ERP USABILITY ISSUES FROM THE USER AND EXPERT PERSPECTIVES[1]

Mari-Klara Oja
Bentley University, MA, United States
oja_mari@bentley.edu

Wendy Lucas
Bentley University, MA, United States
wlucas@bentley.edu

ABSTRACT

Enterprise Resource Planning (ERP) systems are used for streamlining the flow of business processes and information throughout the organization. While industry reliance on these systems has been on the rise, the notorious challenges they impose on their users can severely hinder their successful adoption and use. The scope and complexity of their functionality can be overwhelming, but ERP systems typically provide little in the way of guidance or support. Understanding the specific usability problems experienced by users is essential to the development of more usable systems, yet research in this area has been limited. In this study, we investigate how negative "critical incidents" (i.e., serious breakdowns in human-computer interactions) encountered by users can improve our knowledge and understanding of ERP usability problems. A laboratory-based empirical usability evaluation of a popular ERP system was conducted using both user-reported and expert-observed critical incidents. Having users report on usability issues as they happen provides richer details than are typically available from surveys or interviews alone. Augmenting those accounts with expert observations of user-system interactions yields a deeper understanding of the types of usability issues that must be addressed.

KEYWORDS: Enterprise Resource Planning, ERP, Usability, Critical Incidents, Empirical Evaluation, Contemporaneous Reporting, Retrospective Reporting

INTRODUCTION

Enterprise Resource Planning (ERP) systems automate and integrate processes from throughout the organization. With an estimated worldwide market size of $114 billion in 2011 (Gartner, 2011), companies are increasingly relying on enterprise application software for leveraging information flow and supporting innovative business processes (Woods, 2011). The benefits from having access to data from across the enterprise, however, may not be fully realized because of

[1] *A shorter version of this paper, "Evaluating the Usability of ERP Systems: What Can Critical Incidents Tell Us?," appeared in the proceedings of the Fifth Pre-ICIS workshop on ES Research, St.Louis 2010.*

the usability challenges these systems continue to present to their users, as noted in several industry research reports (see, for example, Hestermann, 2009; Otter, 2008; Hamerman, 2008). Industry studies have also observed the negative effect of poorly designed user interfaces on business performance (Herbert, 2006) and, in particular, on end-user productivity (Iansiti, 2007).

Despite widespread acknowledgement of the poor usability of ERP systems, few research efforts have been directed at indentifying specific usability issues. Hurtienne, Prümper and Rötting (2009) point out the lack of attention paid to ERP software usability even though usable systems can reduce costs for training and documentation, lead to dramatic drops in calls to helpdesks, and prevent accidents and injuries. Topi, Lucas, and Babaian (2005) identified six categories of ERP usability problems that increased learning time, error rates, and user frustration for employees in a Fortunate 500 company. The most common criticism from ERP end-users in case studies of 13 Taiwanese companies was "unfriendly," with interviewees categorizing the systems as complicated and difficult to use (Yeh, 2006). Expert reviewers of an ERP system found its terminology to be confusing at times, the Help functionality to be difficult to use, and guidance to be lacking, particularly for the novice user (Singh and Wesson, 2009). The complexity and scope of the functionality provided by ERP systems coupled with the limited support they provide for aiding in navigation and helping users in error situations underlie the need for users to undergo extensive training on both the ERP system itself and the processes it supports (Jones, Zmud, and Clark, 2008). Users have also been driven to develop detailed notes on how to use the system (Topi, Lucas, and Babaian, 2006) and have come to rely on experts and colleagues for assisting them in error situations and everyday interactions (Babaian et al., 2010).

Improving the usability of ERP systems would undoubtedly benefit both users and companies by improving productivity, reducing costs, and lessening the burden placed on users. The ability to develop such systems is predicated on a comprehensive understanding of the usability issues experienced by users. This understanding is also essential for evaluating and comparing the effectiveness of various usability measurement techniques, yet few research efforts have been directed at in-depth investigations of ERP usability problems. It is this dearth of research that is a prime motivator for the work presented here.

In this paper, we describe a study involving an empirical evaluation of an ERP system that is based on the user-reported critical incidents method (del Galdo et al., 1986; Castillo, Hartson, and Hix, 1997). A critical incident is defined as any event occurring during task performance that is a significant indicator of something positive or negative about usability (Hartson and Castillo, 1998). In our study, critical incidents refer to breakdowns in user-system interaction that have negative consequences ranging from mild to severe. The criticality of incidents, therefore, lies in their ability to reveal something negative about the usability of the system. The critical incidents method involves the user self-reporting on any such incidents that are encountered in performing real tasks under normal working conditions. In our study, three users were asked to report on usability problems, or negative critical incidents, experienced while performing tasks with an SAP ERP system. A usability expert also reviewed all of the user-system interactions to identify any observable problems that were not reported by the users. The 53 unique critical incidents identified using this approach are explored in this study.

OBJECTIVES

The goal of this research is to investigate the applicability of the critical incidents method for unearthing ERP usability issues. We are not aware of any other studies at this time that have applied this approach in the ERP domain. As a first step in this exploratory work, our study was conducted in a laboratory setting with participants reporting on critical incidents encountered during actual usage. The user-reported incidents were augmented with expert-observed accounts to provide a more nuanced understanding of the types of usability issues experienced by users (Hartson, Andre and Williges, 2001). Future studies will build on this work by following this same approach in the workplace.

This study is part of a larger research effort focused on improving ERP system usability by applying the human-computer collaboration paradigm, in which the system is viewed as a collaborative partner with its users, to system design (Babaian, Lucas and Topi, 2006). As part of this research, a novel usability evaluation method called collaborative critique is being developed for evaluating usability from the perspective of human-computer collaboration. Usability problems identified using the approach followed in this study will provide a baseline against which the collaborative critique and other usability measurement methods can be evaluated.

ERP USABILITY STUDIES

There has been extensive research on ERP systems, but most has focused on issues affecting the success or failure of ERP implementations rather than on factors affecting usability. We rely in our research on one of the most widely accepted definitions of usability, which is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-11, 1998, p. 2). This definition views usability as an outcome of user-system interaction within particular environments (Bevan and Mcleod, 1994). While there are other, more product-centered, definitions of usability, in which it is defined as a quality or set of attributes of systems (see, for ex., Cockton, Lavery and Woolrych, 2003, p. 1119; Dumas and Redish, 1999, p. 4), our interest lies in the context of the user's interactions. Our approach is consistent with the general view of system use as a nexus of user, system, and task, where individual-level system usage is defined as "an individual user's employment of one or more features of a system to perform a task" (Burton-Jones and Straub, 2006). However, our goal is to evaluate the efficiency, effectiveness and satisfaction with which individual users can employ features of a system to perform a task, not to measure the duration, breadth and variety of use.

The primary body of work that considers ERP usability falls into the behavioral information systems research category and typically follows the survey-based empirical studies approach. Several of these studies build upon the technology acceptance model (TAM), which posits that perceived usefulness and perceived ease of use determine an individual's intention to use a system. Perceived usefulness is also seen as being directly impacted by perceived ease of use. Studies that build on TAM and its successors for predicting and explaining user acceptance of information technologies (Venkatesh et al., 2003) have shown that positive attitudes toward ERP

systems will result in higher levels of user adoption and implementation success (see, for example, Calisir, Gumussoy, and Bayram, 2009; Bueno and Salmeron, 2008; Amoako-Gyampah, 2007; Amoako-Gyampah and Salam, 2004).

A study by Calisir and Calisir (2004) pursued a TAM-based survey approach for investigating factors affecting ERP usability, where usability was measured solely in terms of user satisfaction. This work extends TAM by focusing on the effects of various interface usability characteristics as well as perceived usefulness and perceived ease of use on the end-user satisfaction of 51 ERP system users in 24 companies. Perceived usefulness was found to have the strongest impact on end-user satisfaction, while learnability had a relatively smaller but still significant effect. Perceived ease of use exerted an indirect effect on satisfaction via perceived usefulness, indicating that users rate ERP systems as less useful if they find them difficult to use. The study also found that a good user guidance scheme improved the learnability of the system and reduced the mental workload, suggesting that easy-to-understand error messages, the possibility of making use of the system without having to learn all of it, the availability of undo and reverse control actions, and the presence of confirming questions before the execution of risky commands may increase both perceived usefulness and learnability.

These behavioral, survey-based studies take the more product-centered view of usability and measure how users perceive certain attributes as useful and/or easy to use, regardless of the particular interaction context. While they provide interesting insights into system usage in general, they are typically at too high a level to enable the identification of particular usability issues with an ERP interface.

Interview-based approaches, in which users can describe in detail the types of problems they encounter with a system, afford the opportunity to delve more deeply and gain greater insights into usability issues than is possible with surveys. Despite these advantages, there are few examples of this type of research in the ERP literature. Topi, Lucas and Babaian (2005) followed this approach in a Fortune 500 company, interviewing ten employees ranging from shop floor workers to upper middle management during the early years of a large-scale ERP implementation. Nine of the employees were users of the system, while one had chosen not to use it. Most of the employees also had access to the ERP system during the interviews so were able to demonstrate the types of usability issues they encounter. The study classified usability issues into six 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. Examples of actual usage problems encountered by the users were provided for each category and highlight the issues faced by ERP users in the workplace.

A benefit of interviews with system users is the detailed descriptions of usability problems encountered during system use that emerge during discussions. A limitation is the reliance on the users' memories of those problems. Users are likely to recall only the more egregious issues; while those are very important, the ones that do not come readily to mind, such as frequent annoyances that the user has learned to live with or problems for which workarounds have been developed, can also have large impacts on usability.

Applying expert inspection techniques is considered a cost-effective alternative to interviews and observations of actual users performing tasks with the system. Outside of industry reports, however, there are few examples of these types of evaluations being applied in the ERP domain. Singh and Wesson (2009) applied an inspection technique called heuristic evaluation to an ERP system. This type of evaluation involves usability experts identifying potential problems with an interface in light of a set of guidelines, or heuristics. In this study, three experts followed a task-based approach for performing a heuristic evaluation of an SAP ERP system based upon Nielsen's commonly used ten heuristics (Nielsen and Mack, 1994) plus five ERP-specific heuristics. To derive the ERP heuristics, the authors first developed a list of seven usability issues by building upon those identified by Topi, Lucas, and Babaian (2005). The seven issues were then mapped to common usability criteria, from which five categories of ERP usability issues emerged: navigation, learnability, task support, presentation (input and output), and customization. The categories were then converted into the heuristics used in the expert evaluation. Results showed that the ERP-specific heuristics identified significantly different potential usability problems with an ERP system than those identified by Nielsen's heuristics.

Like the interview-based method, an expert evaluation provides detailed information on usability issues. A shortcoming is that only potential, rather than actual, issues can be identified since no users are involved (Dumas and Salzman, 2006).

The approach followed in this user study attempts to fill in some of the gaps left by the other methods described here. It is not meant to replace surveys, interviews, or inspection methods, but rather to augment them in order to provide a more complete picture of ERP usability issues. Instead of having to recall usability issues after the fact, users report on problems while performing tasks with the system, and experts base their evaluations on observable problems experienced by those users, as described next.

METHODOLOGY

Our study involved three participants in a laboratory setting who reported on usability problems, or negative critical incidents, that occurred while performing tasks with an ERP system. Video and audio recordings of each user session were obtained with screen-capture software so that all user interactions with the system could be reviewed at a later time for revealing the details of each incident as well as any observable incidents that had not been reported.

Three tasks were specified by the two investigators in order to provide the necessary context for analyzing any critical incidents that were identified. Contemporaneous reporting, in which users reported on incidents as they occurred, and retrospective reporting, in which each user was paired with an investigator for reviewing the recordings of the problems that were encountered, were both used. The combination of contemporaneous and retrospective reporting avoids overloading users during task performance without relying to a great extent on their memories (Akers et al., 2009; Capra, 2002). It also eliminates some of the validity issues that may surface when relying on purely self-reported incidents. The user and the investigator together validate the self-reported issues and uncover unreported ones during the retrospective review process, which is similar to the process followed in Neale, Dunlap, Isenhour, and Carroll (2000). The incident reporting

process therefore involves both subjective and objective aspects: reviewing the recordings can reveal objective measures, such as exploration time, overall time on task, number of errors and mistakes, while the user self-reports provide insights into the users' feelings and perceptions (Hornbaek, 2006).

The user session recordings were also reviewed by one of the investigators at a later date to identify any critical incidents that had not been revealed during either contemporaneous or retrospective reporting. We refer to this process as an expert review, because it identifies issues that were considered problematic from only the investigator's point of view. In usability studies, the term "expert" usually denotes a person familiar with the specific evaluation technique under study and/or someone with general HCI expertise. In our study, the investigator who performed the reviews is both an SAP and a usability expert.

Participants

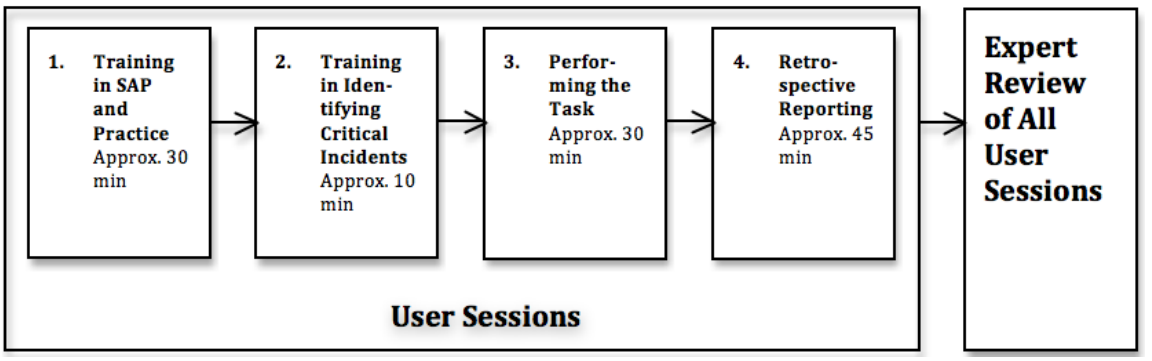
The three participants were males between the ages of 20 and 60 who were enrolled in a graduate program at a business university in the northeastern U.S. Two had prior experience with ERP systems – one with SAP and the other with Oracle. The former had used SAP ERP within a business process management course and was familiar with the interface as well as some of the modules, while the latter had used an Oracle ERP system for about three months. The third user had extensive experience using software applications in general but no experience with any ERP system. Each user received a $25 iTunes gift certificate for participating in the study.

During the study session, each user was equipped with two laptops, one for viewing training videos prepared specifically for this study and the other for practicing and performing tasks with an SAP ERP system. The second laptop had screen-capturing software installed on it that recorded user activities with the system as well as any voiced user comments. Users were instructed to verbalize any issues they encountered during task performance and to also write short comments about them using a reporting feature that was implemented in Java. The constantly visible "Report" button reminded users to keep reporting incidents, while being able to describe the incidents verbally provided a comfortable way for users to go into more detail than they might have been willing to type.

Study Protocol

The study consisted of three approximately 120-minute sessions for gathering data on user-reported critical incidents. Each session consisted of four parts, as shown in Figure 1.

Figure 1 – Overview of Study Protocol



1. *Training in SAP and Practice*: All participants started by simultaneously watching a training video about how to perform a specific task in SAP ERP and practicing the same task with the ERP system. Participants were allowed to take notes and pause the video in order to keep pace with the training activities.
2. *Training in Identifying Critical Incidents*: The participants then viewed a second training video on how to identify and report critical incidents during task performance. In this video, critical incidents were referred to as usability issues to use a more familiar and intuitive terminology. A usability issue was defined as anything that is overly confusing, difficult to understand, requires too much effort, or causes difficulties in task fulfillment. Some problems that an ERP user might run into were demonstrated in the video. These were purposefully not related to any of the tasks the users were asked to perform. Participants were instructed to click on the "Report" button whenever a critical incident was encountered, which would cause a text area to open for entering comments about that incident (see Figure 1), and to verbalize about the incident as well. It was repeatedly stressed that the system, not the user, was being evaluated to encourage full and honest reporting.
3. *Performing the Task (contemporaneous reporting of incidents)*: The three tasks were:

    Task 1. Authorizations task: create a role and a user, assign role to user, test role

    Task 2. Purchasing task: create a purchase order

    Task 3. Sales reporting task: run, sort, export a sales report

    These tasks represent three different ways in which SAP is used in practice: system administration, transaction processing, and reporting. Dumas and Fox (2009) recommend selecting test scenarios by characteristics such as their frequency, their coverage of basic functionality, and the likelihood of usability issues occurring during their performance. Accordingly, the purchasing and sales reporting tasks were chosen because they represent frequently performed scenarios in ERP systems and allow for the testing of two basic functions. The authorizations task, which is considerably longer and more complex than

the other two, was chosen because of the likelihood of usability issues occurring during this task. In a typical work setting, these tasks would be performed by different user populations. In the laboratory setting, two participants were randomly assigned to perform the authorizations task, with the third participant performing the purchasing and sales reporting tasks. Participants were provided with a task description that was detailed enough to enhance their recall of what they had learned in the training but did not provide step-by-step instructions.

4. *Retrospective Reporting:* After the task was completed, one of the investigators and the participant reviewed the video recording and the comments logged during the session. The user was asked to describe in detail what was happening during each incident, if he was confused, what he had expected to happen, and how he had tried to resolve the issue. The user was also asked about any incidents he had encountered but had decided not to report on, with the videotape serving as a reminder of those incidents. Any previously unreported user issues were also documented at that time by the investigator.

After all sessions were completed, an expert reviewed the session recordings. The aim of the separate *expert review* was to identify incidents that actually happened during the session but were not reported by the users either during the task or retrospectively. There are a number of reasons why a user may not have reported an incident, such as considering it to be unimportant or simply forgetting to write it down. The expert identified usability incidents by looking for visible breakdowns in the user's performance, such as when the user was doing something wrong without realizing it, or the user-controlled cursor was observed to be wandering aimlessly around a particular screen for long periods of time. The investigator identified the consequences of each breakdown in terms of lost time, replication of work, erroneous actions/entries, and any effects verbalized by the user.

DATA ANALYSIS

A total of 54 critical incidents were reported, 33 from users and 21 from the independent expert review. One of the 33 user incidents was discarded, as it was related to a change in the screen's color when the "Report" button added to the interface for this experiment was clicked. To facilitate the analysis process, a spreadsheet structure (adapted from Lavery, Cockton, and Atkinson, 1997) was used for describing the remaining 53 incidents. We described each incident in terms of its context (part of the task where the incident occurred), breakdowns in user interactions, and outcomes (user's ability to continue and ways of overcoming the breakdowns). The following describes how critical incidents were transformed into usability problems, which were ranked in terms of their severity. The transformation approach used here was proposed by Howarth, Andre, and Hartson (2007) for reducing the dependency on the skills and experience of the evaluator by having usability problem instances serve as a bridge between the raw data and the usability problems.

Transforming Raw Critical Incidents into Usability Problems

First, general usability problem instances (UPIs) were extracted from the 53 raw critical incidents. Altogether, this process yielded 20 UPIs, eight of which were identified from both user-reported and expert-reported data, another eight were identified from only user-reported data, and four from only expert-reported incidents. The identification process relied on searching for commonalities in the user-system interaction breakdowns across the raw incidents (Table 1 provides illustrative examples). For example, UPI 1: Finding the right buttons is not easy, was extracted from a number of raw critical incidents reported by different users. The common breakdown across these incidents is the difficulty or inability of users to find the correct button. The outcome of a breakdown, however, may differ across incidents and users, ranging from insignificant time losses to not being able to continue without outside help.

Next, UPIs with the same underlying issues were merged into ten general usability problems, or UPs. This two-step process of transforming critical incidents into UPs was performed by both investigators together. Table 2 shows the 20 UPIs and the ten UPs derived from them. Each UPI is identified as being based on user reported critical incidents only, expert-reported incidents only, or both user- and expert-reported incidents. Each UPI is also linked to one or more of the three tasks to show which usability problem instances occurred during which of the tasks. The UPs are shown in sorted order from the most to the least severe. UP3 and UP9 are based purely on user-reported data. All other UPs were derived from both user-reported and expert-reported issues. The data in Tables 1 and 2 add significant contextual detail to our findings and make the analysis and merging processes more transparent.

FINDINGS

UP severity rankings were determined on the basis of problem impact and persistence and on problem frequency. The two investigators, working independently, reviewed each UP and the user-reported critical incidents underlying it (the expert-discovered incidents were not considered in determining impact, frequency, or severity rankings, as these were not reported on by the users). A numeric value was then assigned to each UP using the scale shown in Figure 2. Following the procedure used in Akers et al. (2009), the scores on both variables were added and the final severity ranking was calculated by subtracting one from this sum, resulting in a scale ranging from 1 to 7. The final rankings of the two investigators were the same.

Figure 2 - Scales for Severity Rankings (adapted from Akers et al. 2009)

Problem impact and persistence:
1. minor annoyance, easily learned and worked around
2. bigger problem (at least 3 minutes time lost), but still easily learned or worked around
3. minor annoyance, but will happen repeatedly
4. bigger problem, and will happen repeatedly
5. showstopper (cannot move forward without outside help; data loss; wrong result not noticed)

Problem frequency:
1. some will encounter (at least 1/3, less than 2/3)
2. most will encounter (at least 2/3, less than 100%)
3. everyone will encounter

Table 1 – Illustrative Examples Demonstrating the Merging of Raw Critical Incidents into Usability Problems*

Examples of Raw Critical Incidents	Usability Problem Instance (UPI)	Usability Problem (UP)
User 1: Transaction button was not obvious for adding transactions to role **User 1:** Cannot find Generate button (for auth. profile) (skipped until later) **User 2:** Cannot find Change authorization data button (significant time loss) **User 2:** Cannot find New session button (needed outside help) **User 3:** Find button (in Purch. group search window) was difficult to find	**1.** Finding the right buttons is not easy	**UP1:** It is difficult for users to find the next step (button to push, field to fill, transaction to open) to perform a multi-step task.
User 1: It takes time to find the role creation transaction in main menu **User 3:** It takes ½ of task time to find the sales report transaction	**2.** Findings transactions in the main menu is not easy	
User 1: "No Favorites exist" message (role creation) is not clear **User 1:** There is a beeping sound (no error message) to inform that the username is too long **User 1 and User 2:** Error message "password not downward compatible" (during the setting of initial password for a new user) is not clear **User 3:** Error message "Enter purchase organization" (PO transaction) does not point out where the problem is	**3.** Unclear info/error messages, inappropriate for context	**UP2:** Feedback and information provision is often unclear, unhelpful, and not sensitive to context. Inappropriately placed within the system.
User 2: Information has limited usefulness for novice user (in role creation). Gave overview but not details	**4.** Information tab is unhelpful	
User 3: Warning and information messages appear at the bottom left, are small and easy to miss	**5.** Messages are easy to miss	
User 1: Username is restricted to 12 characters (in user creation), but this is not indicated anywhere.	**8.** Unclear rules for username entry	**UP4:** Basic rules of data entry (formats, restrictions, required fields) are not always obvious to users.
User 1: User enters a username with both lowercase and uppercase letters and cannot log in (case sensitive)	**9.** Unclear login rules of SAP	
User 1: User forgets to enter initial password, tries to go back to Address tab, gets error message "Please enter initial password" (no visual indication that field is required) **User 2:** User clicks Create first without entering a username (in user creation). Message: Enter a user name (no visual indication that field is required and should be filled first)	**10.** Unclear when a field must be filled	

* data on all critical incidents is available from the authors upon request

Table 2 – ERP Usability Problem Instances, Usability Problems and Severity Rankings

Usability Problem Instance (UPI)	Task	Source	Usability Problem (UP)	I*	F*	Sev.*
1. Finding the right buttons is not easy **2.** Finding transactions in the main menu is not easy	1 & 2 1 & 3	Both Both	**UP1:** It is difficult for users to find the next step to perform a multistep task.	4/5	3	6 or 7 severe
3. Unclear info/error messages, inappropriate for context **4.** Information tab is unhelpful **5.** Messages are easy to miss	1 & 2 1 2 & 3	Both Both Users	**UP2**: Feedback and information provision is often unclear, unhelpful, not sensitive to context and inappropriately placed.	3/4	3	5 or 6 severe
6. Only manual entering of wildcards in authorizations tree **7.** Unclear how the authorizations tree has to be filled	1 1	Users Users	**UP3**: Procedures of data entry can be very tedious (and alternatives unknown).	3	2	4 medium
8. Unclear rules for username entry **9.** Unclear login rules of SAP **10.** Unclear when a field must be filled	1 1 1	Users Expert Expert	**UP4:** Basic rules of data entry (formats, restrictions, required fields) are not always obvious to users.	3	2	4 medium
11. Search screen is not clearly differentiable from other screens **12.** The level of abstraction in the purchase order is not clear	1 2	Both Users	**UP5:** It is difficult for users to discern their current location within the system and what is possible at this location.	2	2	3 medium
13. Unclear what exactly the search function does **14.** Unclear why the search function works differently within the same transaction	1 2	Both Users	**UP6**: The functioning of search within transactions is inconsistent and unclear.	2	2	3 medium
15. Visual design of buttons & their placement not always clear **16.** Unclear purpose of the green circle button	1 1	Users Expert	**UP7:** The visual design, purpose and placement of buttons is not clear to users.	3	1	3 medium
17. Function of SAP command line is not clear **18.** It is not obvious how to sort correctly in a report	2 & 3 3	Both Expert	**UP8:** It is difficult for users to understand how some functions in SAP work.	2	1	2 mild
19. Switching between views in the sales report is not easy	3	Users	**UP9**: It is not easy for users to change settings in SAP.	2	1	2 mild
20. Unclear how to select a value from a list	1	Both	**UP10:** Basic navigation and selection within lists is not obvious in SAP.	1	1	1 mild

* I – problem impact; F – problem frequency; Sev – problem severity

Most Severe Usability Problems

Two usability problems, UP1 and UP2, were rated as severe. UP1, *difficulty in finding the next step to perform*, was faced many times by all three participants across the three tasks. In most cases, this problem led to a significant time loss or even a complete breakdown in the interaction. For example, one user spent half of his total task time trying to find the right transaction to begin that task; another forgot how to create a new user in the authorizations task and could not move on until asking for and receiving help from an investigator. Other examples of raw critical incidents related to this usability problem are given in Table 1.

UP2, *the lack of clarity in feedback and information from the system*, was also encountered multiple times by all participants across the three tasks. In general, incidents associated with UP2 were encountered frequently, but they never resulted in the complete breakdown in interactions that resulted from incidents associated with UP1. Often, users just tried to move on and ignore any messages they did not understand. Although this strategy was successful on a number of occasions, it may still leave users confused and can result in unintended consequences. For example, both users performing the authorizations task heard a beeping sound when entering the user name during the process of creating a new user (this happens when a user attempts to enter a name with more than 12 characters). One user became very confused by this and tried a number of different ways to overcome whatever problem was causing the beeping. The user reported the incident as an issue with the system's feedback. The other user just moved on without worrying about the user name being truncated.

Another unclear error message ("*password not downward compatible*"), which was reported by both users performing the authorizations task (see Table 1), was related to the format of the initial password that has to be set for a new user. The third user, who was performing purchasing and reporting tasks, also reported multiple incidents related to the lack of clarity in feedback. For example, he pointed out that it was difficult to see messages when they appeared at the bottom of the system window. In addition, while error messages about missing values in the purchase order transaction did specify the field where the value was missing, they did not indicate where that field was located within the multiple tabs and pages in that transaction.

Medium Severity Usability Problems

Most of the medium severity usability problems (UP3 to UP7) were encountered by two of the three users (sometimes performing the same task, other times performing different tasks) and ranged from bigger but relatively easily overcome problems to minor but repeated annoyances. UP4, *unclear data entry rules*, is merged from instances related to unclear login and username entry rules and unclear indications of required fields. The incidents related to unclear data entry rules (UP4) and tedious data entry procedures (UP3) only occurred during the authorizations task (Task 1). While users reported problems when they did not know the rules, restrictions, or formats for data entry (see UPI 8 in Table 2), which resulted in cryptic feedback (e.g., beeping sounds) from the system, they neglected to report issues related to unmarked required fields or unclear login rules, which resulted in less cryptic messages. For example, when users forgot to

enter a value into a required field, the resulting error messages generally clearly indicated which value was missing; thus, the users could correct their mistakes relatively easily. Such cases were still considered as incidents by the expert, as the users lost time unnecessarily by going back to previous tabs or sections to fill out the required information. More clearly marked required fields would have avoided this time loss.

The log-in incident, identified by UPI 9 within UP4 in Table 1, occurred because usernames are case sensitive, and the user typed in the incorrect case for the user name. He received the standard error message that he could not be logged on because either the username or password were incorrect. While the nature of the mistake was not revealed in the message, it was informative enough for him to realize and correct his mistake in order to successfully log on.

UP5, *difficulty for users to discern their current location within the system and what is possible at this location*, is merged from instances related to the confusing sections in the purchasing order transaction and the difficulty of differentiating the search screen from other parts of a transaction. For example, the user performing the purchasing task (Task 2) reported his confusion about the three levels of data in the purchase order: header, item, and item details. It was difficult for him to see whether he was in the item or item details section, and he was not sure what kind of data should be entered into each section. One of the users doing the authorizations task described during retrospective reporting how he got lost in the search screen without realizing it while trying to create a user. This also happened to the other user performing that task, but he was so confused as to where he was (the user creation screen) and what he was supposed to do (enter a new username and press 'Create') that he failed to report it as an incident. This was reported on later by the investigator during the expert review.

The inconsistent and unclear functioning of search within transactions, UP6, is based on instances related to the difficulty of understanding what the search function does and why it does different things within the same transaction. One user performing the authorizations task ran into trouble when he tried to search for transactions. Instead of getting a dialog box in which to enter the search criteria after clicking on the search button, as was shown in the training video, he got a list of the four most recently searched for transactions. The user reported his confusion, noting: "*Can I search the list of transactions? How do I know that I am viewing the entire list? Was this the entire list? Can you search on type?*" Another user reported an incident related to the inconsistency of the search functionality within the purchase order transaction. To find the purchasing organization, the user can search for a specific value by code or description, but for purchasing groups, there is no way to enter specific search criteria. The user must search through the entire list of groups to find the correct one. Finally, the second user performing the authorizations task ran into the same problem when searching for transactions as the first user, but he did not report this as an incident.

UP7, *unclear visual design, placement and purpose of buttons in the system*, is merged from instances related to the confusing look and feel (i.e., visual properties and associated behaviors), of the buttons, which may make it difficult to discern their purpose. Only one of the users actually reported incidents related to this UP. First, the user indicated that the visual design (i.e., the icon) used for the "Create User" button was very confusing and was in no way related to creating users

in his mind (the icon looks like a sheet of paper). Secondly, the user reported that, in general, the visual design and placement of buttons was inconsistent. For example, "Save" was sometimes indicated by a floppy disc icon and sometimes by a check mark. Both incidents occurred during the authorizations task (Task 1). The expert reported several more incidents associated with a mysterious green button that is visible within most transactions but does nothing when clicked. Users performing the authorizations task clicked on this button repeatedly without any effect but did not report this as an incident. It therefore cannot be determined if and to what extent the button caused any user confusion.

Mild Severity Usability Problems

Mild usability problems (UP8 to UP10) are characterized by low impact and low frequency. Each of the mild problems was encountered by only one user, though some were also reported by the expert. The problem of *difficulty or inability of users to figure out how one or another function works* (UP8) is merged together from instances related to the difficulty in figuring out how the SAP command field, which is used for calling up a transaction by its code, and the sorting in reports work. The user performing the reporting and purchasing order tasks (Tasks 2 and 3) reported the incident related to the command field while trying to search for a specific transaction by name/description. SAP only allows direct access to transactions by typing in the specific code, however, and does not support any searching functionality, so the user was unsuccessful. The lack of available information indicating that this command field did not actually function as a search resulted in user confusion and the need for him to go through the extensive main menu in order to find the right transaction. The same user also encountered UP8 when he was sorting the report incorrectly, as SAP provides no indication as to how the sorting should be performed.

UP9 (*it is not easy for users to change settings in SAP*) and UP10 (*basic navigation and selection within lists is not obvious in SAP*), are both based on only one specific problem instance. In the case of UP9, the user performing the reporting task (Task 3) recorded an incident related to his inability to easily switch between the Excel- and SAP-mode of the report. In the case of UP10, one of the users performing the authorizations task (Task 1) reported his confusion about when to *click* and when to *double click* while navigating and selecting objects within SAP.

The next section summarizes the contributions and implications of these findings.

DISCUSSION

We find that even a small-scale, laboratory-based empirical evaluation combining user-reported and expert-observed data can identify a large number of detailed and real usability issues. These issues provide a useful baseline for comparing the performance of usability evaluation methods and also reveal a rich amount of information for consideration in future ERP design efforts.

The two primary contributions of this study are:

1. The provision of detailed accounts of actually experienced usability issues with an ERP system, resulting in a better understanding of the types and severities of problems

confronting users. Such understanding is critical to the development of more usable systems.

2. Initial testing of a method for unearthing usability problems inherent to an ERP system that does not rely solely on users' memories of past interactions or expert analysis, paving the way for a larger workplace study involving actual ERP users.

With regard to the first contribution, our findings largely corroborate those of Topi, Lucas and Babaian (2005) and Singh and Wesson (2009), the only two studies we are aware of that investigated ERP usability issues in depth. Our ten general usability problems correspond to many of the problem categories described in the two studies. For example, issues with identifying and accessing the correct functionality, lack of transaction execution (or task) support, limited support in error situations, complex layout and unintuitive user interface, confusing search functionality and limited customization opportunities are some of the problem categories identified by both of those studies as well as our own. However, our findings provide a significant amount of novel detail for each of the general usability problems, thereby improving our understanding of user-system interaction issues and providing more user-based input for future design efforts. For example, our findings suggest that functionality problems manifest themselves in the form of users having difficulty in a) finding the correct next step to perform (UP1), b) discerning their current location within a process and the functionalities available at that location (UP5), and c) understanding how some functionalities work in the first place (UP8). For each of these usability problems, we provide even more detailed examples of critical incidents, pointing out, for example, that finding the correct next step is not just about finding the right transaction, but also about finding the right button to push, the next tab to fill out, etc.

We also found specific issues that relate to the lack of transaction execution (or task) support, such as: unclear data entry rules (UP4), tedious data entry procedures (UP3), and inconsistent and unclear navigation within menus and lists (UP10). One of the severe usability problems we found pointed to limited support in error situations: feedback that is unclear, unhelpful, not sensitive to context, and/or inappropriately placed (UP2) is one of the main problems faced by ERP users. As indicated by Tables 1 and 2, there are several different sub-issues here. First, the wording of many messages is confusing and should be clarified. Second, some of the feedback comes in the form of sounds, which is extremely confusing without a corresponding message. Third, many of the messages fail to point out the location of the data causing the error. Fourth, some important messages are easily missed because of poor placement on the screen. Our study also points out specific problems related to unintuitive user interfaces and limited customization opportunities, i.e., the unclear visual design of many icons (UP7) and the limited options for users to adapt the system according to their wishes (UP9) when switching between MS Excel and SAP views of a report.

With regard to our second contribution, we argue that combining user-reported and expert-observed critical incidents ensures that the identified usability problems are based on what really affects users (actual, rather than potential issues), including breakdowns they may not have been aware of, forgot to report, or chose to ignore for whatever reason but were observed to interfere with their interactions with the system. Many of the usability problem instances were extracted from both user-reported and expert-observed data, lending additional support to the importance of

those issues. For example, all users repeatedly recorded incidents related to information and error messages, but the expert review revealed a few additional examples of such incidents. It could be that, as users go through a task and encounter multiple incidents related to the same issue, they becomes less likely to continue reporting on it. Combining the user reported incidents and expert review may, therefore, provide a more realistic view of the impact of these incidents on usability. We also found usability problem instances that were based on purely user-reported incidents. These were mostly related to problems where visible breakdowns or behavioral cues are missing; hence, expert review of recordings cannot easily reveal such incidents. For example, incidents with easy-to-miss information messages or inefficient entry of wildcards are difficult to directly observe from video recordings, as they often do not cause a breakdown, but are nonetheless a significant inconvenience to the user. Finally, four usability problem instances were identified based purely on expert-reported data, with users reporting no equivalent incidents. For example, expert-observed incidents were related to the already-mentioned log-in and unmarked required fields issues. It is unlikely that such problems would be identified in interview studies since the users did not even note them as they happened, let alone after the fact. This does not mean, however, that the incidents are irrelevant, as they may have had a significant impact but the user was too confused to even notice or may have been incapable of reporting on any one specific incident.

Implications

Our detailed findings have both short- and long-term implications for enterprise system design and use. In the short-run, the identified critical incidents and usability problems point to potential improvements in user-system interactions that may be achievable immediately through training. Spending more time during training on introducing new users to the ERP main menu, teaching them how they can create their own list of favorites when such an option exists, familiarizing them with SAP's style of structuring transactions into tabs and views, as well as pointing out some of the more frequent error situations and how to resolve them, if possible, may save them time and frustration later on. Furthermore, our findings related to such problems as tedious data entry procedures and difficulties in changing system settings indicate that training should not be restricted to only novice users. 'Tips and Tricks' training sessions for more expert users or a forum for sharing favorite shortcuts can be very beneficial to those who are already familiar with the ERP system but would benefit from the experiences of other users.

In the longer term, our findings provide rich information that can serve as a basis for designing more usable ERP systems. While the identified usability problems do point to certain design elements for alleviating some of the issues faced by users, it is crucial to point out that an integrated approach that addresses the systemic nature of the poor usability of ERP systems, rather than the application of isolated fixes, is required. For example, when dealing with navigation issues, breadcrumbs can be used to indicate the current location, while a more complex task flow visualization would be useful for indicating both the current location and the next correct step; automatic population of a 'Favorite transactions' list would reduce the necessity of going through the immense main menu of SAP or the memorization of transaction codes; and auto-completing text inputs and system-proposed corrections to data entry errors could save the users a lot of frustration and reduce the time spent seeking external help. Such interventions,

however, must be part of a well-defined approach that considers how to holistically design ERP systems from the ground up in order to provide a consistent and coherent user experience.

LIMITATIONS

The study, which was conducted as part of a pilot evaluation of the Collaborative Critique method, has two main limitations. First, it had only three users. Despite the small number, we feel confident that our results represent a real, rich-in-detail, and informative set of ERP system usability problems. We are encouraged by prior studies on empirical usability evaluations that have also shown that three to five users are enough to determine 80% of usability problems with a system (see Hartson, Andre and Williges, 2001).

Secondly, our participants were students with some ERP experience rather than everyday users of ERP systems; thus, the critical incidents they reported may differ somewhat from incidents that would be reported by users interacting with ERP systems on a regular basis as part of their work. Our findings demonstrate, however, that even students with relevant ERP experience can uncover a significant amount of detailed issues with the critical incidents method, lending further support to the applicability of the method to ERP usability evaluations.

CONCLUSIONS AND FUTURE WORK

The laboratory-based critical incidents approach, which draws on both user experience and expert analysis, provides valuable insights into the usability issues plaguing users of ERP systems. Combining both user- and expert-reported critical incidents enables the identification of a more detailed set of actual ERP usability problems than is possible from evaluations based on either purely user or expert input. Furthermore, expert input that largely confirms what users have reported lends further support to the importance of those issues. Expert input alone can also indicate real issues ranging from mild, such as when the user clicks on a button that does not do anything, to severe, as when the user makes critical mistakes without realizing it. The results of this study and the approach we have followed are an important step in working toward the goal of improving the usability of ERP systems.

We plan on building on this study with additional participants and will ultimately test this approach in the workplace with actual ERP users. Findings from this and future studies will improve our understanding of ERP usability issues, which is essential for enhancing system design, and will provide baselines for evaluating usability measurement methods, such as the collaborative technique currently under development.

ACKNOWLEDGEMENTS

We would like to thank Tamara Babaian for her help in designing and conducting this study. This material is based in part upon work supported by the National Science Foundation under Grant No. 0819333. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Akers, D., Simpson, M., Jeffries, R., & Winograd, T. (2009). Undo and erase events as indicators of usability problems. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*. New York, USA: ACM Press, 659-668.
2. Amoako-Gyampah, K. & Salam, A. (2004). An extension of the technology acceptance model in an ERP implementation environment. *Information & Management*, 41(6), 731–745.
3. Amoako-Gyampah, K. (2007). Perceived usefulness, user involvement and behavioral intention: An empirical study of ERP implementation. *Computers in Human Behavior*, 23(3), 1232-1248.
4. Babaian, T., Lucas, W. T. & Topi, H. (2006). Improving ERP Usability Through User-System Collaboration. *International Journal of Enterprise Information Systems*, 2(3), 10-23.
5. Babaian, T., Lucas, W. T., Xu, J. & Topi, H. (2010). Usability through system-user collaboration: Deriving design principles for greater ERP usability. Paper presented at the 5th International Conference on Design Science Research in Information Systems and Technology, St. Gallen, Switzerland.
6. Bevan, N. & MacLeod, M. (1994). Usability measurement in context. *Behaviour & Information Technology*, 13(1), pp. 132-145.
7. Bueno, S. & Salmeron, J. (2008). TAM-based success modeling in ERP. *Interacting with Computers*, 20(6), 515-523.
8. Burton-Jones, A. & Straub, D.W. (2006). Reconceptualizing System Usage: An Approach and Empirical Test. *Information Systems Research*, 17(3), 228-246.
9. Calisir, F. & Calisir, F. (2004). 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.
10. Calisir, F., Gumussoy, C. & Bayram, A. (2009). Predicting the behavioral intention to use enterprise resource planning systems: An exploratory extension of the technology acceptance model. *Management Research News*, 32(7), 597-613.
11. Capra, M.G. (2002). Contemporaneous versus retrospective user-reported critical incidents in usability evaluation. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 1973-1977.
12. Castillo, J.C., Hartson, H.R. & Hix, D. (1997). The user-reported critical incident method at a glance. *Technical Report: TR-97-13*, Virginia Polytechnic Institute & State University, Blacksburg, VA, USA, 1-7.
13. Cockton, G., Lavery, D., & Woolrych, A. (2003). Inspection-based evaluations. In *The Human-Computer Interaction Handbook*, Jacko, J.A. & Sears, A. (Eds.) Mahwah, NJ: Lawrence Erlbaum Associates.
14. del Galdo, E.M. et al. (1986). An Evaluation of Critical Incidents for Software Documentation Design. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Human Factors and Ergonomics Society, 19-23(5).
15. Dumas, J.S. & Redish, J.C. (1999). A practical guide to usability testing. Bristol, UK: Intellect, Ltd.
16. Dumas, J.S. & Fox, J.E. (2009). Usability Testing: Current Practice and Future Directions. In *Human-Computer Interaction. Development Process*, Sears, A. & Jacko, J.A. (Eds.). Boca Raton, FL: Taylor and Francis Group, LLC.
17. Dumas, J.S. & Salzman, M.C. (2006). Usability Assessment Methods. *Reviews of Human Factors and Ergonomics*, 2(1), 109-140.
18. Gartner, Inc. (2011). Forecast: Enterprise Software Markets, Worldwide, 2008-2015, 2Q11 Update. Retrieved August 21, 2011 from: http://www.gartner.com/DisplayDocument?id=1724639

19. Gould, J.D. & Lewis, C. (1985). Design for usability: Key principles and what designers think. Communications of the ACM 28(3), 300-311.
20. Hamerman, P.D. (2008). ERP applications 2008: The battle goes vertical. Forrester Research.
21. Hartson, H.R., Andre, T.S. & Williges, R.C. (2001). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, 13(4), 373-410.
22. Hartson, H.R. & Castillo, J.C. (1998). Remote evaluation for post-deployment usability improvement. In *Proceedings of the Working Conference on Advanced Visual Interface*, 1- 10.
23. Herbert, L. (2006). Put business applications to the usability test. Forrester Research.
24. Hestermann, C. (2009). Key issues for enterprise resource planning 2009. Gartner Research.
25. Hornbaek, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79-102.
26. Howarth, J., Andre, T., & Hartson, R. (2007). A structured process for transforming usability data into usability information. *Journal of Usability Studies*, 3(1), 7-23.
27. Hurtienne, J., Prümper, J. & Rötting, M. (2009). When enterprise resource planning needs software ergonomics – Some typical scenarios. In *Proceedings of the 17th World Congress on Ergonomics.*
28. Iansiti, M. (2007). ERP end-user business productivity: A field study of SAP & Microsoft: Keystone Strategy. Retrieved March 30, 2011 from: http://download.microsoft.com/download/4/2/7/427edce8-351e-4e60-83d6-28bbf2f80d0b/KeystoneERPAssessmentWhitepaper.pdf.
29. ISO 9241-11 (1998). Ergonomics requirements for office work with visual display terminals (VDTs) – part 11: Guidance on usability. International Standards Organization.
30. Jones, M.C., Zmud, R.W. & Clark, T.D. (2008). ERP in practice: A snapshot of post-installation perception and behaviors. *Communications of the Associations for Information Systems*, 23(1), 437-462.
31. Lavery, D., Cockton, G. & Atkinson, M.P. (1997). Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16(4/5), 246-266.
32. Neale, D.C. et al. (2000). Collaborative Critical Incident Development. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Human Factors and Ergonomics Society, 598-601(4).
33. Nielsen, J. & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people* (CHI '90), Jane Carrasco Chew and John Whiteside (Eds.). NY, USA: ACM, New York, 249-256.
34. Otter, T. (2008). Case study: Ness combines consumer application ease of use with ERP robustness. Gartner Research.
35. Rosson, M.B. & Carroll, J.M. (2002). *Usability engineering: scenario-based development of human-computer interaction.* San Francisco, CA: Morgan Kaufmann Publishers Inc.
36. Singh, A. & Wesson, J. (2009). Evaluation criteria for assessing the usability of ERP systems. In *Proceedings of SAICSIT* 2009. Riverside, Vanderbiljpark, South Africa: ACM, 87-95.
37. Topi, H., Lucas, W. & Babaian, T. (2005). Identifying usability issues with an ERP implementation. In *Proceedings of the 7th International Conference on Enterprise Information Systems*, 128-133
38. Topi, H., Lucas, W. & Babaian, T. (2006). Using informal notes for sharing corporate technology know-how. *European Journal of Information Systems*, 15(5), 489-499.
39. Venkatesh, V., Morris, M.G., Davis, G.B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly, 27*(3), 425-478.
40. Woods, J. (2011). Key issues for ERP and enterprise suite strategies and value realization. Gartner Research.
41. Yeh, J.H. (2006). Evaluating ERP performance from user perspective. IEEE Asia-Pacific Conference on Services Computing (APSCC'06), 2006, 311-314.