In this paper, we discuss one aspect of our SMArTHF study where usability testing afforded us a means of assessing design changes and obtaining user feedback on a recently implemented smart IV pump.

Introduction

In our SMArTHF (Safe Medication Administration Through Technology and Human Factors http://cqpi2.engr.wisc.edu/smarthf/index.html) study we are evaluating the implementation of a new technology – smart IV pumps (and ultimately their integration with bar coding administration technology) – at a large university hospital by assessing the impact of the technology on 1) medication error frequency, type and outcome, and 2) end user quality of working life and perceptions of safety. Smart IV pumps have built-in dose error reduction systems that alert users to medication doses programmed outside of preset medication limits. Because of the organization’s commitment in resources (both monetary and personnel) to ensure a smooth implementation, and the fact that the particular technology is relatively new to the market, close scrutiny of both the pump’s design and its performance occurs. When, in collaboration with the manufacturer, an opportunity arose to improve the physical design of the pumping module, the multidisciplinary research team convened and proposed that the redesigned pump should undergo usability testing and an evaluation of specific design aspects related to human-computer interaction.

Nielsen proposes five facets of conducting usability testing (Nielsen 1993): learnability, efficiency, memorability, errors and satisfaction. Our research team chose to evaluate the redesigned smart IV pump based on its performance as measured by 1) the incidence of errors during testing and 2) subjects’ perceptions of reliability, dependability, and ease-of-use (as collected via a pre-/post-survey questionnaire). Because of feedback from current users, and as a means of evaluating the training all these users were required to undergo, we also assessed the pump’s memorability (as captured through observations made by the evaluation team – ASH, TL and BH).

Factors prompting the evaluation

Prior to implementing the smart IV pump, a group of stakeholders involved in the pump’s implementation, use and follow up conducted a prospective risk analysis, specifically an HFMEA™ (Wetterneck, Skibinski et al. 2004). The team recommended that all pump users (nurses, anesthesiologists and CRNAs) be required to complete training prior to using the new pump. The team had the foresight to recognize both the subtle and
not-so-subtle differences between the “new” pumps and those previously in use as well as some counter-intuitive aspects of the new product’s design. Likewise, they recognized that certain features that offered users significant levels of safety, such as the built-in medication library, would infrequently be used by some users and therefore possibly forgotten.

Within three months of the implementation users at the hospital identified a design issue that affected the pump’s performance and safety. Ongoing dialog between the manufacturer, members of the pump “oversight team”, and the research team occurred until a redesigned pumping module was unveiled and presented to the hospital for the testing. Usability testing conducted by the manufacturer (that included users from a different hospital) suggested that the new design adequately addressed the issues identified.

Seven months after the initial implementation the first round of evaluations began. Within the first seven tests conducted it became obvious that the redesign was not effectively addressing the issue that instigated the redesign process. Usability testing halted immediately. Not only was the design inadequate but it was deemed unwise to expose users to issues that would lessen their confidence in the pump’s performance. Consequently the manufacturer exerted considerable effort to more fully understand the issue at hand and within one month a further-refined design was developed and presented. The product was then shared with the evaluation team for testing and decision making on whether to recommend implementation of the redesigned product.

Design of the evaluation

Approximately two months prior to the first round of tests, representatives from the research team offered input to a smaller group that was responsible for conducting the evaluation. This group included one human factors engineer researcher (ASH), the hospital’s IV pump coordinator (TL) and a biomedical engineer (BH). In turn, they designed four pump use scenarios, the detailed testing plan, the testing script and data collection instrument, and a questionnaire. They also defined each of their roles during the testing. The team agreed that, if necessary, just-in-time training should occur (in most instances at the conclusion of the testing session) to answer any questions that arose and/or point out errors in use that were observed. All testing had to be efficiently carried out since all subjects willing to participate in the tests were interrupting their daily work routine. In an attempt to assimilate the care environment, testing occurred in close proximity to the patient care units for the nurse subjects and all anesthesia personnel testing occurred in an unused but readily accessible operating room (one designated for transplant donor harvesting).

Side-by-side testing of three pumping modules (each having a different physical design) for each scenario, and in some instances, iterative testing of pumping modules within a scenario, would occur. To ensure that all issues would be completely assessed, all levels of users, including staff nurses, super-user nurses, and anesthesia staff would be approached to participate in the testing. The pre-/post-survey questionnaire included selected questions from Shneiderman’s (Shneiderman 1998) instrument that assess reliability, dependability and ease-of-use. The same document was used for both response phases. A black or blue pen was used to answer the questionnaire pre-testing and a red pen was used to record and distinguish the post-testing responses.

Conducting the evaluations

The project was approved by the Institutional Review Board prior to recruiting any subjects. Subjects implied consent by expressing a willingness to proceed with the testing after reading an “information sheet”. Once “consented” each subject was asked
to complete the short questionnaire and, once this was completed, was then re-briefed on the study’s intent. Aspects of the usability testing design were also shared and subjects received instructions to think aloud (Nielsen 1993) to allow the team members to capture cognitive aspects of pump use that cannot otherwise be collected by observation.

The four scenarios addressed general issues related to pump use as well as the relevant issues associated with the pump redesign. Because the redesign affected the physical design of the pump door as well as a recess above the door through which IV tubing is placed, some groups (including the manufacturer) expressed concern that users would have to exert greater force when inserting the tubing and closing the pumping module’s door. A scenario (“A”) was designed but discontinued after the first (initial) round of testing for two primary reasons: the “issue” of note was a “non-issue” as became evident by subjects’ inconsistent responses to the level of difficulty associated with the door-closing process. In addition, this scenario took a considerable amount of time and detracted from what was intended to be the emphasis of the testing.

Scenario B was developed to alert subjects to one of the pump’s automation surprises (Sarter, Woods et al. 1997). As users became adept with pump use and displayed greater confidence in the pump’s “smart” design, concern arose that there was an over-reliance on the alarm features. Alarms were designed to inform users of all pump loading and programming errors – at least this is what all users believed. One aspect of improper loading, placing the tubing in the pumping module incorrectly, a process that could result in an under-infusion, did not result in an alarm. Scenario B confirmed the suspicions of the evaluation team and others: that subjects, although aware they improperly loaded the tubing, did not believe there was any consequence in the amount of medication or fluid being infused because “there was no alarm”. As a result, this has been incorporated in all new-user training and is presented to experienced users on an ongoing basis – both on-the-job and through written (and visual) updates distributed to staff.

Scenario C was the most prescriptive and complicated and was designed to definitively assess whether the redesign accomplished what it was intended to accomplish. In this scenario, subjects were instructed to load the IV tubing in a very exact fashion that was known to cause improper infusion, twice for each of the three pumping modules, and also program the pump. This loading included the manner in which the tubing was “threaded”, how the pumping module door was closed, and how the pump was programmed. The loading outcome was recorded and included: 1) whether or not an alarm sounded, 2) what message accompanied the alarm (if it sounded), 3) the outcome of the set-up including the type of IV flow that resulted – fast, slow, none, and 4) subject comments during the scenario. Two individuals (TL and ASH) observed and recorded subject observations, as well as the loading outcome, using a standard data collection instrument that was compared at the end of each evaluation. The other individual present (BH) manned the audio-recording device, ensured smooth progress between scenarios, and recorded general observations at the end of each evaluation session.

If, after completing the previous scenarios, the subject was willing and available to perform one more scenario she was asked to set up three pumps as she normally would. This scenario (“D”) allowed the team to observe errors in pump use and set-up and then give subjects feedback on their respective technique. Nine of the eighteen subjects completed this scenario, most exhibiting appropriate loading and programming technique. Prior to leaving the room, subjects completed the post-testing questionnaire.

Round two of the usability testing occurred over a three-week period. Weekly the usability team held conference calls with the manufacturer to share results of the testing and obtain insights and suggestions from the manufacturer on continued testing design. Results of both the observational component and the survey questionnaire supported replacing the existing pump with the re-designed pump institution-wide.
Conclusion

The evaluations on the redesigned intravenous infusion pump by the hospital and research team complemented the vendor usability testing. The end result was a better design with less chance of error occurring from improper infusion set-up. We found our application for pump evaluation to be useful and believe it accomplished all of the objectives we had developed prior to initiating the process. Aside from subjects’ general difficulty in “thinking aloud” to share their cognitive perspectives, we would replicate the process if called on again. At this point we plan to design vignettes (comparable to the scenarios used here) to further assess pump learnability and memorability and incorporate findings in future user training.

Acknowledgments

Funding for the SMArT HF (“Safe Medication Administration Through Technology and Human Factors”) project is provided by the Agency for Healthcare Research and Quality (Principal Investigator (PI): P. Carayon, co-PI: T. B. Wetterneck; Grant # 1 UC1 HS014253-01).

References