Master's thesis
Master's Programme in Computer Science

Improving ERP system usability for novice and expert users

Jarmo Kallio

June 10, 2021

Supervisor(s): Mikko Kytö

Examiner(s): Giulio Jacucci

University of Helsinki
Faculty of Science

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Computer Science

Tekijä — Författare — Author

Jarmo Kallio

Työn nimi — Arbetets titel — Title

Improving ERP system usability for novice and expert users

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Master's thesis	June 10, 2021	50

Tiivistelmä — Referat — Abstract

Despite benefits and importance of ERP systems, they suffer from many usability problems. They have user interfaces that are complex and suffer from "daunting usability problems". Also, their implementation success rate is relatively low and their usability significantly influences this implementation success. As a company offering an ERP system to ferry operators was planning to renew the user interface of this system in future, we investigated usability of the current system so this could guide future implementation of the new user interface. We studied new and long time users by conducting sessions where the users told about their experiences, performed tasks with the system and filled usability questionnaire (System Usability Scale).

Many novice and long time users reported problems. The scores from usability questionnaire show all but two participants perceived the usability of the system as below average and in adjective rating "not acceptable". Two users rated the usability as "excellent". We reasoned that there could be a group of users who use the system in such a way and in such context that they do not experience these problems.

The results indicate novices have trouble, for example, navigating and completing tasks. Also some long time users reported navigation issues. The system seems to require that it's users remember lots of things in order to use it well. The interviews and tasks indicate the system is complex and hard to use and both novices and experts face problems. This is supported by perceived usability scores. While experts could in most cases finish all tasks, during interview some of them reported problems such as finding products the customers needed, error reporting being unclear, configuration being tedious, and need for lots of manual typing, for example.

We gave recommendations on what to consider when implementing new user interface for this ERP system. For example, navigation should be improved and users should be provided with powerful search tools. ERP usability is not studied much. Our study supports use of already developed heuristics in classifying usability problems. Our recommendations how to improve usability of the ERP system studied should give some guidelines on what could be done, although not much is backed by laboratory studies. More work is needed in this field to find and test solutions to usability problems users face.

ACM Computing Classification System (CCS):
Human-centered computing → Interaction design → Interaction design process and methods → User centered design

Avainsanat — Nyckelord — Keywords

ERP Systems, Usability, Perceived ease of use

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

1. Introduction

Enterprise resource planning (ERP) systems have been defined as integrated business information systems having the purpose of integrating and managing different business processes and information flows within an enterprise [36]. Infrastructure of ERP system is built upon a database containing data from business processes essential for business operations and decision making [34].

ERP systems have proved useful in many organizations [38]. They improve organizations operational efficiency [36] and are beneficial to business [28]. They are seen as "strong tool that help organizations succeed and rise" [41]. According to Scholtz et al. [34] organizations have need to move to systems that handle the whole business process, not only parts of it. ERP systems are critical here.

Despite benefits and importance of ERP systems, they suffer from many usability problems. Potential users may not use them [9]. These systems have user interfaces that are complex [36][28][37] and suffer from "daunting usability problems" [38]. They are difficult to use [37]. Also, their implementation success rate is relatively low and their usability significantly influences this implementation success [41]. Many ERP projects have struggled to achieve benefits of these systems due to complex user interfaces [34].

As a company offering an ERP system to ferry operators was planning to renew the user interface of this system in future, we investigated usability of the current system so this could guide future implementation of the new user interface. New users had been experiencing problems learning the system. At the same time it was assumed that expert users could use the system efficiently and were satisfied with it. We conducted user study to access the current situation and to then give recommendations: what should be taken into account usability wise when implementing the new user interface. For example: how could new users learn the system faster while it would still remain fast for expert users? The study was conducted by first interviewing 14 users and asking open ended questions relating to using, teaching and learning the system. We also asked them to perform tasks with the system and to fill a usability questionnaire.

In this study we consider usability problems from the perspective of ERP systems. We discuss related work in chapter 2, this includes approaches to measure ERP system usability and improve it. In Chapter 3 we present design for our research. We describe the procedure used and our participants, discuss interview questions used, practical tasks and usability questionnaire. We present results of this study in Chapter 4, these include results from coding interview transcripts and classifying usability problems, task times and completion rates, and results of usability questionnaire. We discuss usability problems found and give recommendations for implementing the new user interface in Chapter 5. We also discuss the effect COVID19 had on our study. We discuss limitations of this study in Chapter 6 and draw conclusions in Chapter 7. Interview questions, script for practical tasks and System Usability Scale questionnaire can be found in appendices A, B and C, respectively.

1.1 Research Questions

Initial discussions with company personnel revealed there had been complaints from clients that new users struggled with the ERP system. Long time users, we were told, were efficient and could use the system without much trouble. Based on this we decided to include both novice and long time users as participants in our study. We formulated the following research questions:

- **RQ1** What usability problems novice and expert users face in current user interface?
- **RQ2** What could be done to the problems found when designing new user interface?

2. Related Work

2.1 Usability

In 1994 Nielsen [22] defined term usability as consisting of measurable components *learnability*, *efficiency*, *memorability*, *errors* and *satisfaction*. He pointed out that usability should be measured relative to certain users and tasks. This is important as different users and tasks could lead to different measurement outcomes for same system. A more recent definition is found in standard ISO 9241-11, which defines usability in terms of how well specified users with certain goals and in certain context can use a system, product or service with effectiveness, efficiency and satisfaction [6]. According to [30] this standard provides internationally accepted definition for usability in many fields. However, Hornbæk [13] points that usability cannot be measured directly but requires operationalization, and this choice of measures "raises the question if that which is measured is a valid indicator of usability".

How is usability measured then? In 2014 Freddy Paz et al. made a systematic review of usability evaluation methods [29]. *Usability testing*, *questionnaires*, *heuristic evaluations* and *interviews* were some of the more often used methods. Here *usability testing* refers to procedure where users perform a set of tasks with the system under study. *Questionnaires* refer to users providing answers to set of questions. It is required that user first had interacted with the system under study. *Heuristic evaluations* consist of a group of usability experts reviewing the system according to "established usability principles". *Interviews* consist of discussions with users to evaluate interface.

2.2 ERP system usability

2.2.1 User satisfaction and technology acceptance

Studies on ERP system user satisfaction include surveys [9] and [41]. In [9] from 2004 Calisir et al. studied usability factors that affect users satisfaction with ERP systems. This was done by forming a questionnaire based on usability literature and recommendations from academics. This questionnaire was used in a survey in whitch 51 end users participated. According to the authors this research indicated that perceived usefulness and learnability affect the satisfaction of users. [41] is a more recent study in user satisfaction where Yassien et al. studied the impact of users satisfaction on ERP project implementation success. This was done by sending questionnaires to 106 persons using or being affected by ERP systems. The authors concluded that usability of a system influences success in implementing ERP project and satisfaction has a mediating role between usability and this success. Here satisfaction refers to *user information satisfaction* (UIS) which has been defined as "the extend to which users believe the information system available to them meets their information requirements" [14].

Scholtz et al. studied the relation of technology acceptance and usability [35]. They conducted a survey of 112 ERP users. The authors found that usability of an ERP system "affects attitudes and intentions to use ERP software". Navigation, presentation and leanability were the usability aspects measured. With these, technology acceptance was measured with related constructs *perceived usefulness*, *perceived ease of use*, *attitude towards usage* and *behavioral intention to use*.

2.2.2 Usability problems and evaluation

Part of the research has focused on developing evaluation criteria for ERP systems and identifying usability problems they have. In 2009 Singh et al. developed five ERP system specific usability criteria or heuristics [36]. These can be found in first part of table 4.2. The authors arrived in these by reviewing multiple usability studies, such as [38]. These criteria were verified in a case study where they were used to identify usability problems with an ERP system. Although highly cited compared to many ERP usability studies, the case study has received criticism from [11] for having only a small number of experts conducting evaluations with the heuristics.

Scholtz et al. [33] studied usability issues of a medium-sized ERP system in context of ERP training in university. The authors employed survey and case study strategies to find out, for example, suitable usability criteria for usability evaluation during learning process. The authors found that *navigation*, *presentation* and *learnability* could be used in evaluation.

In 2014 Lambeck et al. conducted a survey to determine if usability problems identified in past years were still relevant in ERP systems [18]. 184 ERP users participated. The authors concluded that usability problems still existed but were not as critical. To find out whether users in different countries face similar usability problems with ERP systems, Lambeck et al. conducted a comparative user survey for ERP users in Germany and Latvia [16]. No significant differences was found between the countries. The Authors note, however, that the results could not be generalized to all countries at that point.

Choma et al. [11] developed *perspective-based ERP heuristic* based on Nielsen's heuristics [24] and tested them by asking participants perform usability inspection with them. Here word perspective means that one can for example evaluate systems error prevention ability from *perspectives* of presentation or task support. They also added tips or instructions to guide the evaluator. This new approach was tested against evaluation with Nielsen's heuristics. The authors reported the new approach was better for prototype evaluation because of the tips they had added.

Veneziano et al. [39] has investigated how ERP users demographic information influences their rating of these systems. They used *system usability scale* [7] in a survey to get participants usability score for an ERP system. Demographic information used included participants educational level and job experience. The authors conclude that users cognitive diversity should be taken into account in design and evaluation of systems.

Topi et al. in [38] studied usability issues by conducting ten in-depth interviews with ERP system users. A qualitative analysis of the interview transcripts led them to categorize found usability issues in following categories: *Identification of and access to the correct functionality*, *Transaction execution support*, *System output limitations*, *Support in error situations*, *Terminology problems* and *Overall system complexity*. The authors pointed that the problems found can be seen in light of *collaboration theory* as non-collaborative behaviour from the system. They discussed central principles of *collaborative behaviour*: *commitment to mutual support* requires helping other party with task when recognizing its need for it, *commitment to joint activity* requiring that both parties must "be aware of context surrounding their collaboration" and *mutual responsiveness* meaning that collaborating parties adjust their behavior based on behavior of other party.

Oja and Lucas [25] have introduced *critical incidents* method. Part of the motivation was to be able to provide a base line method against which they could evaluate "collaborative critique" method under development. The authors defined *critical incidents* as "serious breakdowns in human-computer interaction" that the users encounter with a system. The method consisted of using a system and reporting these critical incidents as they occurred, and going through the incidents with expert after the use session. This is combined with expert going through the material and reporting what they thought were the critical incidents.

A laboratory study was conducted to test the method. The authors saw it having the strength that, compared to interviews it relied less on users memories and compared to just expert evaluations it provided more information about usability problems.

While showing that the method could uncover many usability issues, it was not compared to other methods in a controlled experiment.

The authors discuss the same topic in [26] and argue that combining user and expert reported issues ensures that identified problems are those that actually affect users. According to the authors the problem with just experts evaluating systems is that identified problems are only *potential*.

A modified version of the critical incidents method [25] was used by Babaian et al. [1] to benchmark usability problems with a system. The authors introduced a new method called *collaborative critique* for accessing how well a system is able to reduce user's cognitive and physical effort while it's being used. It is a cognitive walkthrough method that asks evaluators to answer collaboration related questions when performing tasks with system under study, these try to identify if the system fails to collaborate. In laboratory study the authors found that this new method could predict around 70% of certain usability issues identified in benchmark, making the method promising. Collaborative critique method is further discussed in [19].

2.2.3 Improving ERP system usability

Various suggestions and designs have been made to improve usability of ERP systems. In his 2008 white paper Matthews [20] discusses his ideas for improving ERP usability such as web-like navigation and good searching functionalities.

One approach studied is *adaptive user interfaces*. Eichler et al. [12] describe an adaptive user interface as one that dynamically changes "in order to conform to users needs and usage habits". They created a mockup that had a side bar from which users could find often used items. Also their mockup included *ephemeral visualization* meaning they would animate transaction screens so they would first show only the most often used fields and then all other would start to appear. The authors discussed problems with destructively modifying user interfaces such as having problems getting help when everyone's screen looks different. Their mockup did not have these problems since it had the adaptive recommender part and rest of the user interface remained static. The mockup was for SAP ERP system and the authors performed a brief questionnaire study asking SAP users their opinions about it. According to the authors this study shoved positive results but the amount of participants was small and further study would be needed. For example, whether this kind of user interface could improve task times or navigation.

Another approach, Lambeck and Groh have introduced a concept of *scalable user interface* to "reduce the problem of user guidance and overall system complexity" [17]. In that article they argue that users of different ERP systems have commonalities in their objectives and go on to describe an abstract interface model. In this model the user interface has three layers corresponding users objectives: starting from *orientation and overview* with small amounts of information to *searching and filtering*, where users can compare domain content and find business item he needs, to *editing and execution* with high amount of information and ability to edit selected item.

There is also approach to improve ERP usability by means of improving *collaboration* between ERP systems and their users. This approach was first suggested by Topi et al. in [38]. Based on this human computer collaboration view, Babaian et al. have derived a set of four design principles to guide ERP system design [2]. As discussed in [4] these center on customizing vocabulary, providing navigation and progress guidance, identifying and solving errors, and presenting options to users.

For promoting collaboration between system and user, Babaian and Lucas have presented a way to guide users with tasks by visualizing previous successful task completions to users [3]. These visualizations are based on log data. The authors introduce *Task-Interface-Log (TIL) model* to record and reconstruct interactions. In [4] Babaian et al. discussed their prototype implementation that helped users in the ways of process visualization and playback of previously performed tasks. The prototype used *TIL model* from [3] and rendered business process as interactive graph. No user studies were yet performed. In [5] Babaian et al. discuss laboratory evaluation of the automated playback and process visualization components. According to the authors, experiments show these components helped users learn perform tasks and see how those tasks "fit within the encompassing business process".

As a theoretical development of collaboration view, Xu et al. have proposed a conceptual model for successful user system collaboration in [40]. The authors discuss theoretical considerations when one applies collaboration theory to human computer collaboration. The authors do not provide usability guidelines but suggest use of *collaborative capabilities* that they discuss as "a foundation for practical design principles".

2.2.4 Complexity, task time and completion

Parks [28] has studied how complexity of user interface affects task completion and time in ERP setting. Users performed tasks on two different versions of user interface, one being default interface on an ERP system and other a simpler reworked version. Complexity was measured by human factors models. The author found that task time, not task success, was significantly related to interface complexity. Task success on the other hand was dependent on whether users scanned and verified data before clicking a save button. Parks discussed the importance of interface customization in ERP setting as the default interface contained elements that hindered task completion. The author points how also novice users would benefit it the interface had only needed elements.

The human factors models that Parks used, namely *GOMS-KLM* and *visual complexity*. GOMS-KLM [15] can be used to estimate how long it would take expert user to complete a task. It simply breaks tasks into small parts like key presses and assings times for them. According to Parks, visual complexity can be calculated by counting all objects and vertical and horizontal alignments on a screen and summing the amounts.

2.2.5 Error messages

Sadiq et al. [31] have studied problems with ERP system error messages by conducting evaluations using Nielsen's heuristics [24] and a questionnaire. They collected five error messages from an ERP system and showed study participants the screens where these happened, so more than just one error window could be present. Heuristic evaluations were done by small group from university and the questionnaire was sent to 40 end users of the system. The authors conclude that Nielsen's heuristics were more effective for finding problems in error messages. No ERP specific heuristics were used in this study.

3. Research Design

3.1 Overview

To investigate problems affecting usability of ERP system in question we studied fourteen persons working in different companies by conducting a study session. This session contained interview part and a part where the user performed predefined tasks with the system. At the end of a session a participant was asked to fill a usability questionnaire. The sessions were conducted using remote conferencing software Microsoft Teams. Most of them took about 60 minutes.

3.2 Procedure

Participants were asked to share their desktop so they could show how they use the system. Both sound and video were recorded. The participants took remote control of a machine where we had set up a key logger that printed their key presses on screen.

First the participants were asked open ended questions about their experiences with the system. Open ended questions were needed as we wanted to use grounded theory [10] to analyze the results. Choice to use grounded theory was made as we wanted to be as open as possible about problems the users might be facing with the system.

After the open questions all participants were asked to perform four tasks with the system. This was partly motivated by us wanting to use usability questionnaire and this required use of the system before answering. It is worth noting from the start that the setup was not really suitable for measuring task times since, for example the tasks were read to the participant instead of presenting them on paper. We nevertheless report task times and completion rate for these tasks. For these tasks all used the same test version of the system. Lastly they were asked to fill online usability questionnaire. Participants were urged to answer to this questionnaire as quickly as they could after the interview session had ended.

3.3 Participants

Part of the participants were employees in the client ferry companies working for example in sales, management and support. Part were from company developing the ERP system. Most expert users were from client companies, some from ERP provider.

Unlike we had originally hoped, we did not get any novice participants from the client companies. We recruited people from the ERP provider company who had not used the system much before. Four people accepted the invite and joined the study. All were offered a quick training session where a senior quality assurance engineer instructed them how to perform common tasks with the system. These tasks included those we would ask them to perform in the study sessions at a later date. All but one novice participated.

3.4 Interview questions

Interview questions can be found in Appendix A. They can be divided in base questions, questions for novices and expert questions. First the participants were asked background questions such as age, handedness, how long they had used the ERP system and how they perceived their expertise using that system. In this last question they were given five options: beginner, above beginner but below intermediate, intermediate, above intermediate but below advanced, advanced.

Next the participants were asked the questions either from novice or expert question sets. Depending how they classified themselves in last base question (B2-6), those who responded they would classify their skills as above intermediate but below advanced or advanced were asked questions from expert question set, all others received questions from novice question set. The difference between these sets was that expert set had two extra questions asking whether participants had trained new users and whether they could remember how they had learned to use the system.

The questions were made as open as possible since, as stated by Charmaz in [10], interview questions should be open enough to allow unanticipated material to emerge. Also we did dot want to hint "correct" answers. While answering these questions the participants could use test version of the system used in their company or a test version set up by us. This ment they could show what they were talking about and how they use the system. Participants were also asked to give examples of what they ment with their answers so they would be less abstract and maybe possible to observe and compare with other answers.

In both novice and expert question sets first five questions were about using the system. First was a warm-up question asking if participants could show what they did with the system last time they used it. Next we asked if they saw unfamiliar buttons or terms on booking transaction screen or if they remembered some other screen that had unclear things. The aim here was to see if they had problems with this central transaction screen or could point to problems somewhere else.

Next we asked how participants used the system, how it was to use it and how it was to use it the first time. We thought it could be useful to ask how each participant used the system so we could for example paint a picture for different use cases. The question concerning how it was to use the system was to find out about participants general experience of using the system. Question concerning first time using the system was motivated by comments made before the study that starting with the system was hard.

Continuing with the theme of learning, we next asked when the participants last time learned something new and what they did when facing problems. Reasoning behind the first question was that participants could give example of how they learned things with the system. The next was to find about the ways participants solved problems with the system. Experts were also asked if they could remember how did they learn to use the system. This was motivated by the idea that maybe experts had some ways of learning that could for example be taken into account when designing the system so it would be easier to learn.

The last questions were about teaching and helping others with the system. Both experts and all others were asked if they had helped people use the system and how they would teach someone who has never used the system. Reasoning in the first one was to hear about experiences the participants had had when helping people with the system and second to get their opinions how to teach people how to use it. Experts were also asked additional question whether they had taught people unfamiliar with the system to use it. This was again to hear about their experiences in training people to use the system and maybe uncover some problems or good practices.

3.5 Interpreting interview data

According to Olson [27] p29 grounded theory is a family of methods that can be used to construct theories that are grounded in data. It is useful in describing data and building abstractions based on it. As described by [32] coding and comparing are important part of this data analysis. This means labeling for example events or actions and doing comparisons. These labels or codes are combined into more abstract categories and concepts.

We used grounded theory method in a way expressed in [27] p41 where they describe its use in HCI research to analyze a completed data set. This is in contrast with for example [32] where they see as essential the starting of analysing before all data has been collected. This is done in order to let developing theory guide sampling of data in future and answer questions that have risen from the data. This sampling is called theoretical sampling. In our case we do also coding and comparing and develop our codes into more abstract concepts, but as the data is already gathered, we can see theoretical sampling happening when we for example sort our data differently based on understanding developed so far.

After doing coding and analyzing for some time, for usability issues we used existing usability heuristics to link them to existing literature. We grouped these issues using both the five ERP specific heuristics identified by Singh et al. [36] and Nielsen's 10 heuristics [24].

3.6 Practical tasks

We wanted to see how much more efficient expert users were in using the system compared to novices. Initial discussions indicated that novices had had big problems learning to use the system while experts could use it quickly with key bindings. So we selected four tasks and measured times it took participants to complete them.

When selecting tasks we wanted tasks that would be sufficiently challenging but not too hard.Tasks should also be such that they would be common to many people using the system. We ended up with four tasks. These can be found in Appendix B.

Task one and two dealt with creating new new customer and booking, respectively. Participants needed to locate correct transaction screens and fill data to correct fields, then save successfully. Tasks three and four were about editing existing data: Participants needed to find already created customer or booking and change data in them as asked.

3.7 Usability questionnaire

We used System Usability Scale (SUS) from John Brooke [7] to access the usability of the ERP system. In his article Brooke called it "A quick and dirty usability scale" and described it to be a low cost way to access perceived usability of a system. Since then many studies have used SUS and it has proved to be reliable way to measure perceived usability [8].

SUS questionnaire asks participants to rate how much they agree or disagree with ten claims presented. Participants rate this on five point likert scale where one equals "Strongly disagree" and five "Strongly agree". A score can be calculated from these answers by subtracting 1 from answers to odd numbered questions and subtracting answers to even numbered questions from 5. These resulting score contributions are then summed together and result is multiplied by 2,5 resulting in score between 0 and 100. Questionnaire used in this study can be found in Appendix C.

4. Results

4.1 Answers to base questions

Based on the data from base questions we identified three groups of users: novices, intermediates and experts: Novices had used the system less than year and described their expertise as beginner or above beginner but below intermediate. We classified intermediate users to be those who had used the system longer but did not classify themselves as having above intermediate or advanced skills, whereas we classified experts to be those who classified themselves having above intermediate skills. This grouping yielded four novices, three intermediates and seven experts. Background data such as age group distribution, gender and handedness, as well as length of systems use be found in Table 4.1. Data is provided fro each user group.

Participants roles in their organizations ranged from travel center employees to software developers. In novice group participants were employed in software development and management. This is because we recruited participants from withing the organization manufacturing the ERP system in question. Initial plan called for novice participants from client companies but COVID19 prevented hiring of new seasonal workers.

The intermediate group consisted only of travel center employees. The expert group had these too, but also quality assurance engineers and people responsible of maintaining the ERP system in question.

When asked about their computer skills, less surprisingly novice group members responded having very high and expert skills and being very confident with computers. Intermediates on the other hand reported medium computer skills while expert groups skill descriptions were mostly similar as with novices, meaning excellent. Some reported more basic skills.

Table 4.1: Participant data.

Group	Novices	Intermediates	Experts
n	4	3	7
Age			
< 30	2	0	2
$30 - 50$	2	3	5
> 50	0	0	0
Gender			
Male	3	0	2
Female	1	3	4
No answer	0	0	1
Handedness			
Right	3	3	5
Left	1	0	2
Group avg use of system			
years	<1	3,7	6,9
weekly hours	1,75	20,5	30

4.2 Open coding results

Coding the interview data and processing those codes yielded following five categories: *system use cases*, *learning*, *helping others*, *training novices* and *possible usability issues*. We go now throught these categories.

4.2.1 System use cases

With experts the system use could be divided in testing, making transactions, reporting and configuration. Testing included such things as testing that an external system worked. Transactions included making of bookings and paying them. Reporting included such things as printing amount of passengers on a ferry. Configuring the system ment for example configuring business rules.

> "And we do passenger count at the end of the month, so I make sure all reservations are checked in and boarded."

Intermediate users also did transactions and reporting, but did not mention things related to configuration or testing.

> "Then I can see how many passengers that are walking on board today. And I have another list that shows how many cars so I can see how many people I need for checking in cars for checking in walk-in passengers, so it's good for planning."

Use cases for novice group included making transactions and testing. One had used the system only once before the interviews. Two novices had used the system for software development purposes, like testing if a new feature worked.

> "So I, I I use it to look at look everything up and see how everything looks so that I can break it down and make it better basically."

4.2.2 Learning, helping others and training novices

Most experts reported having helped others with the system and trained novices. Experts told they learned things by using the system, by getting help from colleagues or by using documentation. Learning by doing was seen as essential by many.

> "And it's just repetition seems to be the best way to to learn a system like this for sure. Yeah, nothing can compare to actually just doing it really."

All intermediate users had helped others with the system. One had trained new employees. One novice reported having helped others. Both intermediate and novice users advocated learning by doing.

> "I as I see it, have a manual, and and in that manual is really simple. How you log in when you are in the first screen, what do you do next? And of course that's based on that there are some tasks, so without without any tasks going inside [system] is useless, so you had to have a task of creating a new booking. You have a task of running report or you have a task of maintaining the system or whatever task you have."

4.2.3 Possible usability issues

We list and give examples possible usability issues identified from the transcripts. List of heuristics that were used in grouping the issues can be found in table 4.2. Note that not all heuristics from Nielsen were used. Name on the system has been omitted from quotes presented.

Table 4.2: Heuristics used for grouping potential usability issues.

From Singh
Navigation and access to information
Appropriateness of Task Support
Presentation of Screen and Output
Intuitive Nature of System
Ability to customize

From Nielsen
Visibility of system status
Match between system and the real world
User control and freedom
Consistency and standards
Error prevention
Recognition rather than recall
Flexibility and efficiency of use
Aesthetic and minimalist design
Help users recognize, diagnose and recover from errors
Help and documentation

Navigation and access to information

These issues deal with navigation and finding information. In [36] one of the issues in this category is whether functionality can be found easily. We found some comments from our transcripts that point to problems in finding functions: One expert told the user interface had things worded too similarly. This had resulted in her ending up in wrong places. Another expert said an "overview" of things would be good to find things faster.

> "Things are very similarly worded but mean very different things. Particularly when I've been working in the price group, they may have a page that says product price rule and then there will be another page that says something like rule of the product, but they're completely different things, even though they sound similar. So [???] know where you need to go but then you end up on the wrong spot."

One novice found navigation through systems menus confusing. Same user found starting screen confusing, what was supposed to be done there? Another novice had problems finding an essential transaction screen.

> "..it, uh can be quite confusing to navigate throughout the menus and everything."

Appropriateness of Task Support

One type of issue in this category is whether the system automates tasks that are routine. Following comments indicate lack of task automation in at least some parts of the system: One expert told there being lots of manual typing in a specific part of the system. One intermediate reported it being slow to add price to a customer. Namely she has to input price 25 times for one single customer:

> "So actually to put in a price for a normal truck loaded I have to put in the price five times. But that is only loaded. Then I have to put in the price five times for empty, and that is not all because I do also have to do it for trailer loaded and empty. And then I have to put in the price five times for a lorry up till 13 meters. So to put in a price for a customer I have to put in the price 25 times in [system]. And I'm sorry to say that is not a good solution."

Second type of issue in this category is whether the system is easy to use. To this we found multiple comments from all user groups pointing that the system is hard to use. Some experts told about systems complexity and one expressed it was really complicated at times. Some told certain parts were difficult or hard. One intermediate described problems when she had to do a more complex use case. One Novice expressed the system was hard to use. Here is comment from an expert user:

> "Because as you know. It's pretty difficult system, really complicated sometimes..."

It is worth noting that some experts described the system quite differently: One found it easy to use and other found she could do anything she wanted with it.

Visual layout and Output

Next we consider problems with the design of the visual layout of the user interface. One expert complained the user interface lacked modern things, while same user also said the user interface was logically structured. One novice said there was too much information in single place, and places where there were three levels of "hierarchical information". This ment that selecting a data row from a list showed its related attributes in another part of the same screen, and then again selecting one of these showed its related attributes, these also on different part of same screen.

> "It is little bit too much information squeezed to one single view. ... ... Yeah, and and then also and this this is that was probably like let's say two level hierarchy information. Um, but you have also places where you have three level."

Next type of issues is whether the system supports informed decision making. Here one expert told part timers and even herself had had problems answering customers because the system is complicated:

> "And I can see that some of the.. As well as myself the.. Some of the part time employees that we have for instance working at check in if they are to check the customer profile and then they experience the same issues. They don't. They feel like they can't give an answer directly to the customer because its difficult to to know what everything means on the setup for the customer."

One intermediate complained of similar thing, with there being so many product codes it made it hard to answer to customers:

> "So we have some tools in place for finding the codes. It's just I think it's bit too complicated that it has to be this way. If it was just less codes or always the cheapest one who came up then it would be easier to help the guests. It's.. I don't think it's good customer service that I have to say. Oh, I will call you back in an hour 'cause I need to look through all my my tools to find it."

Other intermediate wondered why the system didn't give her the best offer for a product. One intermediate told it was hard to know when there was a new offer and explained she would find them from company intranet.

More than half of experts gave examples of parts of the system that were hard to understand or unclear. One commented certain part of often used transaction screen telling she wasn't sure what it was used for:

> "...it's not something that we use, and I'm uncertain how someone how one of the companies would use this, since it seems to be repeated information from up here."

Here another expert describes configuring one functionality which sends information to customers:

> "I found that a little bit challenging as well. Like you can see there's a lot of timing options here. Should I set it in daily basis if it to be run by day, but actually it was supposed to be disabled, so I didn't quite understand the logic in that."

Here it is also worth noting that some experts also described the system as understandable or self explanatory.

Most novices described the system with words such as confusing and strange, some parts making no sense or being terrible. When asked how it was to use the system, one novice gave a short answer: "Confusing." When asked to elaborate, he said:

> "Like like a UI, it should be intuitive and user friendly enough so that any kind of user can just pick it up and start using it from the get go. ... However, if you just look at this, then you have basically no idea as complete stranger to this that you can actually make a booking and everything in-between that. Like.. Like the thing that I just did with changing the menu so that it came to the main menu, I didn't even know that that thing existed there. ... So, so it's very confusing, and it's very strange."

Here is another example of the system being non intuitive. A more experienced user from novice group described one part of the system like this:

> "Uh, there's a lot of information, and this is probably not the the hardest part. But once you're here. This part is really, really hard to know what to look for. What? What is what? What kind of information is it here? So so the first time I saw this one I was terrified."

Intuitive Nature of System

Following issues concern learnability, such as need for long introductions or ability to become skillful in short time.

Most novices told one would need a training to use this system or there being a knowledge treshold. One expert told her initial training of two days had been too short. One told there had been too much things to learn at one time. One expert said it takes long time to "get exposed" different things:

> "I feel bad for the employee that has to work by themselves because there's so many things to learn and remember, and it takes a long time to be exposed to every type of customer, every type of ticket that we sell."

One expert told he got confident in couple of weeks, and that initial training had taken couple days. For some experts so much time had passed that they did not remember much from the days they started using the system.

In this same group we have also mentions of system being intimidating to learn. Many experts told the first time they used the system was intimidating or difficult. Common reason given was amount of things such as fields in the user interface or amount of things in general.

> "It was intimidating. It was intimidating because there are so many different fields and I remember one of the first things the when they were training us was: *"Don't worry. You don't need to fill in every field."* And I still say that to new employees now."

Intermediate users told similar things. One said she had gotten used to the complexity of the system even though it was hard when she started.

Novices described their first experiences similarly: being confused, feeling terrified. One said he feared he would break the system:

> "I didn't even know that that that you could open up more windows or that you could make bookings or whatever and when I started jumping around

> with the menu that's when this stuff started to appear and it was almost a little bit scary to use only because of the fact that I might actually end up breaking the system, because I was such a newbie, at it."

Ability to customize

In this group the qualifying question was whether the system could support user-level customization. One expert told she had instructed new users to ignore part of transaction screen as not needed:

> "I honestly tell them to ignore the bottom half of this entire screen, because for the most part we don't need it, so I almost take a piece of paper and cover the whole bottom of the screen so they are not overwhelmed."

One intermediate said there were parts of the user interface that didn't show any information when she was using it. These might indicate that the system lacks customization where user can hide unneeded things from transaction screens. One expert also said she had problems learning changed shortcuts. This also raises the question whether there is ability to customize them.

Consistency and standards

One novice talked of the system being inconsistent. For example here a help button didn't act as expected from previous experience:

> "...normally in a windows you're going to have a, uh, a question mark here. And if I first question mark and hover over something that would give me some kind of help text."

Recognition rather than recall

These possible usability problems deal with users needing to remember things instead of the system assisting.

One set of problems seems to be lack of visibility: Both experts and novices wanted to see what fields were mandatory. In one case one would see that a field was needed only after trying to save and getting error window.

There was also a need to remember lot of stuff to become efficient. Some experts told there was problem with too many category and product codes. This had been problematic to new users, but one expert noted also experienced users could forget if not working on something daily. Here is one talking about remembering and finding correct codes:

> "..you can generally find stuff anyway. But there is a lot of it, and let's say it's a lot easier once you you know these without having to look, which does come eventually.."

And here is another:

> "So if I'm a new employee and I'm not entirely sure what needs to go in this field here, I can hit F9 to search, but the problem is we use so many codes that it's overwhelming. It's supposed to be helpful because it's supposed to help the employee remember. Oh yeah, that's right. I put this code here, but when you have so many different codes, it's almost overwhelming and it's too much information."

Also one intermediate noted that making bookings required lots of familiarity with codes:

> "It's a bit complicated. So you need to. Uh, you need to be very familiar with all the codes to be able to book something."

Flexibility and efficiency of use

It looks like the system does enable experts to be fast with shortcuts while novices can use both mouse and keyboard. Multiple experts reported using shortcuts, some told they used them a lot. One told she liked that there was possibility to use both mouse and keyboard. Thus the system design seems to cater both experts and novices in this respect. However, one expert told she found the user interface "clunky" despite the shortcuts:

> "It's not fast to move through. I've become fast in using it, but it's a bit clunky to move through different pages and access different areas."

Help users recognize, diagnose and recover from errors

According to Nielsen's heuristics [23] error messages should tell what is wrong and suggest a solution. This doesn't seem to be the case at least with all of them since for example one expert found error reporting made no sense to most people. Another expert found error messages vague:

> "... but the error messages are very vague. They don't help you. Sometimes they help you with exactly what's wrong, but other times they're just very vague popups that say product price journey not found. It has 10 different possible places where you might need to go look to find what's wrong."

Another expert said she used service desk to get help with error messages. This might point that error reporting is not sufficient if some experts need help with the messages.

Help and documentation

We got couple comments about the systems documentation. There might be need to make searching information easier. One expert told it took long time to find things in knowledge base. One told finding help was difficult:

> "I also use sometimes the [company name] user knowledge base, but sometimes I find it a little bit difficult to find the answers to the specific questions that I have in the database. I use a lot of time trying to see where can I find the answer to this question that I have."

One expert told he didn't use the knowledge base for problem solving.

4.3 Task time and completion rate

4.3.1 Task times

Task times for practical tasks were gathered from interview recordings and can be seen in table 4.3. Tasks started when interviewer started to read the task and ended when data were successfully saved. Later could be verified from video by seeing a pop up informing of successful save. There was however complications to measuring these times as there was talk additional to giving instructions. For example, sometimes a participant would point to some problem with the system and interviewer then ask questions about it. These parts needed to be subtracted from task times. One should therefore not compare these on the level of individual seconds. Also it could be that some participants thought they were expected to slowly expalin how they would do something. In one case a participant was urged to do task the way they would normally do it.

In case there was a technical problem we mark that with letter T. Technical problems occurred when interviewer made a mistake or participant had to quit the session early. In some cases interviewer for example didn't read part of script and so bookings made like this were missing some parts. We counted these tasks as technical failures.

Task times vary a lot even withing groups. It looked like some users had a routine to do the tasks and were fast with using basically just keyboard. For example users E3 and E4 and E6 were working in customer service positions while E5 and E7 were quality assurance engineers. One would assume those working with customer transactions daily had developed greater level of routine than those with broader scope of use. It is worth noting that for example expert user E3 completion of task 4 in just 37 seconds doesn't leave much time to try different approaches. Even the reading of the script takes time.

Some users like E2 explained things while they were doing the tasks so this definitely lengthened their completion times.

4.3.2 Task completion rate

In case participant could not complete task successfully we mark that with word "FAIL". Task was counted as failure if participant asked and got help when it was not planned to be given, for example if they told they didn't know where data would go. Task was also counted as failure if participant declined to complete task or said he could not do it.

Most novices had serious problems completing the tasks. Only one managed to complete all of them. Most had troubles and ended up asking advice or said they did not know how to do it. This indicates that there indeed is problems with the systems usability.

In other groups task failures were far less common. There was one intermediate I1 who said she could not do a task. In this case she told she did not create customers in her work. There was also one case of an expert user E5 giving up on task. Maybe also this user hadn't used the system this way previously. These occurrences speak against ease of use and intuitive nature of the system: people who have used system a lot say they cannot do a task which was considered common in initial interview when planning the study and selecting these tasks.

4.4 Usability questionnaire

Results from system usability survey using system usability scale (SUS) can be found in table 4.4. 12 out of 14 participants responded to survey. We added *adjective rating* to each score using recommended ranges from SUS scoring template from [21]. In this template scores in range of 0-64 are *not acceptable*, scores in range of 65-84 are *acceptable* and scores in range of 85-100 are *excellent*. As Brooke points in his retrospective, SUS scores *are not* percentages. According to the author a more meaningful way for interpreting them is seeing them as percentiles. Brooke points that average score is 68. This should explain why for example range of *acceptable* scores starts as high as 65 in scoring template.

Table 4.3: Task times in seconds. T = technical error, FAIL = task failure

	task 1	task 2	task 3	task 4
Novices				
$N1$	146	310	47	103
$N2$	T	FAIL	109	FAIL
$N3$	FAIL	T	T	T
$N4$	T	FAIL	FAIL	219
Intermediates				
$I1$	FAIL	162	258	83
$I2$	165	188	40	103
$I3$	108	T	57	52
Experts				
$E1$	114	158	101	T
$E2$	183	146	96	91
$E3$	112	225	52	37
$E4$	88	191	45	75
$E5$	222	223	FAIL	232
$E6$	95	T	69	T
$E7$	209	240	99	97

Table 4.4: System usability scale results.

	SUS score	Adjective rating
Novices		
*N*1	35	not acceptable
*N*2	10	not acceptable
*N*3	55	not acceptable
*N*4	37,5	not acceptable
Intermediates		
*I*2	50	not acceptable
*I*3	52,5	not acceptable
Experts		
*E*2	27,5	not acceptable
*E*3	95	excellent
*E*4	57,5	not acceptable
*E*5	55	not acceptable
*E*6	92,5	excellent
*E*7	60	not acceptable

To summarise the results, most participants viewed usability of the system below average result and as *not acceptable*. Exception to this were two experts who viewed the usability to be *excellent*. Lowest group average score was in novice group.

5. Discussion

5.1 Usability problems

Based on the study we can now say that the system seems to suffer from usability issues. We will now discuss these. As we had assumed, novices reported multiple problems, but unlike initially assumed, also many experts reported problems related to usability. Categories of issues and examples of each can be found in 5.1.

5.1.1 Perceived usefulness

Perceived usefulness measured in SUS scale indicates most experienced users and all novices did not find the system that useful. During interviews, participants from all user groups reported problems that indicate the system is not easy to use. There were, however, some who told it was easy. Novices used terms such as confusing, strange and terrible to describe the user interface. Half of the experts said the system was hard to understand while some described it being understandable or self explanatory.

However, there seems to be a group of long time users that perceive the system highly usable. One could assume therefore that the system might work well in certain context. The fact that both of these participants worked in the same client company hints to this direction. Also these participants did not complain about the number of, for example, product codes. Therefore they may not experience similar complexity than some other experts, who had mentioned the number of these codes being a problem.

5.1.2 Navigation

Novices faced navigation problems in both finding functionality and in performing tasks. During interviews both novices and experts reported problems with navigation, all in all four participants.

In addition to interview reports, most novice and some experienced users problems with completing tasks. We can be therefore somewhat confident that navigation problems do exist.

Table 5.1: Possible usability issues with category of issue and examples

category	**examples**
Navigation and access to information	*functionality can not be found easily*
	navigation is confusing
	desire for overview
Appropriateness of Task Support	*a part that needs lots of manual typing*
	slow to add price to a customer
	system is complicated to use
Visual layout and Output	*lack of modern things*
	too much information in one place
	problems answering to customers: *complicated system* *many product codes*
	problems finding offers
	parts that are hard to understand
Intuitive Nature of System	*need of training*
	many things to learn
	system is intimidating to learn
Ability to customize	*novices instructed to ignore part of UI*
	part of UI shows no information
	problems learning changed shortcuts
Consistency and standards	*non-standard help button*
Recognition rather than recall	*hard to see what fields mandatory*
	need to remember to be efficient
	too many category/product codes
	requires familiarity with codes
Flexibility and efficiency of use	*finding UI clunky despite shortcuts*
Help users recognize, diagnose and recover from errors	*error reporting makes no sense*
	vague error messages
Help and documentation	*slow to find things in knowledge base*
	difficult to find help from knowledge base
	not using knowledge base for problem solving

Also, looking at the recordings where novices perform tasks shows that also people who did not report navigation issues had problems finding functionality. So it seems all people who had navigation problems did not report them in interview. It was not in the scope of this study, but one could go through the recordings and document places where, for example, participant navigates aimlessly, clicking multiple menus to find correct one.

Task completion rates show that novices had problems with the basic tasks most of them had received training to perform. Also, one expert and one intermediate failed a task. This might indicate some experienced users do not feel that comfortable with the system if they are asked to perform something they have not done before. In case of the intermediate she had used the system in freight context, not in that of ferry passengers. The expert had used the system 30 hours a week, but less than one year, and as a quality assurance engineer had probably used many parts of the system but might have missed the task we asked her to perform.

5.1.3 Lack of support

One issue with the system seems to be that it requires users remember many things in order to be efficient. There is also a lack of visibility: what fields are mandatory is not expressed clearly enough. A related finding was that some long time users told they could not give timely customer service because it took so long to find needed information.

Novices did not report problems with *task support*. One expert told that the system required lots of manual typing. One intermediate reported it was slow to do a specific configuration, namely add price.

These might indicate that the system would need smarter solutions to make configuration and data input easier. Also more powerful search capabilities could help with finding information such as products and functions.

5.1.4 Customization and flexibility

It seems the system suffers from poor *customizability*, as there were mentions that new users were instructed to ignore parts of transaction screen. Also error reporting got negative comments: some experts commented the error messages not making sense to most people or them being vague. Help and documentation received also complaints from three experts: finding help was seen as difficult.

The current system seems to cater both long time users and novices in that novices can use both keyboard and mouse while experts can utilize shortcuts. However, one expert commented that the system was "clunky" to use despite shortcuts.

5.1.5 Learning

Many long time users and novices reported the system *intimidating* to learn. This points to problems in learnability. Other thing that was found out from the transcripts was that many advocated learning by doing as means to learn the system. This should be kept in mind when developing new user interface.

Our interview script was quite long and we also included questions about learning and teaching. It is not in the scope of this study, however, to analyze the learning process. Most long time users had trained novices. One can argue that this raises their credibility when they talked about problems novices face.

5.2 Recommendations

As novices and some experts faced navigation issues, it would probably improve usability if future implementations of the user interface supported users in navigation, like finding correct functionality and task steps to perform. This could also help novices in learning. Many participants told learning by doing was the best way to learn system like this, so maybe this could be made easier by improving navigation to let users discover functionality. Other important area would be reducing users memory load and overall user interface complexity. Users should not need to remember many product or category codes to be efficient. They should not need to remember correct places of functions and correct steps to finish a task. Both experts and novices would benefit from implementations reducing memory load or complexity.

Solutions to navigation and progress guidance have been studied by Babaian et al. [4][5] by implementing prototypes of systems that visualize previously completed tasks with *playback mechanism* and use *interactive graphs* to visualize a process consisting on multiple tasks. Functionality like this could help new and also long time users perform tasks by lowering the amount of things they need to remember: they could be show how to perform something and then later they could "play" it again and see how it went. This approach has been studied in laboratory setting and shown that it helped users learn how to perform tasks.

However, this approach was not yet tried on existing full scale ERP system. There could also be challenges when implementing data model needed for functionality like this. For example, if the way a task is done changes, should old "task recordings" be updated somehow?

Adaptive user interface could be helpful in reducing users memory load and finding functionality. Prototype by Eichler et al. [12] had side bar that would show often used items. It also used a visual technique (ephemeral visualization) to show most often used fields. These approaches have the benefit that the system adapts to a specific user without becoming so unique between different users that they cannot ask advice from another. Also, if implemented on existing system, they would probably not disturb an expert user who knows how to use the system already since no changes to layout are done. Ephemeral visualization does, however, require animating capabilities. Also, these techniques could provide help in many situations in multiple screens.

However, we are not aware of laboratory studies conducted to access impact of these methods to usability. Also, this method does not solve the problem of finding functions or objects the first time. One approach might be to design user interface according to Lambeck and Groh's [17] concept of *scalable user interface*. Their model could be useful in answering to complexity and memory load problems as it emphasized the ability to find data and functions, and also compare content. This could enable the users to give faster customer service and help in making informed decisions, as the data would probably be easier to master. Importance of powerful search functionality was also discussed by Matthews in his white paper [20]. His arguments for web-like navigation and search functionalities seem justified. As we do not expect web users to know how everything works, why should this be expected from a user of complex enterprise application?

One could also use other principles and heuristics discussed in this study as aids when developing user interfaces, such as the four design principles [4] to improve human computer collaboration or for example the classic Nielsen's heuristics [24], which seem to be relevant in this context.

5.3 Effect of COVID19

We had initially planned to conduct at least part of the interviews by meeting participants in person and letting them control a machine that we would bring with us. COVID19 changed the situation and we had to think again the interview setup. In the end only one interview session was conducted by meeting participant in her company premises.

All in all the remote setup worked well. The participants controlled our machine remotely and this way we could log their key presses, a thing that would not have been possible with their company machines since installing software like that on those was out of question. There were some problems with participants using remote control software. Some had not used it before and had to get technical help to get started. Some participants did not have remote control software we requested installed and were not allowed to install it. In those cases they used other similar software available. Only in one case we decided to conduct the interview by meeting the participant. This was seen easier than to troubleshoot remote control issues. In this case the participant was given our machine to use.

6. Limitations

We discuss now the limitations of this study. With the interview data one limitation is that we had only one person coding the transcripts. More coders cold have yielded more reliable results. With interviews there is also the problem that we could not recruit any novices from client companies. We do not know how much the novices we had as participants would differ from those arriving to ferry companies to do seasonal summer work, for example. We had one novice who had not used the system before introduction session and would not use it in her work. Other novices had used the system in their work.

System Usability Scale is widely used in usability testing and a reliable way to measure perceived usefulness. Also, Brooke points in [8] to study showing that SUS can give good results with small number of participants. With our remote setup, however, we did not control when users returned their answers and while we asked participants to fill the questionnaire right after interview, some told they would do it later since they had no time.

Shortcomings of our measurement of task time and completion have already been discussed in chapter 3. Our setup would have been more reliable if, instead of reading the tasks, we would have had a text window or something similar giving the instructions. The role of the interviewer should have been only to be there to help if there was a problem with the setup. We could have also given participants more instructions. While task completion rates probably are more reliable than task times, their reliability could maybe have been improved by giving users instructions before the tasks what to do if they get stuck, for example.

7. Conclusions

The aim of this study was to find usability problems of an ERP system used by ferry operators. We studied new and long time users by conducting sessions where the users told about their experiences, performed tasks with the system and filled usability questionnaire.

Many novice and long time users reported problems. The scores from usability questionnaires made with System Usability Scale show all but two participants perceived the usability of the system as below average and in adjective rating "not acceptable". Two users rated the usability as "excellent". We reasoned that there could be a group of users who use the system in such a way and in such context that they do not experience these problems.

The results indicate novices have trouble, for example, navigating and completing tasks. Also some long time users reported navigation issues. The system seems to require that it's users remember lots of things in order to use it well. The interviews and tasks indicate the system is complex and hard to use and both novices and experts face problems. This is supported by perceived usability scores. While experts could in most cases finish all tasks, during interview some of them reported problems such as finding products the customers needed, error reporting being unclear, configuration being tedious, and need for lots of manual typing, for example.

We gave recommendations on what to consider when implementing new user interface for this ERP system. For example, navigation should be improved and users should be provided with powerful search tools. We discussed existing prototypes such as adaptive user interfaces and approaches such as improving human computer collaboration. ERP usability is not studied much. Our study supports use of already developed heuristics in classifying usability problems. Our recommendations how to improve usability of the ERP system studied should give some guidelines on what could be done, although not much is backed by laboratory studies. More work is needed in this field to find and test solutions to usability problems users face.

Bibliography

[1] T. Babaian, W. Lucas, and M.-K. Oja. Evaluating the collaborative critique method. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2137–2164, 2012.

[2] T. Babaian, W. Lucas, J. Xu, and H. Topi. Usability through system-user collaboration. In *International Conference on Design Science Research in Information Systems*, pages 394–409. Springer, 2010.

[3] T. Babaian and W. T. Lucas. Leveraging usage history to support enterprise system users. 2012.

[4] T. Babaian, J. Xu, and W. Lucas. Applying design principles for enhancing enterprise system usability. In *2014 9th International Conference on Software Engineering and Applications (ICSOFT-EA)*, pages 162–169. IEEE, 2014.

[5] T. Babaian, J. Xu, and W. Lucas. Erp prototype with built-in task and process support. *European Journal of Information Systems*, 27(2):189–206, 2018.

[6] N. Bevan, J. Carter, J. Earthy, T. Geis, and S. Harker. New iso standards for usability, usability reports and usability measures. In *International conference on human-computer interaction*, pages 268–278. Springer, 2016.

[7] J. Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.

[8] J. Brooke. Sus: a retrospective. *Journal of Usability Studies*, 8:29–40, 01 2013.

[9] F. Calisir and F. Calisir. The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (erp) systems. *Computers in Human Behavior*, 20(4):505–515, 2004.

[10] K. Charmaz. *Constructing grounded theory: A practical guide through qualitative analysis*. sage, 2006.

[11] J. Choma, D. Quintale, L. A. Zaina, and D. Beraldo. A perspective-based usability inspection for erp systems. In *ICEIS (3)*, pages 57–64, 2015.

[12] Z. Eichler and M. Dostál. Adaptive user interface personalization in erp systems. In *International Conference on Business Information Systems*, pages 49–60. Springer, 2012.

[13] K. Hornbæk. Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies*, 64(2):79–102, 2006.

[14] B. Ives, M. H. Olson, and J. J. Baroudi. The measurement of user information satisfaction. *Communications of the ACM*, 26(10):785–793, 1983.

[15] D. Kieras et al. Using the keystroke-level model to estimate execution times. *University of Michigan*, 555, 2001.

[16] C. Lambeck, C. Fohrholz, C. Leyh, I. Supulniece, and R. MÃŒller. *Commonalities and contrasts: An investigation of ERP usability in a comparative user study.* aisel.aisnet.org, 2014.

[17] C. Lambeck and R. Groh. Mastering erp interface complexity-a scalable user interface concept for erp systems. In *International Conference on Enterprise Information Systems*, volume 2, pages 170–178. SCITEPRESS, 2013.

[18] C. Lambeck, R. Müller, C. Fohrholz, and C. Leyh. (re-) evaluating user interface aspects in erp systems–an empirical user study. In *2014 47th Hawaii International Conference on System Sciences*, pages 396–405. IEEE, 2014.

[19] W. Lucas and T. Babaian. The collaborative critique: An inspection method for expert evaluation of user interfaces. *International Journal of Human-Computer Interaction*, 31(11):843–859, 2015.

[20] D. Matthews. Usability as an erp selection criteria. *IFS White Paper, January*, 2008.

[21] S. McLellan, A. Muddimer, and S. C. Peres. The effect of experience on system usability scale ratings. *Journal of usability studies*, 7(2):56–67, 2012.

[22] J. Nielsen. *Usability engineering.* Morgan Kaufmann, 1994.

[23] J. Nielsen. Ten usability heuristics, 2005.

[24] J. Nielsen and R. L. Mack, editors. *Usability Inspection Methods.* John Wiley Sons, Inc., USA, 1994.

[25] M. Oja and W. Lucas. Evaluating the usability of erp systems: What can critical incidents tell us. *Proceedings of Pre-ICIS Workshop on ES âŠ*, 2010.

[26] M. Oja and W. Lucas. Erp usability issues from the user and expert perspectives. *Journal of Information Technology Case and âŠ*, 2011.

[27] J. S. Olson and W. A. Kellogg. *Ways of Knowing in HCI*, volume 2. Springer, 2014.

[28] N. Parks. Testing & quantifying erp usability. *Proceedings of the 1st Annual conference on Research âŠ*, 2012.

[29] F. Paz and J. A. Pow-Sang. Current trends in usability evaluation methods: A systematic review. In *2014 7th International Conference on Advanced Software Engineering and Its Applications*, pages 11–15, 2014.

[30] D. Quiñones and C. Rusu. How to develop usability heuristics: A systematic literature review. *Computer standards & interfaces*, 53:89–122, 2017.

[31] M. Sadiq and A. Pirhonen. Usability of erp error messages. *International Journal of Computer and Information âŠ*, 2014.

[32] A. Sbaraini, S. M. Carter, R. W. Evans, and A. Blinkhorn. How to do a grounded theory study: a worked example of a study of dental practices. *BMC medical research methodology*, 11(1):1–10, 2011.

[33] B. Scholtz, A. Calitz, and C. Cilliers. Usability evaluation of a medium-sized erp system in higher education. *Electronic Journal of Information Systems Evaluation*, 16(2):148, 2013.

[34] B. Scholtz, C. Cilliers, and A. Calitz. Qualitative techniques for evaluating enterprise resource planning (erp) user interfaces. In *Proceedings of the 2010 annual research conference of the South African Institute of Computer Scientists and Information Technologists*, pages 284–293, 2010.

[35] B. Scholtz, I. Mahmud, and T. Ramayah. Does usability matter? an analysis of the impact of usability on technology acceptance in erp settings. *Interdisciplinary Journal of âŠ*, 2016.

[36] A. Singh and J. Wesson. Evaluation criteria for assessing the usability of erp systems. In *Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, SAICSIT '09, page 87â95, New York, NY, USA, 2009. Association for Computing Machinery.

[37] A. Singh and J. Wesson. Improving the usability of erp systems through the application of adaptive user interfaces. *ICEIS (4)*, 2009.

[38] H. Topi, W. T. Lucas, and T. Babaian. Identifying usability issues with an erp implementation. In *ICEIS*, pages 128–133, 2005.

[39] V. Veneziano, I. Mahmud, A. Khatun, and ... Usability analysis of erp software: Education and experience of users' as moderators. *The 8th International âŠ*, 2014.

[40] J. J. Xu and H. Topi. A conceptual model for user-system collaboration: Enhancing usability of complex information systems. *Communications of the Association for Information Systems*, 41(1):31, 2017.

[41] E. Yassien, R. Masa'deh, M. A. Mufleh, A. Alrowwad, and R. Masa'deh. The impact of erp system's usability on enterprise resource planning project implementation success via the mediating role of user satisfaction. *Journal of Management and Research*, 9:49–71, 2017.

Appendix A. Interview questions

(Name of the ERP system omitted)

BASE QUESTIONS FOR ALL

PERSONAL:

B1-1: In which age group do you belong:

a) 30 and under

b) over 30 and under 50

c) over 50

B1-2: Are you left or right handed or ambidextrous?

B1-3: Which of the following describes you:

a) Female

b) Male

c) Prefer not to answer

B1-4: Could you describe your role in the organization where you are currently employed?

COMPUTING:

B2-1: How would you describe your computer skills?

B2-2: What kind of programs do you use and how much? In work or free time?

B2-3: How long (years) have you used [system] UI?

B2-4: How much you use [system] UI weekly? On average, in hours.

B2-5: How much have you used similar reservation programs?

B2-6: (Expertice with [system] UI)
Think of an advanced, intermediate and beginner [system] UI user. These might be people you know but they need not be. Would you classify your expertice with [system] UI as:

- beginner level of expertice
- above beginner but below intermediate level of expertice
- intermediate level expertice
- above intermediate but below advanced level of expertice
- advanced level of expertice

QUESTIONS FOR EXPERT USERS:

DAILY USE OF [system] UI:

E1-1: When was the last time you used [system] UI? Could you tell us about that? Maybe show a bit?

E1-2: If you look at the UI now, is there some buttons or terms that you do not recognize or know the use for? These can also be on another forms. (First look at Booking form)

E1-3: How is it to use [system] UI? Could you give example?

E1-4: How do you use [system] UI in your work?
Could you tell what features of [system] UI do you use in your daily work? Can you think of some common scenarios? Could you show us a bit?

E1-5: Could you describe how it was to use [system] UI for the first time?

LEARNING NEW THINGS

E2-1: How did you learn to use the UI? Was something especially hard or easy to learn? Can you remember?

E2-2: Do you remember when you last time learned to do a new thing in [system] UI? Could you tell about that and perhaps show us how you did it?

E2-3: What do you do when you have problems with something or don't know how something works? Can you remember a situation like this? Can you show what you did?

TEACHING OTHERS

E3-1: Have you helped someone to use [system] UI? Could you tell us about that?

E3-2: Have you teached people unfamiliar with [system] UI to use it? Could you tell us about that?

E3-3: How would you teach somebody who has never used [system] UI to use it? Can you give an example?

QUESTIONS FOR NOVICE USERS:

DAILY USE OF [system] UI:

N1-1: When was the last time you used [system] inhouse UI? Could you tell us about that? Maybe show a bit?

N1-2: If you look at the UI now, is there some buttons or terms that you do not recognize or know the use for? These can also be on another forms. (First look at Booking form)

N1-3: How is it to use [system] UI? Could you give example?

N1-4: How do you use [system] UI in your work? Could you tell what features of [system] UI do you use in your daily work? Can you think of some common scenarios? Could you show us a bit?

N1-5: Could you describe how it was to use [system] UI for the first time?

LEARNING NEW THINGS

N2-1: Do you remember when you last time learned to do a new thing in [system] UI? Could you tell about that and perhaps show us how you did it?

N2-2: What do you do when you have problems with something or don't know how something works? Can you remember a situation like this? Can you show what you did?

TEACHING OTHERS

N3-1: Have you helped someone to use [system] UI? Could you tell us about that?

N3-2: How would you teach somebody who has never used [system] UI to use it? Can you give an example?

Appendix B. Practical tasks

TASK 1: Creating a new customer

Could you now show us how to create a new customer? I will give you data when needed.

- specify name, phone number, country, language, currency

TASK 2: Creating a new booking

Could you then show us how to create a new booking?

- specify departure and arrival ports
- specify date
- ask for two way trip, return preferably next day afternoon
- book one adult and one car

TASK 3: Editing a customer

Could you show us how to edit the customer you just created? First close booking and customer info if open. Please do not use customer id for finding customer.

- interviewer gives information for finding customer when needed
- update phone number

T4: Editing a booking

Could you show us how to edit a booking? Again close all open booking and customer forms. Please do not use booking id for finding booking.

- interviewer gives information for finding when needed
- change return one day forward

Appendix C. System Usability Scale questions

Usability Questionnaire

(Name of the ERP system omitted)

For each of the following statements, please mark one circle that best describes your reactions to [system] UI today.

```
1. I think that I would like to use [system] UI frequently.
                      1  2  3  4  5
   Strongly disagree  O  O  O  O  O  Strongly agree

2. I found [system] UI unnecessarily complex.
                      1  2  3  4  5
   Strongly disagree  O  O  O  O  O  Strongly agree

3. I thought [system] UI was easy to use.
                      1  2  3  4  5
   Strongly disagree  O  O  O  O  O  Strongly agree

4. I think that I would need the support of a technical person to be
   able to use [system] UI.
                      1  2  3  4  5
   Strongly disagree  O  O  O  O  O  Strongly agree

5. I found the various functions in [system] UI were well integrated.
                      1  2  3  4  5
   Strongly disagree  O  O  O  O  O  Strongly agree

6. I thought there was too much inconsistency in [system] UI.
                      1  2  3  4  5
   Strongly disagree  O  O  O  O  O  Strongly agree

7. I would imagine that most people would learn to use [system] UI very quickly.
```

```
                    1  2  3  4  5
   Strongly disagree  0  0  0  0  0  Strongly agree

8. I found [system] UI very cumbersome (awkward) to use.
                    1  2  3  4  5
   Strongly disagree  0  0  0  0  0  Strongly agree

9. I felt very confident using [system] UI.
                    1  2  3  4  5
   Strongly disagree  0  0  0  0  0  Strongly agree

10. I needed to learn a lot of things before I could get going with [system] UI.
                    1  2  3  4  5
   Strongly disagree  0  0  0  0  0  Strongly agree
```