The Usability of Mobile Enterprise Resource Planning Systems

Thomas Wüllerich and Alexander Dobhan

Department of Business and Engineering, University of Applied Sciences Schweinfurt, Schweinfurt, Germany

Keywords: Usability, Usability Evaluation, User-oriented, Mobile Enterprise Resource Planning Systems.

Abstract: This paper presents a model for end-user-based evaluation of the usability of mobile ERP systems. Recent studies show that the mobile use of ERP software is both, crucial for user satisfaction and still improvable for many ERP systems. Therefore, ERP-specific usability models are necessary to meet the requirements of ERP systems in comparison to e.g., apps for private use. The research objective is therefore to develop a model that enables software providers to measure and benchmark the usability of their software products. Therefore, we introduce after a literature research a usability model for the mobile application of ERP systems (mobile ERP). Our usability model is based on the widely used PACMAD model. We modify the PACMAD model for the context of ERP systems. This results in a new end-user-based model, that differs from existing models, because of its focus on end-users and the ERP context. Subsequently, the model will be tested in an initial study with 19 test persons. The results of the study indicate two main findings. Firstly, the model allows the measurement of the usability of mobile ERP systems. Secondly, some key factors substantially affect the usability of mobile ERP systems.

1 INTRODUCTION

The proliferation of smartphones and tablets has released a discussion on mobile flexibility of business applications, and in particular of enterprise resource planning (ERP) systems (Bahssas et al., 2015; Omar & Gomez, 2016; Tai et al., 2016). Today mobile ERP applications are now part of the ERP standard. Providers of ERP systems are responding to user feedback by increasing the suitability of their software for mobile devices. This enables, for example, sales representatives to access internal data (Bahssas et al., 2015). At the same time, users still see a need for improvement in this area among software providers (Trovarit AG, 2019).

Regardless of whether mobile use is realized as a native application, web application, or hybrid application (Gronau & Fohrholz, 2016), recent research outcomes show that users, who complain about the usability of ERP systems, often reduce overall user satisfaction. The reason for the negative usability perception is the complex and static way of operation (Omar, 2015; Trovarit AG, 2019) caused by a large amount of data and complex functionalities. (Omar et al., 2016).

On the other hand, usability challenges refer to the mobile use of ERP systems, such as limited screen size, the reduced reliability of the mobile data connection, and other aspects. The complexity of ERP systems, which in the past required specific usability approaches for desktop applications (Singh & Wesson, 2009), makes a specific usability model necessary for mobile ERP systems as well (Omar & Gomez, 2017).

Therefore, researchers have already developed usability models for mobile ERP systems. However, these are mostly expert based, while no studies have been conducted from the perspective of end-users (Omar et al., 2016). Therefore, the intention of our research in this paper is to develop an end-user-based model for evaluating the usability of mobile ERP applications.

Thus, the next paragraph contains an overview of recent research outcomes. Subsequently, we introduce a new usability model based on PACMAD. The applicability of the new model is then tested based on an initial study with 19 participants for two ERP systems. The results of the study are described. The results include some indications for interdependencies between the components of our model.

2 LITERATURE REVIEW

The common use of mobile devices leads to usability challenges such as limited screen size, different
screen resolutions, limited processing, and performance capabilities, limited data entry methods, the diversity of mobile operating systems and security in the mobile ERP context (Omar, 2015; Zhang & Adipat, 2005). Another factor is the mobile environment, as interaction with environmental elements causes distraction (Zhang and Adipat, 2005). Mobile connectivity is often a critical feature (Muccini et al., 2012). Furthermore, different levels of end-user knowledge should not be neglected (Nayebi et al., 2012). Another usability challenge refers directly to the back-end ERP system. Compared to other mobile applications, mobile ERP systems process a large amount of data (Omar et al., 2016).

The PACMAD model of Harrison et al. (2013) is one of the most applied usability models for mobile applications. It considers the specific requirements of mobile devices and offers sufficient leeway for adaptations. PACMAD means “People At the Center of Mobile Application Development” and is based on the approaches of Nielsen and ISO 9241-11. The model identifies the three factors user, task, and context that influence the usability of an application. Furthermore, the model has the following seven dimensions (Harrison et al., 2013):

Effectiveness examines the ability of a user to complete a certain task. It is measured by successful task completion (Harrison et al., 2013; Alturki and Gay, 2017). It has been applied similarly several times in a similar way (Alturki & Gay, 2017; Frokjaer et al., 2000).

Efficiency measures the ability of a user to perform tasks with the desired speed and accuracy (Harrison et al., 2013). One of the measurable indicators is the time of completion (Alturki & Gay, 2017; Cooprider et al., 2010). To calculate Efficiency, time is put in relation to Effectiveness.

Satisfaction examines the perceived level of comfort and friendliness of the system (Harrison et al., 2013; Frokjaer et al., 2000; Omar, 2015). This is measured using a questionnaire or other qualitative techniques such as emoji cards (Harrison et al., 2013).

The Errors dimension is used to determine the error rate during use (Nielsen, 1994). In practice, it is measured with the number of errors (Hussain et al., 2018). According to PACMAD, Memorability is the ability of a user to maintain the effective use of an application and avoid repeated learning (Harrison et al., 2013). This can be determined by repeated sessions after a period of inactivity (Zali, 2016) or by using a questionnaire (Hussain et al., 2018). This dimension is like Learnability, which is defined as an experience that the user can gain.

The Cognitive Load refers to the amount of cognitive processing the user can perform (Harrison et al., 2013). This is measured with the use of eye-tracking technology or a NASA TLX test (Alturki and Gay, 2017).

The analysis of usability studies on mobile ERP systems showed that no end-user-oriented approach to usability evaluation exists yet.

3 USABILITY MODEL MERP-U

Our model aims to adapt the PACMAD model to mobile ERP systems. In the first step, the dimensions of the basic PACMAD model are partially summarised and operationalized for the mobile ERP context. After that supplementing, we add further dimensions that are required for the end-user-oriented assessment of mobile ERP systems.

To calculate the Effectiveness (1), the share of successfully completed tasks is quantified. The outcome of the following formula indicates the level of Effectiveness:

\[
\text{Effectiveness} = \frac{\text{number of tasks completed successfully}}{\text{total number of tasks undertaken}} \times 100\% \quad (1)
\]

In the context of ERP systems, the correctness of data maintenance or data extraction is easy to quantify for simple tasks. The number of errors is a measure for the Errors dimension. For the present model, the Effectiveness and the Errors are put into a ratio.

\[
\begin{align*}
\text{Efficiency} &= \frac{\sum \text{number of correct tasks}}{N \times B} \times 100\% \\
\text{B} &= \text{number of users} \\
N &= \text{number of tasks}
\end{align*}
\]

To determine the Efficiency (2) of mobile ERP systems, the time it takes to complete the test, is recorded. Then we take the proportion of correct tasks in relation to the time spent to get the Efficiency as an outcome.

Previous studies on user Satisfaction use a questionnaire with a 5-point Likert scale to measure it (Hussain et al., 2018; Alturki and Gay, 2017; Omar, 2018).

For the model, Learnability is combined with Memorability. The measurement requires the record of the time to reach and maintain a given level of competence (Harrison et al., 2013). It is integrated...
into the questionnaire and estimated by the users themselves.

In the context of the study, the NASA TLX test is an indicator for the Cognitive Load (Harrison et al., 2013; Alturki and Gay, 2017). It relies on a multidimensional construct to derive an overall workload score based on a weighted average of ratings on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration level (Cao et al., 2009).

According to Singh and Wesson, the criteria for the usability of ERP systems are limited due to the small number of corresponding studies. However, they distinguish between six specific basic criteria for assessing usability (Singh and Wesson, 2009):

- Navigation: navigational functions of the ERP system
- Learnability: the degree of learnability of the ERP system
- Task support: the ability of the ERP system to provide effective task support
- Presentation: presentation capabilities of the ERP system
- Customization: the ability of the ERP system to adapt to a specific organization and individual user

From these heuristics defined by Singh and Wesson, we select two dimensions to expand the operationalized PACMAD model. Learnability is not included because it is already considered in combination with Memorability. Furthermore, the heuristic Task Support is included in Effectiveness and Efficiency. The evaluation of Customization does not take place within the scope of the study. Instead, the dimension Presentation considers the main problems of ERP systems: The complex screen display and outputs that are often difficult to understand (Singh and Wesson, 2009). Mobile applications are particularly affected by this due to the smaller display (Omar et al., 2016). The dimension aims to ensure that the layout of menus, dialogue boxes, controls, and information on the screen is appropriate and clear (Singh & Wesson, 2009). Another important dimension for the mobile ERP context is Navigation, as this is a design issue for ERP systems. This examines the application’s ability to access appropriate information, menus, reports, options, and elements (Singh and Wesson, 2009). Both, Presentation and Navigation are subjective criteria of the questionnaire. Along with Satisfaction participants estimate these dimensions by using a 5-point Likert scale.

The modified PACMAD model, MERP-U, is an evaluation model that is especially suitable for mobile ERP systems. It combines the best of both, ERP usability models (Singh and Wesson) and usability models for mobile apps (PACMAD). In summary, the new model includes seven dimensions (see Fig. 1).

4 USABILITY STUDY

For testing the applicability of our model, we applied it for two ERP systems in an initial study.

4.1 Research Design

The study was about two ERP software products available on the European market. System A is a relatively new ERP product, still unknown on most European markets, but established on the eastern European market with growing sales figures. It specializes in production-oriented small or medium-sized businesses. In contrast, System B is a mature ERP software for SMEs that is established on the worldwide market. Both systems are web-based. For System B an app for mobile devices with Android and iOS is available. System A works with all common browsers.

The Participants of the study were 19 research assistants, student assistants as well as Bachelor and Master Students with professional and academic ERP experience. All participants were able to operate in an ERP software system before and to understand its basics. Thus, according to Hofer et al. (2007) the crucial prerequisite for realistic end-users is fulfilled. Nevertheless, it must be considered that the participants have a different level of knowledge regarding the systems.
To record usability-related information, the study includes the application of various methods, which are:

- The (digital) test sheet (SurveyMonkey) contains simple tasks in the mobile ERP system (e.g., changes in the article master). For usability recording, answers on the test sheet and the modified data records in the system were evaluated for correctness and number of errors. The results represent the indicators for the PACMAD dimensions Effectiveness, Errors, and Efficiency. During processing the tasks, a study coordinator observed the behaviour of the participants. He logged the comments of the users and the user’s reactions. The gained observations mainly refer to Satisfaction, Cognitive Load Navigation, and Presentation.

- After the users completed the test sheet tasks, they were asked to fill out some questions on Navigation, Presentation, and Satisfaction with a 5-point Likert scale. In addition, the participants assessed the Memorability and Learnability. Furthermore, they had the opportunity to express further comments. The questionnaire also contains an extended NASA TLX test.

For organizational reasons, participants were assigned to two groups. In the case of the first group, an online meeting took place via Zoom. The study coordinator was available to the subjects during this time for questions and acted as an observer. Group 1 mainly consisted of experienced users.

The members of the second group received their credentials and access to their tenant (each user has their client in both systems). The users also received all instructions and links required for the test by e-mail. The second group primarily included users with less experience.

At the beginning of the individual online meetings, the participants received a short introduction to the study. Any questions were clarified in advance so that the first digital test sheet could be used. The participants worked through the tasks of the test sheet on their smartphones. Additionally, they answered the questions on the test sheet. The study coordinator was available for any questions and recorded comments, reactions, and other conspicuous features.

After completion (about 10 minutes), the participants received a feedback sheet on the ERP system in the chat. As soon as this was completed, the participants turned to the second system in the same way. In the end, a summary feedback discussion took place.

Participants of group 2 received a more detailed introduction and worked through the process steps according to the above structure except for observation and feedback discussion.

4.2 Results

A total number of 19 participants completed the study. 8 of these worked through the questions in a live meeting and 11 completed the tasks outside of an online session. A total of 18 feedback questionnaires were filled out completely.

For Effectiveness and Errors, we have achieved the following results:

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness</td>
<td>77%</td>
<td>63%</td>
</tr>
<tr>
<td>Errors</td>
<td>23%</td>
<td>37%</td>
</tr>
</tbody>
</table>

System A seems to have a higher Effectiveness and a lower Error rate than System B. According to the results, users completed on average one task more correctly in System A than in System B. It turned out that incorrect spelling, forgotten data, or a lack of available (in ERP system) information mostly caused the errors.

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>62%</td>
<td>39%</td>
</tr>
</tbody>
</table>

The results regarding the Efficiency show a better Efficiency for System A. This is related to the fact that the processing time for System A was on average 1.72 minutes shorter than for System B (even though the effectiveness was better).

To capture the Satisfaction of the users with the system, participants were supposed to rate the following statement on a Likert scale: "The application is satisfactory overall".

The outcome on that is given below:

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree at all</td>
<td>0,00%</td>
<td>0,00%</td>
</tr>
<tr>
<td>Disagree</td>
<td>16,67%</td>
<td>22,22%</td>
</tr>
<tr>
<td>Part/part</td>
<td>16,67%</td>
<td>27,78%</td>
</tr>
<tr>
<td>Agree</td>
<td>50,00%</td>
<td>44,44%</td>
</tr>
<tr>
<td>Clearly agree</td>
<td>16,67%</td>
<td>5,56%</td>
</tr>
</tbody>
</table>

When comparing the results regarding the two ERP applications, it is obvious that users are more satisfied with System A than with System B.
The answers regarding the statement about Learnability and Memorability give the following picture: “The functions and steps of the application are memorable and easy to learn”.

Table 4: Results of the Memorability and Learnability.

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree at all</td>
<td>11,11%</td>
<td>5,56%</td>
</tr>
<tr>
<td>Disagree</td>
<td>0,00%</td>
<td>11,11%</td>
</tr>
<tr>
<td>Part/part</td>
<td>0,00%</td>
<td>22,22%</td>
</tr>
<tr>
<td>Agree</td>
<td>38,89%</td>
<td>38,89%</td>
</tr>
<tr>
<td>Clearly agree</td>
<td>50,00%</td>
<td>22,22%</td>
</tr>
</tbody>
</table>

The distribution of results between the two systems is slightly similar. With System B, most users consider the software as easy to learn and memorable. However, compared to System A, the application scores lower. 89% of respondents found System A to be good to very good. 11%, on the other hand, were very dissatisfied.

For measuring the Cognitive Load, we applied the NASA TLX test. The table below contains the test results:

Table 5: Results of the Cognitive Load.

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental effort</td>
<td>Not at all (1,47)</td>
<td>Little (2,29)</td>
</tr>
<tr>
<td>Physical effort</td>
<td>Not at all (1,06)</td>
<td>Not at all (1,37)</td>
</tr>
<tr>
<td>Time pressure</td>
<td>Little (1,71)</td>
<td>Little (2,35)</td>
</tr>
<tr>
<td>Satisfaction of performance</td>
<td>Satisfied (4,11)</td>
<td>Neither (3,27)</td>
</tr>
<tr>
<td>Performance level</td>
<td>Less hard (1,65)</td>
<td>Less hard (2,16)</td>
</tr>
<tr>
<td>Stress level</td>
<td>Little (1,82)</td>
<td>Little (2,37)</td>
</tr>
</tbody>
</table>

Most of the test persons feel a lower strain accompanied by a higher self-satisfaction for System A compared to System B. The greatest difference arises in the mental strain and satisfaction experienced by the participants during the processing.

To record Navigation in the questionnaire, we let the participants rate the following statement: "I like the menu navigation and menu structure of the application".

The table below shows the results on Navigation:

Table 6: Results of the Navigation.

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree at all</td>
<td>16,67%</td>
<td>11,11%</td>
</tr>
<tr>
<td>Disagree</td>
<td>5,56%</td>
<td>33,33%</td>
</tr>
<tr>
<td>Part/part</td>
<td>27,78%</td>
<td>27,78%</td>
</tr>
<tr>
<td>Agree</td>
<td>27,78%</td>
<td>22,22%</td>
</tr>
<tr>
<td>Clearly agree</td>
<td>22,22%</td>
<td>5,56%</td>
</tr>
</tbody>
</table>

System A performs better overall because the users rate the menu navigation and the menu structure better than for System B. The conversations and comments of the test persons confirm that result. The structure of System A is less extensive and the menu navigation with its subdivision is logically arranged. However, some users do not find their way around the menu navigation. This is due to the poor quality of the translation. The poorer tendency for the menu structure of System B is mainly due to the extensive design of the application and the ease of finding functions. This has a negative impact on the search for information.

From the observation and the interviews as well as the comments in the questionnaire, factors for the Navigation were also identified:

Table 7: Factors for the Navigation.

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ drop-down menu</td>
<td>+ successful design</td>
<td></td>
</tr>
<tr>
<td>+ simple structure</td>
<td>+ good scaling</td>
<td></td>
</tr>
<tr>
<td>+ confirmation of transactions</td>
<td>+ mature system</td>
<td></td>
</tr>
<tr>
<td>- not mature, many errors</td>
<td>- overloaded input fields</td>
<td></td>
</tr>
<tr>
<td>- poor translation</td>
<td>- no transaction confirmation</td>
<td></td>
</tr>
<tr>
<td>- poor scaling on a smartphone</td>
<td>- modules difficult to find</td>
<td></td>
</tr>
</tbody>
</table>

The statement for Presentation is as follows: “The application is designed clearly”.

The users of both systems gave the following feedback on this:

Table 8: Results of the Presentation.

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree at all</td>
<td>0,00%</td>
<td>5,56%</td>
</tr>
<tr>
<td>Disagree</td>
<td>22,22%</td>
<td>50,00%</td>
</tr>
<tr>
<td>Part/part</td>
<td>27,78%</td>
<td>16,67%</td>
</tr>
<tr>
<td>Agree</td>
<td>27,78%</td>
<td>16,67%</td>
</tr>
<tr>
<td>Clearly agree</td>
<td>22,22%</td>
<td>11,11%</td>
</tr>
</tbody>
</table>

Users prefer the clarity of System A to System B. The negative user ratings are justified by the fact that the format of System A as a web application is not scalable to smartphones. Nevertheless, some users praise the clear presentation of the individual areas and submenus. In contrast, the scaling of System B is adapted to the smartphone. On the other hand, it is a challenge for most users to obtain an overview to find the desired data.
5 INTERPRETATION OF THE RESULTS

The results of the study show that System A performs better in all usability dimensions. This may be an indicator of how the individual dimensions effect on each other. This means that the positive or negative result of one usability attribute has a corresponding effect on another attribute. For this purpose, the relationships of the dimension are presented taking into account the external influences identified by Omar (2015).

Here, the user forms an opinion about the menu navigation and the presentation of the content. If users perceive these aspects as positive, they feel low Cognitive Load. Because of the low Cognitive Load, there is a high level of self-satisfaction and satisfaction in general. Furthermore, Memorability and Learnability benefit from this, so that data can be learned and memorized more easily. This in turn promotes Effectiveness and Efficiency in the decision (Akiki et al., 2016). The starting points of the effect chain seems to be the newly introduced usability attributes Navigation and Presentation (Babaian et al., 2016).

The reasons for the evaluation of the two new dimensions are the interaction between the software architecture and the personal attitude of the user towards the architecture (Dabkowski & Jankowska, 2003). Therefore, users mostly justify their satisfaction with arguments that evaluate Navigation and Presentation. In between, is the Cognitive Load, which has been shown to influence Satisfaction (Schmutz et al., 2009). Nevertheless, it should be noted that the external challenges identified by Omar (2015) also affect the dimensions within the back-end ERP system.

In the present study, for example, the clarity of System B was criticized by the test persons due to the high depth of functions and the numerous options in the input menus. This had a negative impact on user satisfaction and usability (Singh and Wesson, 2009; Omar, 2015). In a Trovarit study, a similar correlation was found between the range of functions and satisfaction, which reinforces the research conducted.

It was found out that ERP solutions with fewer features and applications from smaller vendors scored best in user satisfaction (Trovarit AG, 2019). However, despite the higher usability of small ERP systems, the decision factor for companies to purchase them is not only usability. These would be, for example, the desire for a high level of functional depth. This is especially true for larger companies, with a large amount of information (Omar and Gomez, 2016).

6 LIMITATIONS OF RESEARCH

For this study, we designed two different settings. The advantages of the online meeting were the direct feedback from the users during the conversation. This was not possible for participants outside of the online session. However, the time flexibility was an important factor for these users to participate in the study. Originally, live conduction with the support of eye-tracking was planned for the research design. This plan was discarded due to the COVID 19 crisis, so an alternative design without the physical presence of the participants was intended. With the new concept, it was consequently no longer possible to use the eye-tracking system to test the Cognitive Load.

![Dimensions of MERP-U](image-url)
Due to the use of digital media, the new approach represents a modern alternative to the classic study with the presence of the participants (Gray et al., 2020).

Nevertheless, this approach brings challenges compared to live execution. On the one hand, it happened that tasks were not fully completed by the users. Furthermore, there is the risk that participants cheat by giving false information when completing the test sheets. The absence of errors was not directly observed.

Another point of criticism is that test persons in the study were partly familiar with one of the systems. Of the 19 users, 15 had practical experience in System B, while one person was familiar with System A. Three subjects only had experience in other ERP systems before the study. Previous knowledge on performing tasks on a particular system can significantly influence the results. It is to be expected that due to experience effects, the known system B performs better. In this study, no positive tendency towards the familiar system was noticeable. This strengthens the result that the usability of system A is higher than that of B.

Likewise, the system sequence during the study should be considered, as the learning effect of the first application probably had a positive effect on the result of the second system. Of the 19 participants, eleven started first with System A and eight with System B. Although System B was mostly used second, it performed worse in Effectiveness and Efficiency. Thus, no clear correlation can be discerned here.

Compared to other studies, the number of participants in this study is lower, at 19 users (Frojkaer et al., 2000; Raptis et al., 2013). This is because this first study was a test of applicability which already gives us interesting insights. Furthermore, the results motivate us to plan a study with far more participants. This study will be designed to examine our hypothesis that the dimensions Navigation and Presentation affect all other usability dimensions.

7 CONCLUSION

In summary, this research aimed to develop an end-user-oriented model for the usability evaluation of mobile ERP systems. This was achieved with the MERP-U model. The interpretation of the study results indicates that there are correlations between the dimensions within the model and that a high functional depth has a negative effect on usability.

Furthermore, the results allow us to suppose that the dimensions Navigation and Presentation are of high importance for mobile ERP systems.

As we emphasized in the paragraph regarding limitations, the sample size was small compared to other studies. Therefore, for future research, we are planning to apply the model to further mobile ERP applications with a higher number of participants. The use of techniques such as eye-tracking or web augmentation is also recommended for analysing aspects like Navigation, Presentation or Cognitive Load. This would allow us to automate the process partly, with the possibility of expanding the analysis to a higher audience. For that study, it makes sense to slightly modify the research design to examine the interdependencies between the dimensions.

REFERENCES


94). IEEE. https://doi.org/10.1109/DEXA.2003.123

524