

EVALUATING THE USABILITY OF ERP SYSTEMS: WHAT CAN CRITICAL INCIDENTS TELL US?

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 notoriously difficult to use. The scope and complexity of their functionality can be overwhelming for users, who must typically undergo extensive training before they can make effective use of these systems. Understanding the specific usability problems experienced by users is essential to the development of more usable systems. This paper investigates 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. The resulting 53 incidents were categorized into 10 usability problems and ranked from mild to severe. We argue that this approach yields a more detailed and representative view of ERP usability problems than that provided by expert evaluations alone, while being less dependent on users’ memories than interview-based studies.

Keywords: ERP systems, Usability, Critical Incidents

Introduction

It is commonly acknowledged that Enterprise Resource Planning (ERP) systems can be difficult to use (Hurtienne, Prümper and Rötting 2009), but few research efforts have focused on identifying the types of usability problems encountered by users. Usage problems are often attributed to the scope and complexity of the functionality supported by these systems, making it difficult for users to find the information and transactions they need. Limited support is offered for aiding in navigation or helping users in error situations, and options for personalization are also lacking (Topi, Lucas and Babaian 2005; Singh and Wesson 2009). Training on both the system and new work processes is therefore necessary for successful ERP usage (Jones, et al. 2008).

Existing research on ERP usability tends to be based on surveys (Amoako-Gyampah and Salam, 2004; Amoako-Gyampah, 2007; Calisir and Calisir, 2004), user interviews (Topi, Lucas, and Babaian, 2005), or expert evaluations (Singh and Wesson, 2009). While such studies have provided valuable data and interesting insights into ERP systems usability in general, the findings are typically at an abstract level, hindering their practical application to improving usability. Knowing more about the specific issues experienced by user is an essential step toward enhancing the design of these systems.

The goal of this research is to develop an approach that extends existing ERP usability studies by examining problems encountered in actual usage from both the user and expert perspective. To this end, we have conducted a pilot study involving an empirical evaluation of an SAP ERP system that is based on the critical incidents method (Castillo, et al. 1997), in which users self-report critical incidents encountered in performing real tasks under normal working conditions. 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).

Our pilot study involved three participants in a laboratory setting who reported on negative critical incidents that occurred while performing tasks with an ERP system (note that prior studies on empirical usability evaluations such as Hartson et al. 2001, have shown that three to five users are adequate for determining 80% of a system's usability problems). All interactions between the users and the system were recorded and reviewed to reveal any incidents that had not been reported. A total of 53 incidents were identified and subsequently categorized into 10 usability problems.

The major contributions of this work are:

- The 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.
- The 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 real ERP users.

This study is part of a larger research effort focused on improving ERP system usability by applying the human-computer collaboration paradigm 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 the extent to which a system collaborates with its users in helping them achieve their system-related goals. The usability problems identified by this pilot, coupled with those from future studies building on this work, will provide a baseline against which the collaborative critique and other usability measurement methods can be evaluated.

Methodology

This study employed the critical incidents method for reporting on the usability problems encountered by three participants performing tasks with an SAP ERP system in a laboratory setting. These tasks were specified by the investigators, providing us with the necessary context for analyzing any identified critical incidents. We used both contemporaneous reporting, in which users reported on the incidents as they occurred, and retrospective reporting, in which the users reviewed the problems they encountered with one or both of the investigators using video and audio recordings of their sessions obtained from screen-capture software. The combination of contemporaneous and retrospective reporting avoids overloading users during task performance without depending to a great extent on their memories (Akers et al. 2009; Capra 2002). Those same recordings were also reviewed by one of the investigators at a later date to identify any critical incidents that had not been reported.

Participants

The three participants were males between ages 20 and 60 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 third had extensive experience using software applications in general but no experience with an ERP system. Each user received a \$25 iTunes gift certificate for their participation.

Setup and Protocol

The study consisted of three approximately 120 minute sessions – one per participant – for gathering user-reported critical incidents. Each user was equipped with two laptops, one for viewing training videos and the other for practicing and performing tasks with the ERP system. The latter laptop had screen-capturing software installed for recording all user interactions as well as any voiced comments.

Each session consisted of four parts:

Training in SAP and Practice (about 30 minutes) Participants watched and practiced along with a training video about how to perform a specific task using SAP ERP. They were allowed to take notes.

Training in Identifying Critical Incidents (about 10 minutes) Participants viewed a training video on how to identify and report on critical incidents, which were referred to as usability issues. 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 a “Report” button whenever a critical incident was encountered, which would cause a text area to open for entering comments about that incident, 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.

Performing the Task (contemporaneous reporting of incidents) (about 30 minutes)

The three tasks were:

1. Authorizations task: create a role, create a user, assign role to user, test role
2. Purchasing task: create a purchase order
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. The first task was considerably longer and more complex than the other two; two participants were randomly assigned to perform that task, with the third participant performing the second and third 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.

Retrospective Reporting (about 45 minutes)

After the task was completed, one or two investigators reviewed the video recording and the comments logged during the session with the participant. The user was asked to describe in detail what was happening during the 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, which were documented by the investigator.

Expert Review

The aim of the expert review was to identify incidents that actually happened during the session but were not reported by the users, who may have considered them to be unimportant or could simply have forgotten to report them. 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 reviewer (one of the two study investigators) was both an SAP and usability expert and identified incidents by looking for visible breakdowns in the user’s performance, such as when the user was doing something wrong without realizing it or wandering aimlessly around the same screen for long periods of time. The investigator identified the consequences of each breakdown in terms of lost time, replication of work, and any other 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 we added was clicked. To facilitate the analysis process, a spreadsheet structure (adapted from Lavery et al. 1997) was used for describing the remaining 53 incidents in terms of their context, breakdowns in user interactions, and outcomes.

Transforming Raw Critical Incidents into Usability Problems

Critical incidents were transformed into usability problems (UPs) using a structured approach proposed by Howarth (2007), which reduces the dependency on the skills and experience of the evaluator by having usability problem instances (UPIs) serve as a bridge between the raw data and the UPs. The investigator who performed the expert review performed this transformation, first by noting the underlying larger issues of each of the 53 incidents, and then by grouping the incidents on the basis of those shared issues into 20 UPIs. Next, UPIs with the same underlying issues were merged into the 10 general usability problems.

Severity Rankings

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, as their impact was not reported on by users). A numeric value was then assigned to each UP using the scale shown in Figure 1. 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 from 1 to 7. The final rankings of the two investigators were the same (see Table 1).

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

Figure 1: Scales for Severity Rankings (adapted from Akers et al. 2009)

Findings

The 10 UPs identified from the 53 critical incidents are shown in Table 1, sorted from most to least severe. UP1, *difficulty in finding the next step to perform*, was faced many times by the three participants. In most cases, this UP resulted in a significant time loss or even a complete breakdown in the interaction. 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 an investigator for help.

UP2, *the lack of clarity in feedback and information from the system*, was also encountered multiple times by all participants. Sometimes a participant would just ignore an unclear message and continue with the task; while this was often a successful strategy, it left users feeling confused. Both users performing the authorizations task, for example, heard a beeping sound when entering the user name (this happens when a user name with more than 12 characters is entered). One user became very confused in trying to resolve this issue. The other user chose to ignore the beeping and move on.

Most of the medium severity usability problems (UP3 – UP7) were encountered by two users and ranged from bigger but still easily learned problems to minor but repeated annoyances. For UP4, *unobvious data entry rules*, incidents were related to users not knowing the rules, restrictions, or formats for data entry. For UP7, *unclear visual design and placement of buttons*, one user indicated that the icon for the "Create User" button was very confusing (it looks like a piece of paper) and that the visual design and placement of buttons was inconsistent,

with “Save” sometimes indicated by a floppy disc icon and sometimes by a checkmark. UP7 also included an expert-revealed UPI concerning a mysterious green button that is visible within most transactions but causes no discernable action when clicked. Users clicked on this button repeatedly to no effect but did not report this as an incident.

Mild usability problems (UP8 – UP10) were characterized by low impact and low frequency. An incident associated with UP8, *difficulty in determining how some functions work and their purpose*, was encountered by a user who was trying to search for a specific transaction by its name using the “shortcut to transactions” function. The user was unsuccessful because this function requires typing in the transaction’s code. No information was provided indicating this, resulting in confusion and a subsequent search through the extensive main menu for finding the right transaction. The expert review revealed another issue related to UP8 for this same user, which occurred when the user was attempting to sort a report. The user did not notice he had sorted incorrectly and thus failed to report the incident.

UP	Usability Problem (UP) Description	Severity
1	It is difficult for users to find the next step (i.e., the button to push, the field to fill, the transaction to open) in performing a multi-step task.	6 or 7 (severe)
2	Feedback and information provision is often unclear, unhelpful, not sensitive to context, and inappropriately positioned within the system.	5 or 6 (severe)
3	Procedures of data entry can be very tedious (with alternatives unknown to users).	4 (medium)
4	Basic rules of data entry (i.e., formats, restrictions, required fields) are not always obvious to users.	4 (medium)
5	It is difficult for users to discern the current location within the system and what is functionally possible at that location.	3 (medium)
6	The functioning of the Search feature within transactions is inconsistent and unclear.	3 (medium)
7	The visual design (i.e., labels, icons) and placement of buttons in the interface are often unclear to users.	3 (medium)
8	It is difficult for users to understand how some functions actually work, and the purpose of these functions is unclear.	2 (mild)
9	It is not easy for users to change certain settings or adapt the system according to their wishes.	2 (mild)
10	Basic navigation and selection within lists is not obvious or consistent.	1 (mild)

Table 1: ERP Usability Problems and Severity Rankings

The above examples demonstrate the types of detailed usability problems that even a small pilot study can reveal. An initial comparison of our results with those from other studies suggests that approaches employing experts, such as heuristic evaluations (Singh and Wesson 2009), are useful for providing generic examples of problem categories but are less useful for revealing the realities of system use. In-depth interviews (Topi, Lucas and Babaian 2005) are more likely to bring usage problems to light, but they are subject to users’ memories of past problems. We argue that the critical incidents-based methodology outperforms both heuristic evaluations and interviews because it identifies more detailed usability problems (e.g. unclear data entry rules, tedious data entry processes) and offers reasonable coverage of the evaluated system without relying purely on either user memory or expert data.

Conclusions

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. We plan on building on this study with additional participants for improving the repeatability of our findings and will ultimately test this approach in the workplace with ERP users. Findings from this and future studies will improve our understanding

of ERP usability issues, which is essential to improving system design, and 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 the pilot 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

- Akers, D. et al. "Undo and erase events as indicators of usability problems," In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* 2009. New York, USA: ACM Press, pp 659-668.
- Amoako-Gyampah, K., and Salam, A. "An extension of the technology acceptance model in an ERP implementation environment," *Information & Management* (41:6) 2004, pp 731-745.
- Amoako-Gyampah, K. "Perceived usefulness, user involvement and behavioral intention: an empirical study of ERP implementation," *Computers in Human Behavior* (23:3) 2007, pp 1232-1248.
- Calisir, F., and Calisir, F. "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) 2004, pp 505-515.
- Capra, M.G. "Contemporaneous Versus Retrospective User-Reported Critical Incidents in Usability Evaluation," In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* 2002, pp 1973-1977.
- Castillo, J.C., Hartson, H.R., and Hix, D. "The User-Reported Critical Incident Method at a Glance," *Technical Report: TR-97-13* 1997, Virginia Polytechnic Institute & State University, Blacksburg, VA, USA, pp 1-7.
- Hartson, H.R., Andre, T.S., and Williges, R.C. "Criteria for Evaluating Usability Evaluation Methods," *International Journal of Human-Computer Interaction* (13:4) 2001, pp 373-410.
- Hartson, H.R., and Castillo, J.C. "Remote evaluation for post-deployment usability improvement," In *Proceedings of the Working Conference on Advanced Visual Interface* 1998, pp 1- 10.
- Howarth, J. "A Structured Process for Transforming Usability Data into Usability Information," *Journal of Usability Studies* (3:1) 2007, pp 7-23.
- Hurtienne, J., Prümper, J., and Rötting, M. "When Enterprise Resource Planning Needs Software Ergonomics – Some Typical Scenarios," In *Proceedings of the 17th World Congress on Ergonomics* 2009.
- Jones, M.C., Zmud, R.W., and Thomas, D.C. "ERP in Practice: A Snapshot of Post-Installation Perception and Behaviors," *Communications of the Associations for Information Systems* (23:1) 2008, pp 437-462.
- Lavery, D., Cockton, G., and Atkinson, M.P. "Comparison of evaluation methods using structured usability problem reports," *Behaviour & Information Technology* (16:4/5) 1997, pp 246-266.
- Singh, A., and Wesson, J. "Evaluation Criteria for Assessing the Usability of ERP Systems," In *Proceedings of SAICSIT* 2009. Riverside, Vanderbiljpark, South Africa: ACM, pp 87-95.
- Topi, H., Lucas, W., and Babaian, T. "Identifying usability issues with an ERP implementation," In *Proceedings of the 7th International Conference on Enterprise Information Systems* 2005, pp 128-133