Title
Detection and characterization of usability problems in structured data entry interfaces in dentistry.

Permalink

Journal

ISSN

Authors
Walji, Muhammad F
Kalenderian, Elsbeth
Tran, Duong
et al.

Publication Date

DOI

Peer reviewed

Detection and characterization of usability problems in structured data entry interfaces in dentistry

Muhammad F Walji, PhD[1,*], **Elsbeth Kalenderian, DDS, MPH**[2,*], **Duong Tran, DDS**[1], **Krishna K Kookal, MS**[1], **Vickie Nguyen, MS**[1], **Oluwabunmi Tokede, DDS, MPH**[2], **Joel M White, DDS, MS**[3], **Ram Vaderhobli, BDS, MS**[3], **Rachel Ramoni, DMD, SCD**[4], **Paul C Stark, ScD**[5], **Nicole S. Kimmes, DDS**[6], **Meta E. Schoonheim-Klein, DDS, PhD**[7], and **Vimla L Patel, PhD, DSc, FRSC**[8]

[1]The University of Texas Health Science Center at Houston

[2]Harvard School of Dental Medicine

[3]University of California, San Francisco

[4]Harvard Medical School

[5]Tufts University

[6]Creighton University

[7]Academic Centre for Dentistry at Amsterdam (ACTA)

[8]The New York Academy of Medicine

Abstract

Background—Poor usability is one of the major barriers for optimally using electronic health records (EHRs). Dentists are increasingly adopting EHRs, and are using structured data entry interfaces to enter data such that the data can be easily retrieved and exchanged. Until recently, dentists have lacked a standardized terminology to consistently represent oral health diagnoses.

Objectives—In this study we evaluated the usability of a widely used EHR interface that allow the entry of diagnostic terms, using multi-faceted methods to identify problems and work with the vendor to correct them using an iterative design method.

Corresponding author: Muhammad F Walji PhD, Associate Professor, The University of Texas Health Science Center at Houston, School of Dentistry, Department of Diagnostic and Biomedical Sciences, 7500 Cambridge Street, Room 4178, Houston, TX 77054, Tel: 713-486-4276, Muhammad.F.Walji@uth.tmc.edu.
*These authors contributed equally.

Statements on conflicts of interest

The author(s) declare that they have no competing interests.

Methods—Fieldwork was undertaken at two clinical sites, and dental students as subjects participated in user testing (n=32), interviews (n=36) and observations (n=24).

Results—User testing revealed that only 22–41% of users were able to successfully complete a simple task of entering one diagnosis, while no user was able to complete a more complex task. We identified and characterized 24 high-level usability problems reducing efficiency and causing user errors. Interface-related problems included unexpected approaches for displaying diagnosis, lack of visibility, and inconsistent use of UI widgets. Terminology related issues included missing and mis-categorized concepts. Work domain issues involved both absent and superfluous functions. In collaboration with the vendor, each usability problem was prioritized and a timeline set to resolve the concerns.

Discussion—Mixed methods evaluations identified a number of critical usability issues relating to the user interface, underlying terminology of the work domain. The usability challenges were found to prevent most users from successfully completing the tasks. Our further work we will determine if changes to the interface, terminology and work domain do result in improved usability.

Keywords

Electronic health record; dental; usability; interface; terminology; structured data entry; diagnosis

INTRODUCTION

Poor electronic health record (EHR) usability has been shown to reduce efficiency, decrease clinician satisfaction, and even compromise patient safety (1–5). Clinicians often face usability challenges when entering structured data in an EHR that use standardized terminologies. Structured data entry tasks are burdensome, as clinicians must successfully navigate and choose a pre-defined concept from a long list of possibilities(6). Although clinicians may enter data in narrative form in an EHR, structured data entry is critical for being able to derive many of the benefits of EHRs such as decision support, quality improvement, and reuse for research purposes (7, 8). Structured data entry is also a key requirement for certified EHR systems and for enabling clinicians to achieve meaningful use and associated incentive payments (9). Before interface solutions can be generated, there is a need to better detect the specific usability issues involved with structured data entry.

For structured data entry tasks, usability challenges involve both the underlying standardized terminology and the user interface(8). Rosenbaum et al define an interface terminology as a "systematic collection of health care-related phrases (terms) that supports clinicians' entry of patient-related information into computer programs"(8). Zhang and Walji have proposed the TURF unified framework for EHR usability where usability is defined as "how useful, usable, and satisfying a system is for the intended users to accomplish goals in the work domain by performing certain sequences of tasks" (10). TURF asserts that overall system usability is a result of both extrinsic difficulty and intrinsic complexity. Intrinsic complexity refers to the actual work conducted in a domain, and can be assessed by determining the usefulness of a system in supporting this work. Extrinsic complexity refers to the challenges users face when trying to accomplish specific tasks while using a user interface and can be assessed by determining the usableness of a system. Several techniques and methods exist to assess the usefulness, usableness and user satisfaction components of TURF. For example, in-depth qualitative techniques such as observations and interviews can provide rich contextual understanding of the work clinicians carry out in their day-to-day activities. User testing (11), expert inspection such as heuristic evaluation (12), and the development of cognitive models (13) are suited to determine the usableness of a system as they are primary concerned with identifying user interface related challenges as users try to accomplish their

tasks effectively and efficiently. Validated surveys can be used to determine user satisfaction (14). Usability challenges may best be detected using mixed method approaches rather than relying on any one technique(15).

Entering a diagnosis in a structured format is one example of a critical and frequently performed activity in an EHR. Until recently, dentistry has lacked the benefit of a standardized terminology to describe diagnoses. Previous efforts to standardize diagnostic terms in dentistry have included the Leake codes (16), and SNODENT developed by the American Dental Association (17). While the International Classification of Disease (ICD) incorporates some oral health concepts, they do not provide the necessary specificity to describe dental conditions (18). Attempts at standardizing diagnostic codes in the past (16, 19, 20) have not yet gained traction in part due to fragmentation of efforts to create coding systems as well as the absence of meaningful incentives (21). A standardized diagnostic terminology also plays an important role in stimulating widespread adoption of EHR systems (22). To fill this gap, a consortium of twenty-one dental schools has developed and implemented a standardized controlled terminology for dental diagnoses called EZcodes (23). This group of dental schools also uses a common EHR platform (24). In the past, these diagnoses would only be described in narrative form in a clinical note within the EHR, but now clinicians must formally use the standardized terminology and select a preferred concept. Users often continue to also describe the diagnoses in their clinical notes but are not obligated to do so. Early efforts to introduce these terms yielded lower than anticipated utilization rates and increased error rates. Analyses have shown that clinicians do not often enter terms and, when they do, they are often entered incorrectly (23). Although the usability of dental EHRs has been explored (1, 25–27), the use of a structured standardized diagnostic terminology has dramatically changed the way diagnoses are documented in these clinical settings and little is known about the impact from a user's perspective.

Therefore, the purpose of this research was to identify usability challenges of clinicians finding and entering diagnoses selected from a standardized diagnostic terminology in a dental EHR. We sought to analyze challenges associated with the (1) use of the terminology itself, (2) use of the EHR interface and (3) use of the terminology as part of clinic workflow. This study was conducted in collaboration with the EHR vendor (axiUm, Exan Corporation, Vancouver, BC) who has committed to iteratively improve the usability of the EHR as part of a 5-year grant funded project (1R01DE021051). This EHR is used by approximately 85% of all dental schools in North America, and improvements in the interface would impact a large number of users in the field.

METHODS

Sites

The usability assessments were conducted at two dental schools: Harvard School of Dental Medicine (HSDM) and University of California, San Francisco (UCSF). Both institutions have university owned clinics to train dental students as well as residents (post graduate students). Both dental schools also have a private faculty practice, use the same EHR system and have been early adopters of the EZcodes dental diagnostic terminology.

Participants

Participants of the study included a subset of third and fourth year dental students actively involved with delivering patient care and therefore one of the primary users of the EZcodes. Dental students are responsible for updating the medical record under the supervision of attending faculty. In addition, a subset of residents and faculty were also invited to

participate. Participants were recruited based on their availability during the dates of the site visits.

Procedures

Appropriate IRB approval was obtained at all institutions and all participants signed consents before entering the study. Usability evaluation researchers from The University of Texas Health Science Center at Houston (UTHealth), who were not known to the participants at each site, were responsible for conducting the assessments, and traveled to each site over a 3-day period. Subjects who agreed to participate in the study completed a short survey to document demographics and experience in using computers and the EHR. The TURF framework helped to inform the selection of methods. Observations and interviews were used to primarily assess intrinsic complexity and the corresponding usefulness of the EHR. User testing and cognitive modeling were primarily used to determine the extrinsic difficulty and associated usableness of the EHR. Observations were conducted in the dental clinics while user testing and interviews were conducted in conference rooms within the dental school building.

User testing with think-aloud—Participants were asked to think aloud (28) and verbalize their thoughts as they conducted two pre-defined scenarios. Two clinical scenarios and associated tasks were developed to assess the two approaches in which diagnoses could be entered into the system. The first scenario, consisting of two tasks (task 1a and task 1b), involved the user adding a diagnosis and an associated procedure using the "chart add" feature of the EHR. The "chart add" feature may be used to document a single diagnosis and associated treatment or a few uncomplicated diagnoses with an uncomplicated treatment plan existing of very few treatment options, i.e some cavities that need fillings or a cracked tooth that needs to be treated with a simple crown and some localized gum surgery or four partially impacted wisdom teeth that need to be removed. The second scenario involved using a more expansive treatment planning module in the EHR (task 2), which allowed a user to sequentially build a comprehensive treatment plan for a patient. Dentists at each of the two sites developed the clinical scenarios for user testing through a consensus-based approach. The scenarios outlined a patient's chief complaint, health history, and other pertinent information to complete the tasks. These scenarios were designed to accommodate the knowledge and skill level of third and fourth year dental students. The scenarios were tested to ensure the tasks could be performed through the EHR.

As part of user testing, we captured quantitative data to assess if tasks were completed successfully (a measure of effectiveness) and the amount of time spent in accomplishing the task (a measure of efficiency). Both the attempt time, and the task completion time were captured. After completing the tasks, participants were asked to provide additional feedback on the use of their system, and complete a user satisfaction survey using the validated and widely used System Usability Scale (29). The participants' computer screen and audio of the verbal comments were recorded using Morae Recorder Version 3.2 (Techsmith). In this study we aimed to recruit a minimum of 24 subjects (12 subjects from each site). Although there is some debate in the literature about sample size calculations for usability studies(30), our recruitment goals were consistent with recent recommendations for assessing EHR (31) and medical device usability (32).

In order to determine if users successfully completed the tasks, it was important to predefine the correct path to complete the tasks. Hierarchical Task Analysis (HTA) was used to model the tasks after gaining input from expert dentists at each site. HTA allowed the complex tasks to be defined in terms of a hierarchy of goals and sub goals with the aim of determining how users accomplish tasks within the EHR (33). After determining the

appropriate path to complete the tasks, the time was calculated that it would take an expert (who makes no errors) to complete the tasks (expert performance time). CogTool is an open source software tool that allows an evaluator to use screenshots of an application and a specified path to predict performance time for a specific task based on the keystroke level model (KLM). KLM is a type of GOMS technique (34) where a specified set of physical operators such as keystrokes (typing a diagnosis) and mouse operators (double clicking on a button) are specified. (35) In addition, the model can incorporate mental operators such as the process of "thinking" while trying to locate a specific item on an interface. Both the physical and mental operators have pre-defined task times (for example single clicking mouse button takes 0.2 seconds) that are used by CogTool to automatically generate a model to predict expert task performance (36). In prior work we have used KLM to predict EHR task performance time (13). These expert times derived from KLM provided a baseline in which to compare actual user performance from user testing.

Observations using ethnography—Ethnography, as a research method, is commonly used in sociology and anthropology to acquire detailed accounts of a particular environment, the people involved, and individuals' interactions within the environment. Observational data were collected over a three day period by a trained researcher in order to provide insight into the clinical workflow, information gathering and diagnostic decision-making process in the clinical environment where the dentists and dental students worked. To minimize any impact on patient care, a non-participatory observational technique was used. The researcher engaged with the participants only if there was a need for any clarification or during downtime such as when a patient did not show for an appointment. Observational data were captured using pen and paper. Each set of observations occurred for approximately four hours, in two separate shifts (morning and afternoon). The primary purpose of the observations was to capture overall clinical workflow and to identify how diagnoses were made and captured in the EHR using the EZcodes, and to identify any associated challenges. Actual clinical work was not part of the observation. We used a purposeful sampling technique to observe participants who would be engaged in diagnosis and treatment planning during the observation period. An initial sample size target of 10 participant observations was set based on prior experience in conducting similar observations. The research team had the ability to expand the sample if the observer felt that additional cases would provide new insight.

Semi-Structured Interviews—In order to capture subjects' experience of dental diagnostic terminology, workflow and interface, semi-structured interviews were used with open-ended questions. We used a semi-structured format to ensure uniformity of questions asked while allowing the interviewees to express themselves. These interviews were guided conversations with broad questions, which do not constrain the conversation, and new questions were allowed to arise as a result of the discussion. The prepared questions focused on two broad themes including (1) the perception and internal representation of the clinic, patient care and role of dentists/students within the clinic; and (2) the nature of the workflow and environment of care within this dental clinic with the use of EHR. The questions were influenced by the knowledge gained from the observations. Interviews were conducted with the members of the clinical team, and lasted thirty minutes each. The interview data were collected in order to assess information on the role, situational awareness and general work philosophy of the subjects in the dental clinic. The sample was representative of those who are usually present in the clinical environment and as such included dental third and fourth year students, residents and faculty.

Data Analysis Approach

Observational data that were collected were converted into a structured table with time stamps. Interview data were transcribed verbatim. User testing data were collected using Morae Recorder, and were imported into Morae Manager for further analysis. For qualitative data, a priori, the data analysis team agreed to focus on analyzing the data records based on three pre-defined themes: (1) dental diagnostic terminology, (2) EHR interface and (3) clinical workflow. A grounded theory approach was used to analyze the data (37). Written data from transcripts were conceptualized line by line and the key points were marked with a series of codes, usually a word or short phrase that represented the associated data segments. The codes were compared, renamed, modified and grouped into similar concepts. The researcher went back and forth while comparing data, constantly modifying, and sharpening the growing themes. Data were triangulated to determine common themes that emerged based on the different approaches for data collection. Data quality was assessed checking data transcription by randomly choosing three transcriptions and reading over the transcriptions while listening to the audiotapes.

Collaborative process for identifying, prioritizing and addressing usability problems—We used a participatory approach where study investigators and the EHR vendor met face to face during one day, and additionally over three, one-hour teleconferences to further identify and prioritize the usability problems that emerged from the study. The purpose of these sessions was to allow the group to identify specific problems, rank their criticality and importance for resolution, and to develop a timeframe and plan for addressing the concerns. The identified usability problems were classified first in terms of impact on a three-point scale (high, medium, and low). High impact items were those that could have a dramatic impact for improving usability, medium impact were those that would have a significantly positive impact and should be addressed but were not deemed urgent, and low impact issues were those that would have a minimal impact on user experience and could be addressed based on the availability of extra time and resources available. The usability problems were then assessed to provide a suggested timeframe for being addressed based on a three-point scale (short, medium and long term). Short-term issues were those that should be addressed immediately, medium-term addressed by the next major release of the EHR or within one-year. Longer-term issues were those that would be addressed within a 2-year timeframe. Based on the categorization, a three-phased timeline was developed to identify the usability issues that could be 1) addressed immediately by using existing features in the EHR, 2) addressed in the next release of the EHR (within 6 months), and 3) addressed in a future release of the EHR or within one year.

Results

Demographics

Thirty-four dental providers from UCSF and thirty-four dental providers from HSDM participated in the study (Table I). At UCSF, the majority of the participants were students. At HSDM, half of the participants were students and the remainder were residents or faculty. On average, providers practiced in three clinical sessions per week, and most were experienced in using computers and the EHR. More providers at UCSF reported that they often documented a diagnosis by using diagnostic terminologies except HSDM faculty (43%). In total 32 number of subjects participated in user testing, 36 were interviewed, and 24 were observed.

User Testing

Quantitative results from user testing provided an assessment of effectiveness (task success), efficiency (task attempt and completion times) and satisfaction of using the EHR for

completing diagnoses related tasks. As shown in Table II, users struggled to successfully complete the tasks. Task 1 had a higher success rate (22–41%) compared with task 2 (0%). No significant differences were found for task success between either school for task 1A (p=0.76) or task 1B (p=0.25). There was a large discrepancy between the optimal or best possible completion time calculated using CogTool and actual user completion times. For example for task 1A, users who successfully completed the task took over 3 times as long compared to the best possible time. The difference was less pronounced for completion of task 1B because users had learned how to complete the task from their previous attempt. Although no users successfully completed task 2, the attempt time was over 15 minutes compared with the possibility of users completing the task within about 4 minutes. With the exception of task 1B (p= 0.04), no significant differences were found between task completion times at either school.

The mean System Usability Scale (SUS) score measured on a scale from 0–100 was 56.9 (SD: 14.3). A higher SUS score indicates increased user's satisfaction with the system. No significant differences were found in user satisfaction between the 2 sites (p=0.09).

Identification of Usability Problems—Based on the findings from user testing, observations and interviews, 24 high level usability problems were identified (see Appendix 1). User testing detected most (83%) of the themes. While interviews and observation detected 63% and 38% of the total themes respectively. Twelve of the 24 themes are provided as examples in table III. Thirty-eight percent of the usability problems were associated with the user interface, 25% diagnostic terminology, and 38% with the work domain. Sixty-three percent of these problems were classified as high priority, 37% as medium priority and 0% as low priority. Further, based on group consensus between the usability evaluators, clinicians and vendor, a timeframe was developed to determine when the issue would be addressed.

User interface-related problems—The user interface posed several challenges for the dental providers to successfully find and select a diagnosis. Finding a diagnostic term could be accomplished using three approaches in the interface. First users could pick a term from a "quick list" that was an abbreviated version of the concepts. Users could find a term from the "full list" that displayed all the terms. These terms were listed in 13 categories. Categories could have one or more sub-category. Alternatively, users could use a search box to type a keyword, which would display a list of terms that matched syntactically. We found that users generally used the default option to find a concept. At one institution, the default was set to the quick list, and many users failed to even recognize the existence of the full list. When users did use the search feature, no assistance was provided to the user regarding misspelled keywords. Synonyms or alternative names for the same concept were also not searchable. For example users often tried to search for "missing tooth" instead of the preferred concept of "partial anodontia" or "partial edentulous". Users requested the use of synonyms, auto completion, and keyword suggestions to improve the search functionality.

One of the most frequently found problems discovered during user testing and participant interviews was the illogical *ordering of the terms* in the interface. The user interface displayed both the identifier/code, and the preferred term. The concepts were ordered numerically based on a meaningless identifier. Therefore the concepts were not displayed in any logical order from a user perspective. Furthermore, users had no ability to reorder the concepts based on the preferred term name. Users requested the ability to customize ordering by: (1) alphabetical order or (2) most common terms arranged on top or (3) general to specific or (4) frequency of use or (5) little to big (minor to major) or (6) vague/ ambiguous towards end of list (e.g. not otherwise specified- NOS).

The *lack of complete visibility of the category and concepts names* also posed serious problems for users (see example in Figure 1). Due to the limitations of the interface, the descriptions of the terms were truncated, or abbreviations were used to conform to the maximum allowable characters that could be displayed. Users often encountered a long list of terms. For example the sub-category "Caries" had 29 variations listed, which require the elevator bar to scroll through. They include 3 variations of primary caries, categorized by depth, three variations of incipient caries (white spots) categorized by depths and six variations of recurrent caries categorized by depth. The sub-category "gingival diseases - non plaque induced" has 33 terms classified by the various diseases, materials, trauma or habits that induce the disease.

User testing revealed a frequent use error in which participants lost a diagnosis that they had previously selected in the interface. Users would often single click a diagnosis so that the entry was highlighted, and then click "OK" to close the dialogue box, expecting the diagnosis to be saved. However, the system required a user to double click the term so it would then move to another display window on the same screen. This unexpected approach to selecting a diagnosis forced users to go back to the entry screen to retype the diagnosis. The inconsistent placement of action widgets (such as save), and the spatial relationships between action buttons and content being manipulated also caused user problems in entering and saving data. Naming of action buttons was also found to be confusing. For instance in order to add a diagnosis, a user would need to click a button labeled "Update". Some non-editable fields appeared editable, and users sometimes tried to select a category name as a diagnosis. Users were unable to verify if the correct diagnosis term had been saved in the system because only the code (number) of the diagnostic terms was shown on the completion step. This also caused problems for attending faculty who were responsible for verifying a correct diagnosis had been entered as the screen that required their approval only showed the number of the diagnostic term entered by the student and not the full description.

User interface related problems that were identified were generally associated with the lack of visibility in the interface, unexpected actions by the system, and generally a time consuming process to select and enter the diagnostic terms. It was determined that two of the 10 issues could be addressed immediately using existing functionality in the system, 2 issues could be addressed in the next software release (within 6 months), and the remaining required substantial changes and could be addressed within one year.

Diagnostic terminology-related problems—The usability analyses revealed that the granularity and specificity of the terms were not optimal for users to be able to clearly distinguish between the meaning of some of the categories and concepts. Label names of categories and subcategories did not have a clear meaning. Dental students who have limited experience and knowledge with dental diagnosis were often confused by distinct terms with similar names. Although the nuances of these concepts may be obvious to experienced clinicians, the provision of a definition would have helped the dental students ascertain their meanings. Abbreviations also caused confusion for users.

Users reported difficulty in finding a number of diagnostic terms such as: missing tooth, partial tooth loss, and generalized chronic gingivitis. Users also suggested that diagnostic terms for fracture, aesthetic concerns, para-functional habits should be expanded and be more specific. In addition, users also suggested improvements to the categorization of diagnosis terms. For example "gingivitis of generic origin - NOS", and "pericoronitis" were suggested to be re-classified into the sub-category "periodontics" instead of under the "pain/ altered sensation" sub-category.

Diagnostic terminology related problems were mainly associated with missing or mis-categorized concepts, and the need for more explanatory information such as definitions and synonyms. It was determined that most of the usability problems could be addressed either immediately or within 6 months.

Work domain and workflow related issues—In addition to usability problems associated with the interface and the terminology, some issues were due to missing functionality or incompatibilities with the activities and work conducted in the clinical domain. For example, the system only allowed the entry of one diagnosis to support a particular treatment. However, in dentistry multiple diagnoses are often used to support specific treatments. For example an endodontic treatment may require both a pulpal and peri-apical diagnosis. Pulpal diagnoses are those that directly affect the pulp tissue such as (a)symptomatic (ir)reversible pulpitis, pulp stones and internal resorption. Peri-apical diagnoses are concerned with the tissue surrounding the apex of the tooth and include acute or chronic apical abscess, fistula and external resorption. Users were also not able to distinguish between a differential or working diagnosis and a definitive diagnosis. This issue was seen as important, as a dentist often does not know the final diagnosis until after treatment has provided the opportunity for exploration and conformation of the initial diagnosis.

Dentists are not required to provide a diagnosis to justify a procedure for third-party reimbursement and therefore some users did not feel it was important or necessary to enter structured dental diagnoses. User testing also showed the propensity of users to either quickly select a miscellaneous or not otherwise specified (NOS) code, a very general diagnosis, or skip adding diagnosis altogether. For example one user commented: *"I don't want to go through all of it. I just choose one that is similar to [the]thing I think it is".* Dental students also reported that they didn't expect attending faculty to verify the correct entry for the diagnosis and observational data confirmed this. However, some users did comment on the importance of documenting a diagnosis for legal purposes: *"I have to be able to prove to you five years later why I did it. If you can't read my note then I can't prove to you why I did it. I am liable in the court of law. There must be adequate information in my descriptive note to sustain that I diagnosed and treated you. If it is not there and 5 years later the tooth fails and you get angry with me, I may lose because I didn't justify it."*

The work domain related usability challenges included missing functionality (e.g. need to add multiple diagnosis), functionality that should not have been included (e.g. ability to add free text diagnosis) and limitations in user's level of knowledge and training. A longer-term outlook was required to address most of the work domain related issues due to their complexity.

Discussion

Our mixed method usability evaluation of the structured data entry interface revealed challenges with the user interface, the underlying terminology, and how the system as a whole supported the work conducted by clinicians. Our approach was designed to identify issues related to usefulness and usableless of the structured data entry interface that needed improvement to allow users to efficiently and effectively enter diagnoses in the EHR. Therefore, although there were many instances of well-designed aspects of the application, our approach was not designed to detect these components. Overall more user interface and work domain related issues were detected than usability challenges associated with the underlying diagnostic terminology. Less than half of the users at both clinical sites were able to complete the first task successfully, despite having the name of the diagnosis clearly specified in the clinical scenario. No user who participated in usability testing successful

completed the task of finding 5 specific diagnoses while creating a treatment plan. Only a small number of users were able to enter three of the 5 diagnoses successfully. Although this task was developed to assess a more complex activity in which the participants had to infer a correct diagnosis, the results clearly demonstrated the challenges users faced in accurately documenting dental diagnoses which was further supported by the identification of 24 usability challenges from user testing, interviews and direct observations. These results also are consistent with our prior analysis which showed a lower than desired utilization rate for the diagnostic terminology, and when used, a high number of errors between diagnosis and procedure pairings(38).

Several user interface related challenges were discovered that were preventing users from successfully finding and selecting a diagnosis and as a result, degrading efficiency and causing errors. One of the most common problems experienced by users was finding an appropriate concept from a long list of items due to the illogical ordering of the terms in the display screen. The original diagnostic terminology was developed to include a meaningless identifier to confirm to best practices in terminology development (39, 40). However, in the user interface, ordering was displayed based on this numeric code. Before the inclusion of the diagnostic terminology in the interface, this was not a problem for the EHR, as treatment terminologies such as the Current Dental Terminology (CDT) use identifiers to denote ordering and categorization and dentists know many of the most often used treatment codes by their number. Showing the numeric code instead of the concept-preferred term in the screens that showed which diagnosis has actually been selected reinforced this now outdated custom. Although some clinicians may still remember numeric identifiers, as more and more terminologies make these identifiers irrelevant it is critical for the user interface to expose the term names rather than the numeric codes.

The lack of visibility of items on the interface also posed challenges for users and in some cased resulted in errors when concepts were mistakenly chosen. Users repeatedly struggled to ascertain the full name of a diagnosis or diagnosis category, as the term was either truncated, or had been abbreviated due the character limitations. Abbreviations are often misunderstood in healthcare, and have to be used with great caution (41). The EZcodes was developed as a pre-coordinated terminology and therefore the terms were often complex and lengthy. A terminology that allows post co-ordination may have mitigated some of the usability challenges of displaying and selecting these pre-coordinated items in the interface. Benefits of post-coordination also include greater expressivity and flexibility (42). However, composing a new concept from constituent terms also adds additional steps and cognitive burden for users, which can negatively impact usability (8, 43).

Another source of error was found when users attempted to link a diagnosis with a planned treatment. In order to accomplish this task users would need to navigate a tabbed user interface first to select a diagnosis, and then to select a treatment. However, once a diagnosis was selected, a user navigated to the next tab listing the treatments. This new tab failed to provide any indication of which diagnosis was actually selected in the previous screen. User testing showed the cognitive difficulty users had on retrieving from memory the specific diagnosis, and many errors were made during this phase (1, 25). This was especially prevalent for the second task where users had to document 5 diagnoses and related treatments. While users often corrected these errors after they were discovered in subsequent screens, there was a negative impact on performance. Other sources of errors were due to inappropriate naming of buttons and inconsistent placement of interface widgets, issues that have been documented in other studies also (1, 25, 44).

There were differing opinions from users relating to the comprehensiveness of the dental diagnostic terminology. Missing concepts, ambiguous preferred term names, and need for

improving the categorization scheme were identified as key areas for improvement of the dental diagnostic terminology. Although the terminology incorporated both synonyms and definitions to help distinguish and define concepts as recommended for contemporary clinical vocabularies (45), the user interface was not able to support them. As a standardized dental diagnostic terminology had not successfully been adopted before, the entire concept of formally recording a diagnosis was relatively new to both the attending faculty, residents and students who participated in the study. Therefore, although the EZcodes had been in use for over two years at each institution, we found a certain degree of ambivalence from dental students about the need to accurately document a dental diagnosis. Further, when faculty approved students' treatment plans, they could not verify the correct name of the chosen diagnostic term because only the numeric code was shown, resulting in either them ignoring checking for diagnosis-treatment match or spending more time finding the name of diagnostic terms which was a frustrating process. Therefore it was clear from this study that a concerted effort is needed to not only educate and gain buy-in, but also improve the workflow and interface in order to improve the usability of the dental diagnostic terminology for faculty and students.

Several work-domain related issues were also uncovered through our evaluation. These issues often related to a mismatch between the desired functions of users and the actual functions available in the interface. Many of the interface and terminology related usability problems were anticipated to be addressed in a shorter time frame than those that were related to the work domain.

Usability evaluations were conducted by evaluators who were completely independent of both the vendor of the EHR and of the clinical sites. However, collaboration with the vendor of the EHR system was essential for better understanding, prioritizing, and planning how the usability issues would be addressed in the future. Our collaborative approach led to the identification of several issues that could be immediately addressed by better customizing the system with features already available in the version that was tested. We then planned how to address the remaining issues either in the very next software release, or in future releases. We expect our relationship with the vendor to be a major asset in moving the results of this study into actions steps that will positively impact the usability of the diagnostic terminology and the system overall.

Although we used a mixed method approach to conduct usability evaluation, limitations exist that concern the generalizability of the study. Usability issues that were detected may have also been detected using alternative usability techniques. Inspection techniques, for instance, such as heuristic evaluation or cognitive walkthroughs are considered effective in detecting a broad range of interface related problems at a relatively low cost, especially when conducted early in a product development lifecycle (46). However, such inspection techniques are conducted by experts, and cannot determine the degree to which users can successfully complete tasks(a measure of effectiveness), or determine measures of efficiency. Our findings from the usability testing component of the research were also dependent on the clinical scenarios that were constructed for users. We would expect that findings relating to the user interface would have been similar regardless of the specific clinical tasks. We did however discover a learning effect for task 1B, where users performed better after learned how to complete the task from their previous attempt. Randomizing the order of task 1a and 1b would have minimized such an effect. We used a think-aloud protocol as part of user testing which likely provides an overestimate in terms of the time to complete tasks. However, comparison between the optimal time and observed time provides a useful metric for comparison purposes.

While we detected several important issues related to the diagnostic terminology itself, our findings were limited by the nature and content of the user testing scenarios, or what was actually observed during the site visits. Additional issues relating to comprehensiveness and coverage of the terminology may have been discovered with the inclusion of more clinical scenarios. Other more comprehensive evaluation techniques exist to determine concept coverage, term accuracy, and term expressivity (42). For example, concepts extracted from clinical documents can be assessed against a terminology (47, 48). In addition, clinicians can be asked to search concepts they would expect to be in a terminology, and allow them to determine an appropriate match (49). Such comprehensive terminology evaluations of the EZcodes will be conducted in future work. We also expect to evaluate the diagnostic terminology through clinician surveys, as well as through a feedback mechanism to a formal workgroup which is responsible for updating and maintaining the vocabulary. While three evaluators were primary responsible for conducting the initial analyses, each team member focused on primarily one methodology (usability testing, observations or interviews). Additional team members did not independently transcribe data or identify usability problems therefore we did not calculate inter-rater reliability. However, frequent team meetings were used to discuss methodological issues and findings.

Conclusion

Mixed methods usability evaluation allows for the identification of usability issues that are impeding the use of structured data entry interface. We discovered critical usability issues relating to the user interface, the underlying terminology, and to the work domain. An independent usability evaluation coupled with a close collaboration with the EHR vendor provided an opportunity to better understand, prioritize and provide a timeframe to address usability issues. In future work, we will reassess usability to determine the impact of diagnostic term entry post improvements in EHR interface and the diagnostic terminology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Thyvalikakath TP, Monaco V, Thambuganipalle HB, Schleyer T. A usability evaluation of four commercial dental computer-based patient record systems. J Am Dent Assoc. 2008; 139(12):1632–42. Epub 2008/12/03. [PubMed: 19047669]

2. 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):519–26. Epub 2005/07/27. [PubMed: 16043081]

3. Horsky J, Kaufman DR, Patel VL. The cognitive complexity of a provider order entry interface. AMIA Annu Symp Proc. 2003:294–8. Epub 2004/01/20. [PubMed: 14728181]

4. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. Journal of the American Medical Informatics Association: JAMIA. 2001; 8(4):299–308. [PubMed: 11418536]

5. Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. Journal of biomedical informatics. 2004; 37(3):193–204. Epub 2004/06/16. [PubMed: 15196483]

6. Khajouei R, Jaspers MW. CPOE system design aspects and their qualitative effect on usability. Stud Health Technol Inform. 2008; 136:309–14. Epub 2008/05/20. [PubMed: 18487749]

7. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. Journal of the American Medical Informatics Association: JAMIA. 1997; 4(3):213–21. Epub 1997/05/01. [PubMed: 9147340]

8. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006; 13(3):277–88. Epub 2006/02/28. [PubMed: 16501181]

9. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010; 363(6):501–4. Epub 2010/07/22. [PubMed: 20647183]

10. Zhang J, Walji MF. TURF: Toward a unified framework of EHR usability. J Biomed Inform. 2011 Epub 2011/08/27.

11. Khajouei R, Peek N, Wierenga PC, Kersten MJ, Jaspers MW. Effect of predefined order sets and usability problems on efficiency of computerized medication ordering. Int J Med Inform. 2010; 79(10):690–8. Epub 2010/09/14. [PubMed: 20833104]

12. Khajouei R, Peute LW, Hasman A, Jaspers MW. Classification and prioritization of usability problems using an augmented classification scheme. Journal of biomedical informatics. 2011 Epub 2011/07/26.

13. Saitwal H, Feng X, Walji M, Patel V, Zhang J. Assessing Performance of an Electronic Health Record (EHR) Using Cognitive Task Analysis. AMIA Proceedings. 2009:5.

14. Viitanen J, Hypponen H, Laaveri T, Vanska J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability. Int J Med Inform. 2011; 80(10):708–25. Epub 2011/07/26. [PubMed: 21784701]

15. Khajouei R, Hasman A, Jaspers MW. Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system. Int J Med Inform. 2011; 80(5): 341–50. Epub 2011/03/26. [PubMed: 21435943]

16. Leake JL, Main PA, Sabbah W. A system of diagnostic codes for dental health care. Journal of public health dentistry. 1999; 59(3):162–70. Epub 2000/01/29. [PubMed: 10649589]

17. Goldberg LJ, Ceusters W, Eisner J, Smith B. The Significance of SNODENT. Stud Health Technol Inform. 2005; 116:737–42. Epub 2005/09/15. [PubMed: 16160346]

18. Ettelbrick KL, Webb MD, Seale NS. Hospital charges for dental caries related emergency admissions. Pediatric dentistry. 2000; 22(1):21–5. Epub 2000/03/24. [PubMed: 10730282]

19. Gregg TA, Boyd DH. A computer software package to facilitate clinical audit of outpatient paediatric dentistry. International journal of paediatric dentistry/the British Paedodontic Society [and] the International Association of Dentistry for Children. 1996; 6(1):45–51. Epub 1996/03/01.

20. Orlwosky, R. Dental Diagnostic Codes. Winston-Salem; N.C: 1970. Report No

21. Kalenderian E, Ramoni RL, White JM, Schoonheim-Klein ME, Stark PC, Kimmes NS, et al. The development of a dental diagnostic terminology. J Dent Educ. 2011; 75(1):68–76. Epub 2011/01/06. [PubMed: 21205730]

22. Rector AL. Clinical terminology: why is it so hard? Methods of information in medicine. 1999; 38(4–5):239–52. Epub 2000/05/11. [PubMed: 10805008]

23. White JM, Kalenderian E, Stark PC, Ramoni RL, Vaderhobli R, Walji MF. Evaluating a dental diagnostic terminology in an electronic health record. J Dent Educ. 2011; 75(5):605–15. Epub 2011/05/07. [PubMed: 21546594]

24. Stark PC, Kalenderian E, White JM, Walji MF, Stewart DC, Kimmes N, et al. Consortium for oral health-related informatics: improving dental research, education, and treatment. J Dent Educ. 2010; 74(10):1051–65. Epub 2010/10/12. [PubMed: 20930236]

25. Hill HK, Stewart DC, Ash JS. Health Information Technology Systems profoundly impact users: a case study in a dental school. J Dent Educ. 2010; 74(4):434–45. Epub 2010/04/15. [PubMed: 20388817]

26. 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):209–10. 12–8. Epub 2007/02/03. [PubMed: 17272376]

27. Atkinson JC, Zeller GG, Shah C. Electronic patient records for dental school clinics: more than paperless systems. J Dent Educ. 2002; 66(5):634–42. Epub 2002/06/12. [PubMed: 12056768]

28. Ericsson, KA.; Simon, HA. Protocol analysis; Verbal reports as data. Cambridge, MA: Bradford books/MIT Press; 1993. Revised Edition

29. Brooke, J. SUS: a "quick and dirty" usability scale. In: JORDAN, PW.; THOMAS WEERDMEESTER, BA.; McCLELLAND, IL., editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996. p. 189-94.

30. Faulkner L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods. 2003; 35(3):379–83.

31. Lowry, SZ.; Quinn, MT.; Ramaiah, M.; Schumacher, RM.; Patterson, ES.; North, R., et al. Technical Evaluation, Testing, and Validation of the Usability of Electronic Health Record. National Institute of Standards and Technology; 2012. NISTIR 780

32. Wiklund, M.; Kendler, J.; Strochlic, A. Usability Testing of Medical Devices. CRC Press; 2011. p. 374

33. Diaper, D.; Stanton, N. The Handbook of Task Analysis for Human-Computer Interaction Mahwah. New Jersey: Lawrence Erlbaum Associates; 2004.

34. Card, SK.; Moran, TP.; Newell, A. The Psychology of Human-Computer Interaction. London: Lawrence Erbaum Associates; 1983.

35. John, BE. Information processing and skilled behavior. In: Carroll, JM.; Kaufman, Morgan, editors. Toward a multidisciplinary science of human computer interaction. 2003.

36. John, B.; Prevas, K.; Salvucci, D.; Koedinger, K., editors. Proceedings of CHI. Vienna, Austria: 2004. Predictive Human Performance Modeling Made Easy.

37. Strauss, A.; Corbin, J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage; 1990.

38. White JM, Kalenderian E, Stark PC, Ramoni RL, Vaderhobli R, Walji MF. Evaluating a dental diagnostic terminology in an electronic health record. J Dent Educ. 2011; 75(5):605–15. Epub 2011/05/07. [PubMed: 21546594]

39. Cimino JJ. In defense of the Desiderata. Journal of biomedical informatics. 2006; 39(3):299–306. Epub 2006/01/03. [PubMed: 16386470]

40. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine. 1998; 37(4–5):394–403. Epub 1998/12/29. [PubMed: 9865037]

41. Brunetti L, Santell JP, Hicks RW. The impact of abbreviations on patient safety. Jt Comm J Qual Patient Saf. 2007; 33(9):576–83. Epub 2007/10/06. [PubMed: 17915532]

42. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. A model for evaluating interface terminologies. J Am Med Inform Assoc. 2008; 15(1):65–76. Epub 2007/10/20. [PubMed: 17947616]

43. Bakhshi-Raiez F, de Keizer NF, Cornet R, Dorrepaal M, Dongelmans D, Jaspers MW. A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. Int J Med Inform. 2011 Epub 2011/10/28.

44. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res. 2010; 10:231. Epub 2010/08/10. [PubMed: 20691097]

45. Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp. 1997:640–4. Epub 1997/01/01. [PubMed: 9357704]

46. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform. 2009; 78(5):340–53. Epub 2008/12/03. [PubMed: 19046928]

47. Chiang MF, Casper DS, Cimino JJ, Starren J. Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies. Ophthalmology. 2005; 112(2): 175–83. Epub 2005/02/05. [PubMed: 15691548]

48. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. J Am Med Inform Assoc. 1996; 3(3):224–33. Epub 1996/05/01. [PubMed: 8723613]

49. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc. 1997; 4(6):484–500. Epub 1997/12/10. [PubMed: 9391936]

Summary Table

What was already known on the topic:
• Structured data entry interfaces in EHRs pose usability challenges to clinicians
• Poor EHR usability can negatively impact patient safety as well as provider efficiency and satisfaction

What this study added to our knowledge:
• The successful use of an dental interface terminology is dependent on the usability of the interface and how it supports the work of clinicians
• The structure and design of a pre-coordinated terminology has a direct impact on how it can be used within an existing EHR
• Poor interface and terminology usability has a detrimental impact on the ability of users to complete critical tasks

Research highlights

- A combination of usability methods can detect usability problems in structured data entry interfaces

- Usability challenges discovered relate to the user interface, terminology and work domain

- Usability problems severely impact clinical users ability to complete routine tasks
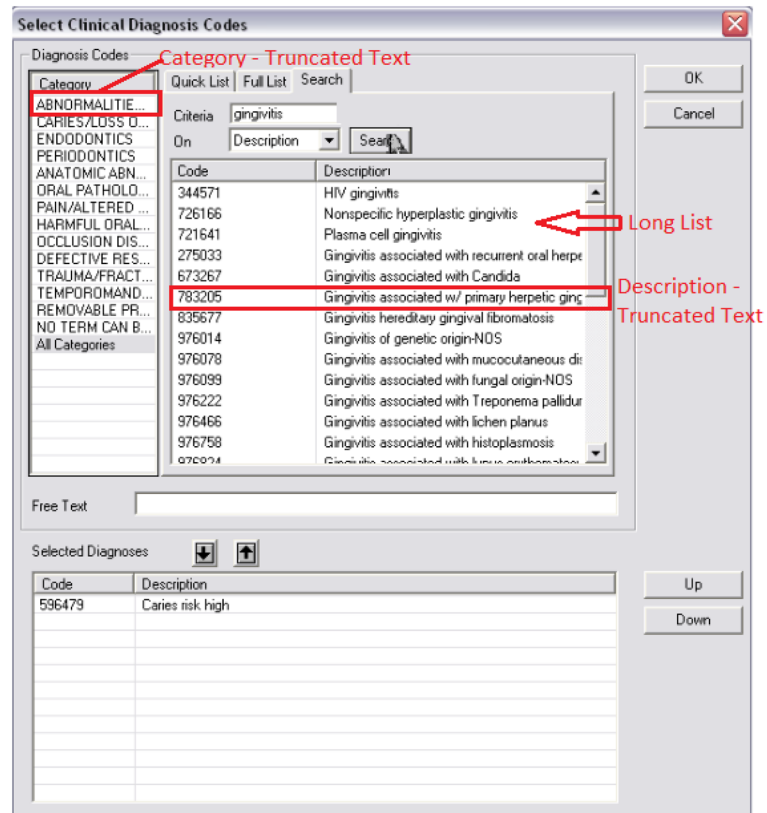
Figure 1.
Examples of limited visibility of the diagnostic term categories and concepts in the user interface

Table I

Demographics of Participating Dental Health Providers

Participants	UCSF (n=34)			HSDM (n=34)		
	Faculty (n=5)	Student (n=28)	Hygienist (n=1)	Faculty (n=7)	Student (n=17)	Resident (n=10)
Average clinic sessions/week	3.2	6.6	2	4.5	5.0	3.6
Average years of using axiUm	8.3	2.2	1	1.8	1.3	1.8
Experienced using computer (%)	100	89	0	86	94	80
Experienced with the use of the EHR (%)	100	86	0	57	87	50
Often documenting a diagnosis by using diagnostic terminologies (%)	100	93	0	43	94	100

Table II

Task Performance of users at two dental schools

	Task success (N=32)	Optimal time calculated using CogTool	Successful completion time				Attempted time (when task not completed successfully)			
			N	Mean	SD		N	Mean	SD	
Task 1A	41%	54 sec	13	3 min 8 sec	1 min 37 sec		29	2 min 50 sec	1 min 28 sec	
Task 1B	22%	50 sec	7	1 min 55 sec	1 min 14 sec		27	2 min 18 sec	1 min 50 sec	
Task 2	0%	4 min 39 s	n/a	n/a	n/a		32	15 min 36 sec	11 min 56 sec	

Table III

Select examples of usability problems identified, and associated priorities and timeframe to address and implement solutions

	Usability Problem(PRIORITY) Description/Example	Timeframe to Implement Solutions
Interface	**Illogical ordering of terms(HIGH)** Terms are ordered based on numeric code rather than alphabetically	**Immediate**: Reorder alphabetically; **< 1 year**: Users to customize ordering
	Time consuming to enter a diagnosis (HIGH) User must navigate several screens and scroll through a long list to find and select a diagnosis	**< 1 year**:
	Inconsistent naming and placement of user interface widgets (HIGH) To add a new diagnosis, a user must click a button labeled "Update"	**< 6 months**
	Search results do not match users expectations (MEDIUM) A search for "pericoronitis" retrieves 3 concepts with the same name but a different numerical code	**< 1 year**
Terminology	**Some concepts appear missing/not included (HIGH)** Examples of missing concepts according to users include: missing tooth, arrested caries, and attritional teeth	**< 6 months**
	Abbreviations not recognized by users (HIGH) Example: F/U, NOS, VDO	**< 6 months**
	Visibility of the numeric code for a diagnostic term (HIGH) Although the numeric code is a meaningless identifier, users had an expectation that the identifier should provide some meaning	**Immediate**: Use Quicklist to hide code **< 1 year**: Remove numeric code in UI
	Users not clear about the meaning of some concepts (MEDIUM) Novice users (students) had difficulty distinguishing between similar terms, and definitions and synonyms were not provided	**< 1 year**
Work	**Free text option can be used circumvent structured data entry (HIGH)** Instead of selecting a structured term, some users free text the name of the diagnosis.	**Immediate**: Disable option **< 1 year**: Remove option altogether
	Knowledge level of diagnostic term concepts and how to enter in EHR limited (HIGH) Users appear to have had little concerted education and training either by institution or vendor	**< 1 year**
	Only one diagnosis can be entered for each treatment (HIGH) Endodontic discipline require that treatments are justified using both a pulpal and periapical diagnosis	**< 1 year**
	No decision support to help suggest appropriate diagnoses, or alert if inappropriate ones are selected(MEDIUM)	**< 2 year**