text entry among all data entry elements that require an exact format for the data to be entered (e.g., phone numbers, credit card).
Feedback	Percentage of meaningful labels (less than 3 words).
Information	Percentage of action elements with built-in icons.
Density	Percentage of action with feedback response.
Brevity	Percentage of the screen occupied by objects.
Navigability	Percentage of input elements that present a default value.
Message Quality	Average of navigation element per interface.
Legibility	Average of word per error message.
Cancel Support	Percentage of a pointer target element with a recommended size (greater or equal than 44pt x 44pt).
Undo Support	Percentage of interface elements with a recommended font size (greater or equal than 16pt).
Explicit User Action	Percentage of actions that can be canceled without harmful effects.
Error Prevention	Percentage portion of actions that can be undone without harmful effects.
Font Style	Percentage of validation action following data entry.
Uniformity	Percentage of enumerated input elements that use the primitive, which represents a list.
Color Uniformity	Total number of used font styles per user interface.
Consistency	The total number of used colors per user interface.
Balance	Percentage of the repeated elements that have the same label.
	The difference between the total weighting of components on each side of the horizontal and vertical axis.

Table 3 shows the list of indicators defined for the aforementioned metrics. Note that the indicator values for these metrics are proposed by analogy to some metrics that are previously validated empirically in related work. The future work that we plan is to conduct an experiment to empirically validate these indicators or identifies those that need to be adjusted.

Table 3: Proposed indicators for usability metrics

Metric	G	M	B
STE	$STE \geq 0.8$	$0.7 \leq STE \leq 0.8$	$STE \leq 0.7$
BI	$BI \geq 0.8$	$0.7 \leq BI \leq 0.8$	$BI \leq 0.7$
DV	$DV \geq 0.2$	$0.1 \leq DV \leq 0.2$	$DV \leq 0.1$
TeS	$TES \geq 0.8$	$0.7 \leq TES \leq 0.8$	$TES \leq 0.7$
TxS	$TS \geq 0.8$	$0.7 \leq TS \leq 0.8$	$TS \leq 0.7$

3.4. Usability report generation

This step consists of inserting all usability problems detected during the evaluation process using a well-defined template. To be considered as a usability problem, the value of a metric must be Bad. For other values, we consider them as accepted and do not raise a usability problem.

To improve the readability of the usability report, we propose a template to represent a usability problem. The template contains the following fields:

- Problem identifier: A numeric identifier for the usability problem with the format UPXX.
- Description: A short description of the usability problem.
- Affected usability attribute: The usability attribute that is affected by the usability problem. This may help us to identify the conceptual primitive source of the problem.
- Recommendation: A short description of the changes that are required to fix such a problem.

These changes are usually related to the conceptual primitive source of the problem.

At the end of the evaluation, the recommended changes are made, and a re-evaluation may be conducted to check whether the previously detected problem is fixed or need more rework.

4. A priori validation of the early usability evaluation

This section explains the experiment that we used to evaluate our proposal for early usability evaluation of mobile applications. The experiment aims to compare the usability measure obtained by our proposal and those perceived by end-users who interact with two mobile applications.

Fig. 2 summarize the general process to be used to empirically validate the early usability evaluation process. It consists of comparing the values of usability issues obtained by our proposal and those perceived by end-users.

4.1. Objectives

Following the goal-question-metric (GQM) template, the objective of the experiment was to analyze internal measure of the usability for the purpose of evaluating them with regard to their coherence with regard to users' perception.

The research question that can be derived from this objective is the following: Is there a significant coherence between usability measures obtained by our proposal for early usability evaluation and those perceived by end-users?

Two hypotheses related to this question are identified:

- H0: There is no significant difference between the usability measures obtained with our proposed method (internal usability IU) and that perceived by the end-user (external usability EU).

$$H0: \mu IU = \mu EU.$$

- H1: There is a significant difference between the usability measures obtained with our proposed method (internal usability IU) and that perceived by the end-user (external usability EU).

$$H1: \mu IU \neq \mu EU.$$

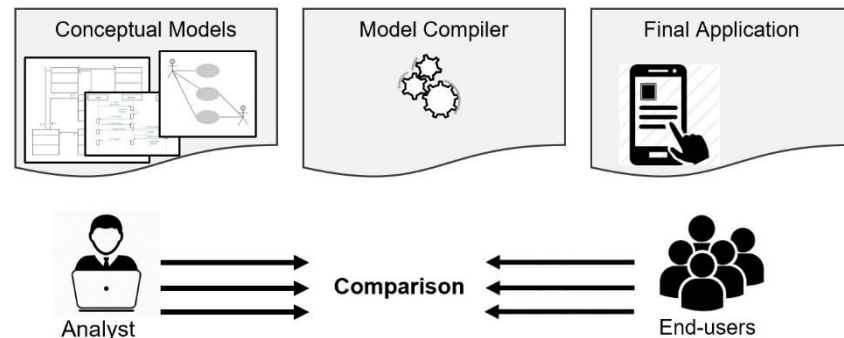


Fig. 2: General process for empirically validate the early usability evaluation process

4.2. Subjects and objects

The objects were two mobile applications that are largely used in Saudi Arabia country. The first one is the individual eService mobile application Absher. It allows citizens and residents in Saudi Arabia to use a variety of governmental services (e.g., Hajj permit, passport info, traffic violation, etc.). The second one is the mobile banking application, AlMubasher Retail. It allows AlRajhi customers to do most of their banking transactions through their mobile phones (e.g., Currency Converter, ATM Locator, Transfer, Payment, etc.).

The subjects were 18 representative users (doctors and students) from the Prince Sattam bin Abdul-Aziz University. All subjects are familiar with the first application Absher. However, some of them are not familiar with the second one since they are not customers of AlRajhi bank and are familiar with other similar applications provided by their banks. All subjects had a high level of knowledge in the mobile application domain used in the experiment; however, they did not have any experience in conceptual modeling. Their age ranged between 20 and 45 years old.

4.3. Identification of variables

While designing the experiment, we identified two types of variables:

- **Response variables** that correspond to the outcome of the experiment. Usability was the target of this experiment and was measured in terms of sub-characteristics such as Learnability, Understandability, and Operability. Hence, each one of these sub-characteristics is considered as a response variable for the experiment.
- **Factors** that correspond to any characteristics that may vary during the experiment and affect the response variables. In this work, the usability evaluation technique was identified as a factor that

affects the response variables and had two alternatives: 1) early usability evaluation from conceptual models, and 2) usability evaluation with end-users.

4.4. Instruments

We used the following instruments to carry out our experiment:

- A demographic questionnaire to know the level of experience of each user about the application domain used in the experiment.
- A list of specific tasks for the test. Users are asked to try to accomplish these tasks without any assistance. They can ask for help only if they felt unable to complete the task.
- A user satisfaction questionnaire that contains a list of questions to be answered by end-users to capture their perception on a 5-point Likert scale. Note that each question refers to a metric defined in our proposal, which is included in the experiment. We use a specific question for each metric and not any existing questionnaire due to the specificity of the used metrics.
- A spreadsheet to accelerate the calculation of measures obtained from the conceptual models and made a comparison with those perceived by end-users.
- SPSS software to perform the statistical analysis allowing us to study the comparison in depth.

Note that we used a specific survey and not anyone from existing surveys such as SUMI, ISOMETRICS, or QUIS because of the specificity of the proposed metrics. Our instrument survey included twenty closed questions, one for each metric proposed. The question was formulated using a 3-Likert scale. Once users have filled the survey, we obtain their opinions related to each question since each question refers to one metric. Fig. 3 illustrates some questions from the survey.

The application provides me enough information about which entries are permitted in a particular situation?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The amount of information displayed in the screen is adequate?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Error messages are helpful?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reading characters on the screen is easy?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 3: Example of items from the survey

4.5. Design process

As depicted in Fig. 4, the process for empirically evaluating our proposal started with filling in a demographic questionnaire to capture the users'

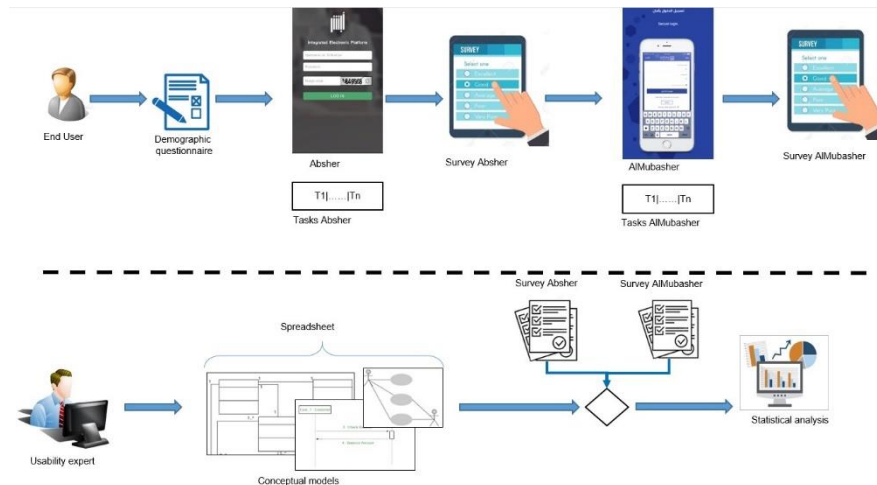


Fig. 4: The process for empirically evaluate the proposed method

In the third step, our proposal for the early usability evaluation is applied by two usability experts to calculate the usability measures using the conceptual models of each application. Note that the conceptual models of each application are designed using the MDE method proposed by Bouchelligua et al. (2010), taking into account their final user interfaces as they are presented in the application.

In the last step, the outcomes of the surveys were compared to the outcomes of the evaluation based on the conceptual models. Statistical analysis was done to fulfill this objective.

4.6. Reliability and validity

According to the experimentation literature, there are two main properties of the instrument used (e.g., interview, questionnaire, test) that indicate its quality and usefulness and must be examined; reliability and validity.

- **Reliability** refers to the extent to which the test is consistent. It checks whether the test provides similar scores or not if a person takes it again. It is usually indicated by the reliability coefficient. Cronbach's alpha is the most widely used coefficient to evaluate the reliability of an instrument.
- **The validity** refers to the extent to which the test measures what it purports to measure. Different types of validity can be evaluated during an

attitude concerning the domain of application. In the second step, users are divided into two groups and are asked to interact with the applications alternatively. The first group started with the Absher application and the second group with the banking application AlMubasher. Only users that finished all tasks successfully are asked to fill in a survey to capture their usability perception concerning the application. After that, each group repeated the same work with the other application. Note that the same questions are used for the two applications, even the tasks for each application were different.

experiment, and construct validity is the most important one when the test uses a measure as an index of a variable that itself is not directly observable. It defines how well a test or experiment measures up to its claims. It is essential to the perceived overall validity of the test. One way to test the validity of the questionnaire is to use the Pearson correlation coefficient.

In our experiment, we used the SPSS tool to analyse data and tested the validity and reliability of the experiment. We used the reliability analysis function provided by SPSS, which will estimate the reliability of our test and the extent to which the items correlate well with one another. It will also help us to identify the troublesome items, items that have a low item-total correlation, and their deletion would go up the alpha. Note that accepted values for alpha are those greater or equal to 0.7.

The appendix illustrates all the outputted documents for the reliability and validity test. Table 4 resumes the important information considered while validating this experiment.

Taken into account the obtained results, we decided to exclude the item that is suggested by the SPSS assistant to obtain a good Cronbach's alpha.

4.7. Data analysis

In this section, the outcomes of the two evaluation methods for each metric are compared.

The first step in this stage is to compare the average of the users' perception with the value obtained by our proposed method for each usability metric. Fig. 5

depicts the comparison for the first application (Absher), and Fig. 6 depicts the comparison for the second application (AlMubasher).

Table 4: Reliability and validity test

Sub-characteristic	Mobile apps	Cronbach's alpha	Item to be deleted	Cronbach's alpha if item deleted
Learnability	Absher	0.586	Informative feedback	0.735
	AlMubasher	0.781	None	0.781
Understandability	Absher	0.668	Text size	0.712
	AlMubasher	0.600	Brevity	0.709
Operability	Absher	0.708	None	0.708
	AlMubasher	0.689	Explicit User Action	0.807
Attractiveness	Absher	0.687	Balance	0.723
	AlMubasher	0.623	Balance	0.756

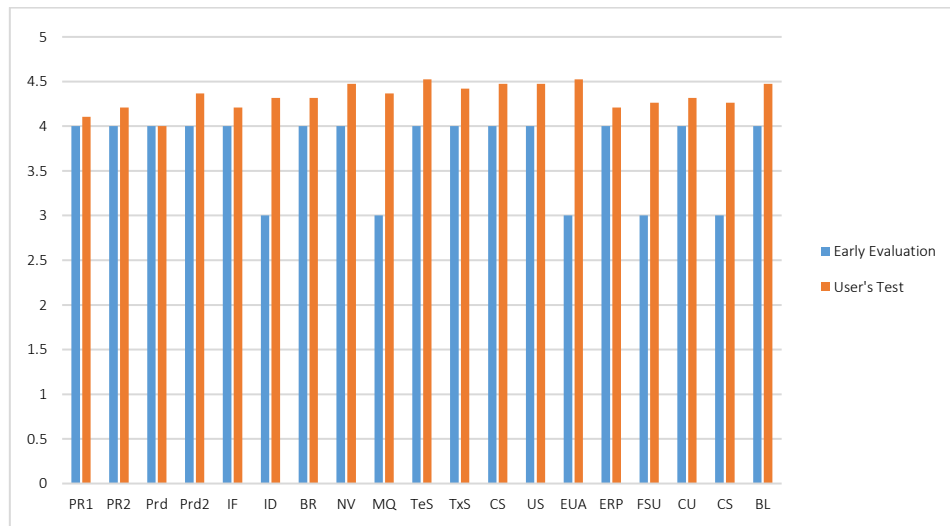


Fig. 5: Comparison of the users' perception with the early evaluation for the Absher apps

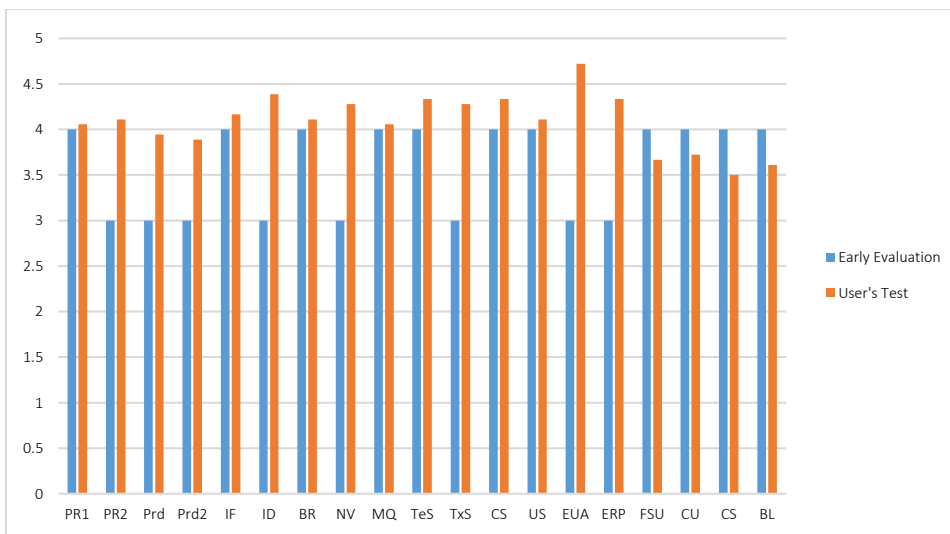


Fig. 6: Comparison of the users' perception with the early evaluation for the AlMubasher apps

Based on the observed values, the following remarks can be identified:

- Users' perception is more constant than the early evaluation, and values are usually close to optimality, especially for the first apps Absher. This can be explained by the fact that these applications are owned by the government and one of the most populated banks in Saudi Arabia: AlRajhi. They continuously receive feedback from their customers/citizens and consider this feedback to improving the usability of their applications.

- Trends are similar for most metrics for the first apps Absher contrarily to the second apps, where half of the metrics are not similar to the users' perception. This can be explained by the fact that all participants are familiar with the Absher apps, which is not the case for the AlMubasher apps since they are not all customers of AlRajhi bank. Thus, they felt more satisfied with the first apps.
- The observed values can be considered as encouraging to build on it and conduct other experiments to improve the study.

The second step in the comparison is to perform the statistical analysis called One Sample T-test. The objective of the One-Sample T-test is to study the comparison in depth. It determines whether the sample mean is statistically different from a known or hypothesized population mean. For our experiment, the sample was composed of the evaluation performed with the 18 subjects, and the population mean was the value obtained by the early evaluation. Since the One-Sample T-test is a parametric test, the first step is to check whether the response variables follow a normal distribution. The SPSS assistant provides a one-sample Kolmogorov-

Smirnov (K-S) test allowing us to test the normal distribution of the response variables. The results show that all the response variables follow a normal distribution.

Table 5 shows the results of the level of significance obtained for each metric in the Absher apps. The null hypothesis can be accepted when the significance level is higher than 0.05. Consequently, we can state that the early usability evaluation fits in with the users' perception for 9 metrics among 19 for the Absher apps: PR1, Prd, Prd2, Br, NV, TeS, UOU, ERP, CU.

Table 5: Significance of the one-sample T-test for the Absher apps

Metric	PR1	PR2	Prd	Prd2	IF	ID	BR	NV	MQ
Sig. (2-tailed)	0.430	0.004	0.381	0.495	0.000	0.000	0.055	0.717	0.000
Metric	TeS	TeX	UOU	UOC	EUA	ERP	FSU	CS	CS
Sig. (2-tailed)	0.104	0.002	0.000	0.331	0.000	0.163	0.000	0.269	0.001
Metric	BL								
Sig. (2-tailed)	0.042								

Table 6 shows the results of the level of significance obtained for each metric in the AlMubasher apps. In this case, 11 metrics among 19 allows us not to reject the null hypothesis. For the

other metrics, the null hypothesis is rejected, and we can state that there is no correspondence between the early evaluation and the users' perception.

Table 6: Significance of the one-sample T-test for the AlMubasher apps

Metric	PR1	PR2	Prd	Prd2	IF	ID	BR	NV	MQ
Sig. (2-tailed)	0.579	0.331	0.749	0.495	0.187	0.004	0.163	0.020	0.331
Metric	TeS	TeX	UOU	UOC	EUA	ERP	FSU	CS	CS
Sig. (2-tailed)	0.010	0.020	0.083	0.163	0.000	0.083	0.010	0.056	0.004
Metric	BL								
Sig. (2-tailed)	0.001								

4.8. Result analysis and discussion

The experiment aimed to test the null hypothesis: there is no significant difference between the usability of a mobile application obtained by our proposal and that perceived by end-users. Based on the results obtained by the experiment, we can conclude that the null hypothesis can be accepted for most of the metrics. For those metrics where the null hypothesis could not be accepted, the following reasons may be in the causes of the problem:

- The indicators used for some metrics (e.g., NV, EUA, FSU) are extracted from related research works that treat other types of applications (desktop, web). Their values may need to be adjusted for mobile applications.
- The indicators of some metrics that are added to the usability model because of their relevance to mobile context (e.g., TeS, TxS, BL) are estimated by analogy to other metrics that are previously validated empirically because of their similarity.
- Some metrics are not valid during the reliability and validity test. This raises a new challenge about identifying other threats that may affect the experiment.

Finally, we can conclude that the experiment was important and allow us to identify the potentialities and limitations of our proposal. The results of the experiment show that it is possible to predict the

usability of a mobile application using conceptual models. Once the indicators are adjusted and validated empirically, we can define an automatic evaluator that determines the usability of a mobile application using the conceptual models. This would be a significant advantage with regard to saving time and resources.

5. Conclusion

This paper presents an empirical evaluation of a model-based approach for mobile usability evaluation. The proposed approach takes the conceptual models as the main input of the evaluation process. The empirical evaluation was carried out through a comparative experimental study where the findings of our proposal are compared to those perceived by end-users. The objects of the experiment were two mobile applications that are largely used in Saudi Arabia. The results of the experiment show that the proposed approach can indeed be a useful complement to standard techniques of usability evaluation early usability evaluation. Besides, the experiment has been a key factor in guiding the improvement of our proposal. As future works, we plan to conduct more experiments with other mobile applications to adjust the indicators and define more realistic ranges based on the users' perception. We also plan to automate the early evaluation process once indicators are improved and develop a tool that

predicts the usability of a mobile application based on the conceptual models that represent such an application.

Acknowledgment

This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project #2017/01/7379.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Abrahao S and Insfran E (2006). Early usability evaluation in model driven architecture environments. In the 6th International Conference on Quality Software, IEEE, Beijing, China: 287-294.
<https://doi.org/10.1109/QSIC.2006.26>
- Abrahao S, Iborra E, and Vanderdonckt J (2008). Usability evaluation of user interfaces generated with a model-driven architecture tool. In: Law ELC, Hvannberg ET, and Cockton G (Eds.), *Maturing usability*: 3-32. Springer, London, UK.
https://doi.org/10.1007/978-1-84628-941-5_1
- Ammar LB (2019). A usability model for mobile applications generated with a model-driven approach. *International Journal of Advanced Computer Science and Applications*, 10: 140-146.
<https://doi.org/10.14569/IJACSA.2019.0100218>
- Ammar LB, Trabelsi A, and Mahfoudhi A (2016). A model-driven approach for usability engineering of interactive systems. *Software Quality Journal*, 24(2): 301-335.
<https://doi.org/10.1007/s11219-014-9266-y>
- Balagtas-Fernandez FT and Hussmann H (2008). Model-driven development of mobile applications. In the 23rd IEEE/ACM International Conference on Automated Software Engineering, IEEE, L'Aquila, Italy: 509-512.
<https://doi.org/10.1109/ASE.2008.94>
- Barros AC, Leitão R, and Ribeiro J (2014). Design and evaluation of a mobile user interface for older adults: Navigation, interaction and visual design recommendations. *Procedia Computer Science*, 27(27): 369-378.
<https://doi.org/10.1016/j.procs.2014.02.041>
- Bouchelligua W, Mahfoudhi A, Mezhoudi N, Daassi O, and Abed M (2010). User interfaces modelling of workflow information systems. In: Barjis J. (Eds.), *Enterprise and organizational modeling and simulation*. EOMAS 2010. Lecture Notes in Business Information Processing, Vol 63. Springer, Berlin, Germany.
https://doi.org/10.1007/978-3-642-15723-3_10
- Fernandez A, Abrahão S, and Insfran E (2013). Empirical validation of a usability inspection method for model-driven web development. *Journal of Systems and Software*, 86(1): 161-186.
<https://doi.org/10.1016/j.jss.2012.07.043>
- Harrison R, Flood D, and Duce D (2013). Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science*, 1: 1.
<https://doi.org/10.1186/2194-0827-1-1>
- Hussain A and Kutar M (2009). Usability metric framework for mobile phone application. In the 10th Annual PostGraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, Liverpool, UK.
<https://doi.org/10.1145/1497308.1497412>
- ISO/IEC (2001). ISO/IEC 9126-1: Software engineering-Product quality-1: Quality model. International Organization for Standardization, International Electrotechnical Commission, Geneva, Switzerland.
- Jindal G and Jain M (2012). A comparative study of mobile phone's operating systems. *International Journal of Computer Applications and Information Technology*, 1(3): 10-15.
- Moumane K, Idri A, and Abran A (2016). Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. SpringerPlus, 5: 548.
<https://doi.org/10.1186/s40064-016-2171-z>
PMid:27190747 PMCID:PMC4851667
- Nayebi F, Desharnais JM, and Abran A (2012). The state of the art of mobile application usability evaluation. In the 25th IEEE Canadian Conference on Electrical and Computer Engineering, IEEE, Montreal, Canada: 1-4.
<https://doi.org/10.1109/CCECE.2012.6334930>
- Panach JI, Condori-Fernandez N, Vos T, Aquino N, and Valverde F (2011). Early usability measurement in model-driven development: Definition and empirical evaluation. *International Journal of Software Engineering and Knowledge Engineering*, 21(03): 339-365.
<https://doi.org/10.1142/S0218194011005311>
- Reis RA, Fontão ADL, Gomes LL, and Dias-Neto AC (2015). Usability evaluation approaches for (Ubiquitous) mobile applications: A systematic mapping study. In the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Nice, France.
- Umuhoza E and Brambilla M (2016). Model driven development approaches for mobile applications: A survey. In the International Conference on Mobile Web and Information Systems, Springer, Vienna, Austria: 93-107.
https://doi.org/10.1007/978-3-319-44215-0_8
- Zhang D and 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