



Published in final edited form as:

*Stud Health Technol Inform.* 2009 ; 143: 322–327.

## Comparative study of heuristic evaluation and usability testing methods

Thankam Paul Thyvalikakath, BDS, MS<sup>a</sup>, Valerie Monaco, PhD<sup>b</sup>, Himabindu Thambuganipalle, BDS<sup>c</sup>, and Titus Schleyer, DMD, PhD<sup>a</sup>

<sup>a</sup>Center for Dental Informatics, School of Dental Medicine

<sup>b</sup>Department of Biomedical Informatics, University of Pittsburgh

<sup>c</sup>College of Dentistry, New York University

### Abstract

Usability methods, such as heuristic evaluation, cognitive walk-throughs and user testing, are increasingly used to evaluate and improve the design of clinical software applications. However, there is still some uncertainty as to how those methods can be used to support the development process and evaluation in the most meaningful manner. In this study, we compared the results of a heuristic evaluation with those of formal user tests in order to determine which usability problems were detected by both methods. We conducted heuristic evaluation and usability testing on four major commercial dental computer-based patient records (CPRs) which together cover 80% of the market for chairside computer systems among general dentists. Both methods yielded strong evidence that the dental CPRs have significant usability problems. An average of 50% of empirically determined usability problems were identified by the preceding heuristic evaluation. Some statements of heuristic violations were specific enough to precisely identify the actual usability problem that study participants encountered. Other violations were less specific, but still manifested themselves in usability problems and poor task outcomes. In this study, heuristic evaluation identified a significant portion of problems found during usability testing. While we make no assumptions about the generalizability of the results to other domains and software systems, heuristic evaluation may, under certain circumstances, be a useful tool to determine design problems early in the development cycle.

### 1. Introduction

Computer-based patient records have been shown to provide significant benefits to patient care and outcomes. However, poor user interface design is a barrier to using clinical information systems effectively. Many problems can be traced to weaknesses in usability and human-computer interactions (HCI) design [1,2,3]. Usability and HCI methods are considered important components of the system development process outside of healthcare. In medicine, several studies describe cognitive and HCI methods for evaluating and improving clinical systems [3,4,5]. Examples are cognitive task analysis, heuristic evaluation, cognitive walkthroughs and usability tests, which are used to provide insights to developers about potential usability problems. These methods can also be used for the summative evaluation of clinical systems. As part of developing a multimodal interface for dental computer-based patient records (CPR), the Center for Dental Informatics at the University of Pittsburgh conducted heuristic evaluation [6] and usability testing [7] of four commercial dental CPRs. Both heuristic evaluation and usability testing yielded strong evidence that the dental CPRs have significant usability problems.

Previous studies in other fields have suggested using a combination of different usability methods to identify design problems [8,9]. Several studies have shown that heuristic evaluation

can predict major usability problems that could potentially occur during usability tests [10, 11]. Jeffries et al. found that heuristic evaluation and usability testing performed better than cognitive walk-through and software guidelines in identifying usability problems and stressed the importance of choosing evaluators who are experienced in providing usability feedback to product groups [10]. Given this background, the objective of this study was to determine the extent to which heuristic evaluation and usability tests revealed the same types of usability problems in the four dental CPRs.

## 2. Methods

We conducted heuristic evaluation and usability evaluation methods on four major commercial dental CPRs during the period from January 2005 to July 2005. We briefly describe our application of the two methods below.

### 2.1. Heuristic evaluation

For the heuristic evaluation study, a set of ten heuristics published by Jakob Nielsen [12] was used to evaluate the four dental CPRs. Two dental informatics postgraduate students and one dental informatics faculty member evaluated each of the four dental CPRs. The systems were Dentrix Version 10.0.36.0 (Dentrix, American Fork, UT), EagleSoft Version 10.0 (Patterson Dental, St. Paul, MN), SoftDent Version 10.0.2 and PracticeWorks Version 5.0.2 (both Kodak Corp., Rochester, NY).

All evaluators were dentists with significant background in informatics and information systems. The faculty member was an expert in heuristic evaluation, while the postgraduate students had completed a course in human-computer interaction evaluation methods, including heuristic evaluation. All evaluators were familiar with the CPRs in general, but had no experience through routine use. Evaluators verbalized the heuristics that they considered violated while completing the tasks. An observer [TT] wrote down the violations and helped record illustrative screen shots when necessary (using a recorded macro function in MS Word [Microsoft, Redmond, WA]). While the evaluation was grounded in three clinical documentation tasks, evaluators were free to explore other clinical (not administrative) program functions in order to increase the coverage of the heuristic evaluation. For further details, please refer to the paper published previously [6].

### 2.2. Usability evaluation

We conducted usability assessments [4,9] on the charting interfaces of working demonstration versions of Dentrix Version 10.0.36.0, EagleSoft Version 10.0, SoftDent Version 10.0.2 and PracticeWorks Version 5.0.2 with four different groups of users consisting of five novice users in each group. Each participant used only one software package and worked through nine clinical documentation tasks using a think-aloud protocol [4,9,13]. The tasks were explained in detail in a previously published paper [7]. The purposive sample of novice users for each system consisted of one full-time faculty member, two practicing dentists and two senior dental students from the School of Dental Medicine (SDM) and the Pittsburgh area. After the completion of all sessions, two researchers coded usability problems based on an established coding scheme [9]. For each task, both the task outcome (rate of completed tasks, incomplete tasks and incorrectly completed tasks) as well as the type(s) of usability problems that occurred were coded.

### 2.3. Comparing heuristic evaluation and usability evaluation results

Heuristic evaluation results were reviewed to identify violations that led to usability problems during usability testing. The results were then summarized and described using descriptive statistics. The heuristic violations statements were classified into two groups: one group

consisting of specific violations that directly predicted actual usability problems, and the second consisting of general violations that suggested, but did not directly predict, observed usability problems.

### 3. Results

The number of usability problems identified through heuristic evaluation ranged from a low of 17 (39%) in PracticeWorks to a high of 61 (64%) in Dentrix (see Table 1). On average, heuristic evaluation predicted 50% of the usability problems found empirically. While in some cases, such as for EagleSoft and Dentrix, a significant majority of heuristic violations was specific enough to predict the actual usability problem, most heuristic violations found for PracticeWorks and SoftDent only suggested usability problems.

In Table 2 and Table 3, we illustrate specific (Table 2) and general (Table 3) heuristic violations. As is evident from the examples, specific heuristic violations identified design features, such as buttons and menu items, that could be directly tied to the failure or difficulty to complete a task. General heuristic violations, on the other hand, tended to highlight visual and functional designs that could have resulted in a number of usability problems.

### 4. Discussion

Our study demonstrated that, at least when applied to dental CPRs, heuristic evaluation identified a significant percentage of the usability problems found in an empirical study. This result is encouraging given the variability of results provided by different usability evaluation methods produced by other studies [5,8,9,11]. However, heuristic evaluation did not always predict usability problems in a very specific manner. Many of the heuristic violations reported by evaluators only suggested the potential of a range of different usability problems.

Usability issues identified by both methods often resulted in problems that were severe enough to cause users either to fail completing the task or to commit one or more errors in completing it. Previous research has suggested that using a combination of different usability methods is most useful to identify the majority of problems [4,8]. Problems identified by more than one method may indeed be more severe than those identified by a single method. However, support for this position is equivocal.

Unfortunately, our study produced no insights into which heuristic violations, a priori, were more likely to produce actual usability problems than others. While it would be highly desirable to be able to flag truly serious problems as early as possible in the development process, it is currently an open question on whether this is possible using heuristic evaluation. Future research should continue to investigate the relationship between findings of usability problems using different methods, and how the most significant problems can be identified as early as possible in the development cycle.

### Acknowledgements

The research in this manuscript was supported in part by National Library of Medicine award 5T15LM07059-17 and by grant 1 KL2 RR024154-02 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. Information on NCRR is available at "[www.ncrr.nih.gov/](http://www.ncrr.nih.gov/)".

## References

1. Ash J, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc* 2004;11(2):104–112. [PubMed: 14633936]
2. Elting LS, Martin CG, Cantor SB, Rubenstein EB. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ* 1999;318(7197):1527–1531. [PubMed: 10356010]
3. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. *Int J Med Inform* 2005;74(7–8):519–512. [PubMed: 16043081]
4. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform* 2004;37(1):56–76. [PubMed: 15016386]
5. Johnson CM, Johnson T, Zhang J. Increasing productivity and reducing errors through usability analysis: a case study and recommendations. *Proc AMIA Symp* 2000:394–398. [PubMed: 11079912]
6. Thyvalikakath TP, Schleyer TK, Monaco V. Heuristic evaluation of clinical functions in four practice management systems: A pilot study. *J Am Dent Assoc* 2007;138(2):209–218. [PubMed: 17272376]
7. Thyvalikakath TP, Monaco V, Thambuganipalle HB, Schleyer T. Usability evaluation of four commercial dental computer-based patient records. *J Am Dent Assoc* 2008;139(12):1632–1642. [PubMed: 19047669]
8. Law, L.; Hvannberg, ET. Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform. *ACM International Conference Proceedings; Vol 31 Proceedings of the second Nordic conference on Human-computer interaction; Aarhus, Denmark. 2002. p. 71-80. ISBN: 1-58113-616-1.*
9. John BE, Mashyna MM. Evaluating a multimedia authoring tool. *J Am Soc Inform Sci* 1997;48(11):1004–1022.
10. Jeffries, R.; Miller, JR.; Wharton, C.; Uyeda, K. User interface evaluation in the real world: a comparison of four techniques. *Conference on Human Factors in Computing Systems; Proceedings of the SIGCHI conference on Human Factors in computing systems: Reaching through technology; New Orleans, Louisiana. 1991. p. 119-124. ISBN: 1-89791-383-3.*
11. Tang Z, Johnson TR, Tindall RD, Zhang J. Applying heuristic evaluation to improve the usability of a telemedicine system. *Telemed J E Health* 2006;12(1):24–34. [PubMed: 16478410]
12. Nielsen, J.; Mack, RL. Executive Summary. In: Nielsen, J.; Mack, RL., editors. *Usability inspection methods*. Vol. 1st ed. New York, NY: John Wiley & Sons Inc; 1994. p. 1-24.
13. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inform* 2005;38(1):75–87. [PubMed: 15694887]

**Table 1**

The number of usability problems found through usability testing by system, and the number and percentage of usability problems predicted by heuristic evaluation (separated into specific and general categories)

| Number of usability problems ... | ... found through usability testing | ... predicted by heuristic violations |                      |  |
|----------------------------------|-------------------------------------|---------------------------------------|----------------------|--|
|                                  |                                     | Specific (percentage)                 | General (percentage) | Total (percentage of usability problems) |
| System                           |                                     |                                       |                      |  |
| EagleSoft                        | 60                                  | 20 (77%)                              | 6 (23%)              | <b>26 (43%)</b>                          |
| PracticeWorks                    | 44                                  | 0 (0%)                                | 17 (100%)            | <b>17 (39%)</b>                          |
| SoftDent                         | 86                                  | 5 (15%)                               | 34 (87%)             | <b>39 (45%)</b>                          |
| Dentrix                          | 96                                  | 41 (67%)                              | 20 (33%)             | <b>61 (64%)</b>                          |
| <b>Total</b>                     | <b>286</b>                          | <b>66 (46%)</b>                       | <b>77 (54%)</b>      | <b>143 (50%)</b>                         |

**Table 2**Sample “specific” heuristic violations which *directly predicted* a usability problem

| Specific heuristic violation  | Corresponding usability problem  |
|---|--|
| Button highlighting is the inverse of the customary design (button greyed out when selected) (Error prevention; EagleSoft).   | When asked to select roots of a single tooth for root canal treatment, users deselected the roots which they were supposed to select due to the non standard design.                       |
| Numbers 1–6 represent the six surfaces of the tooth that are normally identified by anatomical terms, e.g. bucco-mesial (Match between system and the real world; EagleSoft).                   | The onscreen numerical keypad to navigate tooth surfaces was mistaken as a means to enter dental pocket depths in mm.  |
| Switching between restorative and periodontal charts is difficult (Recognition rather than recall; EagleSoft).  | Several users experienced difficulty when switching from the restorative to the periodontal chart. They suggested providing mechanisms to perform this action more easily.                 |
| Trying to record caries on a tooth does not produce a result unless user clicks on one of several poorly labeled buttons (“Eo,” “Ex,” “Tx,” and “Comp”) (Visibility of system status; Dentrix). | Most users experienced difficulty completing tasks that used one of the buttons. The system provided neither feedback nor guidance.  |
| The tool tip for the button to enter root canal therapy (RCT) on a molar tooth indicates that the procedure applies to incisors (Error prevention; Dentrix).                                    | The tool tip misguided users, who, as a result, failed to locate the icon for molar tooth RCT.   |
| Deleting a finding on a tooth should only require selecting the finding and pressing the Delete key or similar action (Consistency and standards; Dentrix).                                     | Most users tried multiple times to delete an amalgam restoration by selecting the tooth and by pressing the delete button on the keyboard, an action that was not supported by any system. |

**Table 3**Sample “general” heuristic violations which *suggested* a usability problem

| General heuristic violation  | Corresponding usability problem   |
|--|---|
| The periodontal chart represents teeth by lines while the restorative chart depicts them more naturally (Consistency and standards; EagleSoft).  | Users were confused by the poor graphical presentation of teeth and surfaces on the periodontal chart which led many of them to record pocket depths incorrectly.   |
| Poorly designed periodontal chart with boxes and lines does not resemble teeth and results in visual clutter (Aesthetic and minimalist design; PracticeWorks).   | Users had difficulty identifying the buccal and lingual tooth surfaces which in turn led to failure in recording pocket depths for all tooth surfaces.  |
| The touch screen panel at the bottom of the periodontal chart is small, which makes it difficult to determine what findings are entered and where they are recorded (Recognition rather than recall; PracticeWorks). | Users experienced difficulty in recording bleeding gums because the icon to record bleeding was not easily recognizable   |
| The function of each icon is hard to recognize. Icons are small and the pictures on it do not convey specific meaning (Recognition rather than recall; SoftDent).  | Often users could not locate the specific icon needed to complete a task.   |
| The icons D, M, P, C in the periodontal chart are not helpful (Aesthetic and minimalist design; SoftDent)  | Users assumed that the icons D, M, P, and C helped in recording the different periodontal findings, but instead they indicated only a mode change. This led to users failing to record pocket depths and bleeding gums. |