

Text Prediction on Structured Data Entry in Healthcare

A Two-group Randomized Usability Study Measuring the Prediction Impact on User Performance

L. Hua^{1,2}; S. Wang³; Y. Gong¹

¹School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, USA; ²Informatics Institute, University of Missouri, Columbia, MO, USA; ³Department of Nursing, Tianjin First Central Hospital, Tianjin, China

Keywords

Structured data entry, text prediction, performance measurement, patient safety event reporting, usability

Summary

Background: Structured data entry pervades computerized patient safety event reporting systems and serves as a key component in collecting patient-related information in electronic health records. Clinicians would spend more time being with patients and arrive at a high probability of proper diagnosis and treatment, if data entry can be completed efficiently and effectively. Historically it has been proven text prediction holds potential for human performance regarding data entry in a variety of research areas.

Objective: This study aimed at examining a function of text prediction proposed for increasing efficiency and data quality in structured data entry.

Methods: We employed a two-group randomized design with fifty-two nurses in this usability study. Each participant was assigned the task of reporting patient falls by answering multiple choice questions either with or without the text prediction function. t-test statistics and linear regression model were applied to analyzing the results of the two groups.

Results: While both groups of participants exhibited a good capacity of accomplishing the assigned task, the results were an overall 13.0% time reduction and 3.9% increase of response accuracy for the group utilizing the prediction function.

Conclusion: As a primary attempt investigating the effectiveness of text prediction in healthcare, study findings validated the necessity of text prediction to structured date entry, and laid the ground for further research improving the effectiveness of text prediction in clinical settings.

Correspondence to:

Yang Gong MD, PhD
7000 Fannin St Ste 600
Houston, TX 77030
Email: Yang.Gong@uth.tmc.edu
Phone: 713-500-3547

Appl Clin Inform 2014; 5: 249–263

DOI: 10.4338/ACI-2013-11-RA-0095
received: November 13, 2013
accepted: 18. January 18, 2014
published: March 19, 2014

Citation: Hua L, Wang S, Gong Y. Text prediction on structured data entry in healthcare: A two-group randomized usability study measuring the prediction impact on user performance. Appl Clin Inf 2014; 5: 249–263 <http://dx.doi.org/10.4338/ACI-2013-11-RA-0095>

Introduction

Many attempts have been made to investigate the difficulties with data entry in order to promote the acceptance and quality-in-use of clinical information systems [1-3]. Structured data entry plays an indispensable role because of its merits of interoperability and reuse for research purpose. This is the rationale behind the initiative of a structured data capture project for the meaningful use of Electronic Health Records (EHR) [4] and the continued effort to develop and refine the standardized structured forms for patient safety event reporting [5]. As a process of selecting options from a predefined list, however, structured data entry is restrictive and inflexible compared to clinical report narratives, with respect to ambiguity tolerance and argument making. Consequently, a built-in field for narrative comments given as the last option of the predefined list becomes a common remedy. This remedy often comes along with the increase of physical and mental loads for text completion and may create a challenge for optimizing the overall performance of structured data entry. A solution proposed and examined in this study is the use of text prediction in the build-in narrative fields.

Text prediction, also known as word, sentence or context prediction originated in augmentative and alternative communication (AAC) to increase text generation rates for people with the disabilities of motor or speech impairment [6]. The advance of natural language processing techniques has brought text prediction into a broad scope of daily computing activities, such as mobile computing [7] and radiography reports [8]. However, the text prediction technique has two concerns when being applied in healthcare. First, there is a scarcity of research regarding the impact of text prediction on the quality of data entry that clinicians value. Second, despite text prediction having proven effective in reducing the motor requirement for text generation, whether this alone translates into an increased efficiency remains unclear [9-11]. In this study, a two-group randomized design was employed to examine the impact of text prediction on data entry quality and efficiency in a clinical setting, and to determine the effects of text prediction on clinician's overall performance in structured data entry.

Background

This study was grounded in a user-centered design for the development of a patient safety event reporting system. Such systems have shown the problems of underreporting [12] and low quality of reports [13, 14] for a decade although patient safety organizations at local and national levels have advocated the systems for years [15, 16]. The design aimed to increase the efficiency and the data quality by using text prediction in the system. Heuristic evaluation, cognitive task analysis and think-aloud user testing were conducted sequentially [17-19] to address interface representational issues. To further deepen the design at the functional level, text prediction functions on both structured and unstructured data entries were proposed to bridge the information gaps induced by work domain complexity and user disparity [20].

According to a preliminary study, the task of using multiple-choice questions (MCQs) to collect details of process-oriented events was the most time consuming and error-prone step in the course of safety event reporting, due to a great number of cognitive problems such as language ambiguities, mental model mismatches, etc. [19]. In this study, we developed a stand-alone prototype with the task to evaluate the impact of text prediction on the task-related structured data entries. The study can be instructive to the researchers who work on the optimization of prediction accuracy and interface representation to ascertain their efforts associated with text prediction were worthwhile.

Methods and Materials

Participants

Potential candidates who were nurses and experienced in reporting and analyzing patient safety events in the Tianjin First Central Hospital (TFCH) in Tianjin, China were identified and invited to participate in the study. Two candidates were on a leave of absence during the study period, and three candidates felt not confident with operating computers. As a result, the study enrolled 52 nurses from 21 clinical departments. All of the nurses were females and between 30 to 52 years old. On average, they had around 20 years of nursing experience and reported patient safety events for at least four years since the implementation of a citywide computerized reporting system in 2009. None of them used the interfaces for the study before. During the enrollment, each participant signed an informed consent form approved by the Ethics Committee at the TFCH. This study was also approved by the Institutional Review Board at the University of Texas Health Science Center at Houston.

Interfaces and task

Two experimental interfaces were developed as an easy control over the configurations and a means of data collection. The contents and layouts of two interfaces were identical, carrying the same task of the 13 MCQs for the collection of patient fall details [21]. One single exception was the provision of text prediction between the interfaces. The text prediction lists appear on four out of 13 MCQs (question 5, 6, 9 and 10) in the treatment interface primarily targeting the cognitive issues revealed in our preliminary study [19]. The interfaces were developed using PHP 5.2.6, JavaScript, MySQL 5.0.51b plus a JavaScript library (JQuery 1.7 [22]) and two open source modules (SlidesJS [23] and Tag-it [24]). ►Figure 1 as shown below, illustrates a typical page of the treatment interface utilized for answering the MCQs.

In the study, the text prediction lists were manually prepared case by case by domain experts (X.L. & Y.S.) as did similar studies [9, 25] and added to narrative fields for four of the 13 MCQs. Each list offered five possible choices of predictions (as shown in ►Figure 1, part C). The number of five is a trade-off between the efficiency and effectiveness of text prediction [26]. At least one of the five choices was considered accurate by expert agreement, and the other choices were less relevant ones. The participants were able to select a predicted text entry or type in unique comments for these four questions of their own free will. On the control interface, participants were only able to type to add their narrative comments using the keyboard.

Testing cases

In the study, every participant documented five patient fall cases in a randomized sequence. The cases were selected from two sources – a case depository with 346 fall reports from a previous study [14] and a public database of Morbidity and Mortality (M&M) [27]. Five selected cases were translated into Chinese and rephrased by the domain experts (X.L. & Y.S.) for the purpose of quality and readability of text. The difficulty of the five cases was managed at the same level. As an example, the following narrative excerpted from one of cases, shows here in English.

“... patient was alert and oriented X3 (person, time and location) upon assessment, and instructed on admit not to getting up without assist. He had been sleeping and attempted to get up to go to the bathroom. He forgot to call staff to have plexipulses (a device) undone, and tripped on plexi tubing and attempted to catch self on overhead bars. He landed on the floor...”

Experimental design

With a permuted-block algorithm [28], the 52 participants were randomly assigned to two groups. Twenty-five participants were allocated into the control group using the interface without text prediction; twenty-seven were assigned to the treatment group with text prediction. The presenting se-

quence of five cases for each participant was randomly determined at the time of allocation by the identical algorithm. The training combined verbal instruction and practice. Participants were trained and then practiced using both interfaces to document a sample case repetitively until all questions were accurately answered. Since the training was prior to grouping and the grouping procedure was blind to both the participants and the trainer, this arrangement prevented confounding implications delivered consciously or unconsciously by the trainer leading to a training bias.

Initiating a report upon witness' word-of-mouth information is one of typical scenes for safety event reporting in the hospital. This study simulated what commonly occurs by using the five cases with each appearing on the first page of the interface. Participants read the case descriptions and started the 13 MCQs upon recall. Pauses and pop-up questions were discouraged except when the participant switched between reports. Keystroke level operations (mouse clicks and keystrokes) for each participant trial were time stamped and logged into a MySQL database. All reporting sessions were done on a laptop with 800x600 resolution and recorded using Camtasia Studio[®] 7 for data reconciliation.

The processing of data

In this study, four dependent measures as shown in ►Table 1 were investigated to survey participant's performance and variation on the structured data entry. The calculations of the three measures in terms of time on question, prediction list active frequencies, and keystroke savings were conducted by using SQL queries upon logged and time stamped keystroke level operations.

The answers in the built-in narrative fields were reviewed and manually graded by the same experts (X.L. & Y.S.) measuring the response accuracy. Specifically, a single-response question n if correctly answered would result in an integer score $s_n = 1.0$, otherwise $s_n = 0$; a question n that accepts multiple responses could have an integer score $s_n = 4.0$ maximally in this study. Considering Q_n is the correct response for question n and q_n is the response given by participants, $Q_n \cap q_n$ indicates the degree of matching that is either a binary number for single-response questions or decimal for multiple-responses questions. The equations of calculating the response score S_n of an individual question and the overall response accuracy A_s across all questions for a report used in the study are shown as below.

$$S_n = (Q_n \cap q_n) s_n \quad (\text{Equation 1})$$

$$A_s = \frac{\sum_{n=1}^{13} S_n}{\sum_{n=1}^{13} s_n} \quad (\text{Equation 2})$$

To examine the significance of text prediction impact on the two primary measures of response accuracy and time on the questions, t-test was conducted using the group as the between-participants factor. Kernel density statistics were applied to examining the distributions of times on questions between the groups. Interactions between the measures and the experimental factors such as the level of difficulty of cases, the number of question options and the allowance of multiple responses were examined by linear regression models to identify factorial impacts on the overall performance and the effectiveness of text prediction lists. All statistical computing in the study was performed in R Studio v0.97.

Results

The participants successfully concluded the study with 260 reports comprising 2,849 questions, 3,194 question responses, which accounted for 3,999 mouse clicks and 3,868 keystrokes in total. Comparing the control and treatment group, although the means of participants' ages were 43.6 ± 5.8 versus 41.1 ± 6.6 ($p = 0.189$), the performances of them exhibited significant disparities on several levels. For instance, total mouse clicks were 1,669 versus 2,330 (39.6% increase) and keystrokes were 3,426 versus 442 (87.1% decrease) respectively. The detailed findings are shown in ►Table 2.

► Table 2 shows the results on two key measures of completion time and response accuracy. Completing a report of 13 questions on average took 131.0 ± 50.0 seconds in the control group and 114.0 ± 41.7 seconds in the treatment group. The overall response accuracies (A_s) were 79.4% and 83.2% respectively. According to the t-test results, both the differences were statistically significant ($p < 0.01$), while no significant difference between the groups on either efficiency or response score was found on the questions not associated with the text prediction function. As for the questions with the prediction lists, t-test results were significant on question 5 and 9, and insignificant on question 6 and 10. The active frequencies of prediction lists on these questions were 90.5% and 70.4% versus 32.8% and 44.0% respectively. On one hand, the four results support the text prediction largely increased participant's performance in efficiency and data quality; on the other hand, these effects might be mediated by the active frequency of prediction list.

► Figure 2 illustrates the distributions of times on three questions between groups, which presented three typical relationships between prediction lists and questions in the study. These relationships were: uninfluenced (question 1), influenced significantly (question 9), and influenced insignificantly (question 10). Regardless of the time differences between the groups, the text prediction list if used, showed a trend of bunching up values on the right side of the bell curve and a trend of narrowing the curve and tail as ► Figure 2 indicated on question 9 and 10. It means that the participant who spent much longer time on completing a report than the average were more likely from the control group than the treatment group. ► Figure 3 visually presents the mean differences between and within the groups in terms of time efficiency, response score and accuracy across the questions and cases. Two stacked lines are notably divergent at the questions where the prediction lists involved. In a completed report, the treatment group always reached higher response scores and shorter completion time than the control group. Within either of the groups, the performance variations across the questions and cases are large at the significant level ($p < 0.01$). This indicates the differences among cases and the MCQ features in terms of the number of options per question and the allowance of multiple responses had significant effects on participant's performance, as did the group factor. Therefore, the coefficients of these factors were further scrutinized by linear regression statistics. As a result, the coefficient of the group factor was significant ($p < 0.01$) which supports the effectiveness of text prediction despite the influences induced by the other factors in the experiment.

Discussion

Clinicians working under time constraints are usually expected to document data in a timely manner [29, 30]. The quality of entered data is critical to the decision-making and creation of actionable knowledge. This study attempted to promote efficient and accurate patient safety event reporting by introducing a narrative field supported by text prediction. A two-group randomized experiment was successfully developed and conducted to justify the impact of text prediction on data accuracy and time of completion of the structured data entry for patient safety event. As for a single patient fall report, the improvements in efficiency and data quality perspectives were small in absolute values and seemingly uncritical to care delivery. However, given the facts of millions of safety event reports generated each year [31, 32] and documentation demands in lethal situations such as medication issues, the text prediction could save practitioner's time, reduce cost and improve the quality of care in clinical settings.

Time efficiency, keystroke savings and response accuracy

Text prediction in the study has proved effective in increasing time efficiency on two questions, question 5 and 9 in the treatment group. As for the other two questions 6 and 10 with text prediction lists, the reason for lacking statistical significance remained unclear throughout the study. We believe that the low active frequencies of prediction lists and the large number of options per question somehow diminished the significance of impacts of the function, yet none of the conjectures were tested in the study.

The relationship between text prediction and time efficiency shows that the text prediction was most helpful in reducing the time expense when the reporting process was cumbersome and took

much longer time (e.g. over 30 seconds on question 9 and 40 seconds on question 10). A cumbersome situation could be defined as when a proper response was not in the predefined option lists or the participant failed to recognize the correct response due to cognitive issues. When the participant encountered few cumbersome issues and was able to respond rapidly (e.g. shorter than 10 seconds on question 9), the text prediction did not make the response even faster.

The analysis also implied that keystroke savings might play a vital role in increasing time efficiency in this type of data entry. A great portion of keystrokes, as high as 87.1% of total keystrokes, was reduced in the treatment group. This finding is consistent with the results of peer studies in a variety of fields [8, 33]. Nevertheless, whether keystroke savings alone could translate into increased efficiency remains unclear. There are mixed studies reporting contradicted results for the increased cognitive loads, eye gaze movements and mouse clicks [34-36]. The central value of investigating keystroke savings in this study is the savings that could be amplified for data entry with on-screen keyboards as more and more health information systems are migrating from desktop to mobile terminals. Usually, keystrokes with on-screen keyboards have a much greater time cost than those with regular computer keyboards.

In contrast to time efficiency, data quality has often been an ignored measure and underreported in text prediction research. This is partly because that measuring quality is not as straightforward as quantifying the numeric values for time efficiency. In addition, in the originated fields such as AAC and mobile computing, the data quality is much less of value than the time efficiency for daily normal activities, unworthy of the laborious manual analysis for the measurement. However, it is not the case in healthcare where the quality of data matters greatly.

There are multiple dimensions in measuring data quality [37] and one of the dimensions that we focused on is the accuracy of question responses. In this study, the response accuracy could be undermined in many ways, such as typographical errors, memory decay, casual attribution and hindsight biases [38]. Though no relations were systematically established by the study, somehow the text prediction offset these difficulties and resulted in significant improvements ($p<0.05$) on the response accuracy and two response scores as ►Table 2 and ►Figure 3 demonstrate. This evidently supported that text prediction would advantage the data quality in structured data entry, despite the drawbacks such as the over-reliance on predicted text might exist.

Limitations and work in progress

In the study, the prediction list was populated manually with at least one accurate answer among other four candidate answers. In reality, there are no matched techniques approaching such a high predictive accuracy. Moreover, the number of predicted answers could be various, more or less than five. Usually the longer the list is, the longer the time it would take for the participant's inspection and the greater the chance of missing correct responses. In this study, whether a longer text prediction list would have a lower accuracy of data entry was not investigated.

The active frequency of the prediction list seemed to be influenced by the number of options per question, the allowance of multiple responses or other features. Further studies should examine the relations between the active frequencies and question features.

Conclusion

Structured data entry, as an important format for documentation in healthcare, was demonstrated to be enhanced by a text prediction function in terms of time efficiency and data quality in a two-group randomized experiment. This groundbreaking study disclosed the necessity of developing and implementing the text prediction function even for experienced domain users in structured data entry.

Acknowledgments

The authors express heartfelt thanks to Xiao Liu, RN and Ying Sun, RN at the Office of Nursing at Tianjin First Central Hospital in China who made numerous efforts helping review and edit the testing cases, derived proper text predictor responses, scheduled the testing sessions, and graded the

participants' commentary data. The authors also thank anonymous reviewers for their constructive comments.

Clinical Relevance Statement

The study provides evidences supporting the usefulness of text prediction technology in healthcare documentation and justifying the importance of relevant research in terms of predicting algorithm and interface usability that are still lacking in the field.

Human Subject Research Approval

This study was approved by the Ethics Committee at the Tianjin First Central Hospital in Tianjin, China (No.E2012022K) and the Institutional Review Board at the University of Texas Health Science Center at Houston, Texas, the United States (No. HSC-SBMI-12-0767).

Conflicts of Interest

The authors of this paper certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Did the patient sustain a physical injury as a result of the fall?

CHECK ONE:

Yes No Unknown

A

What type of injury was sustained?

CHECK ONE; IF MORE THAN ONE, CHECK MOST SEVERE:

Dislocation Fracture Intracranial injury Other: PLEASE SPECIFY **B**

Laceration requiring sutures

C

D

May have the injury of

<input type="checkbox"/> Abrasion	46%
<input type="checkbox"/> Tears of skin	17%
<input type="checkbox"/> Laceration of skin	12%
<input checked="" type="checkbox"/> Deep incised wound	1%
<input type="checkbox"/> Open wound	0%

Reminding content or content
categorical entries for the
specification

Prediction list that reminds
of the content or content
categories of reportable data
(# of cued items <= 5)

Main component lists multiple-choice questions in slide-in mode

Fig. 1 The layout of the structured data entry interface with a text prediction list. The child question appears only when the corresponding item in its parent question is checked (A). The prediction list (C) is activated as the associated commentary field (B) is checked. Clicking the bottom button (D) would slide one page of new question(s) in, which helps identify transitions between questions in capturing time on questions in the study.

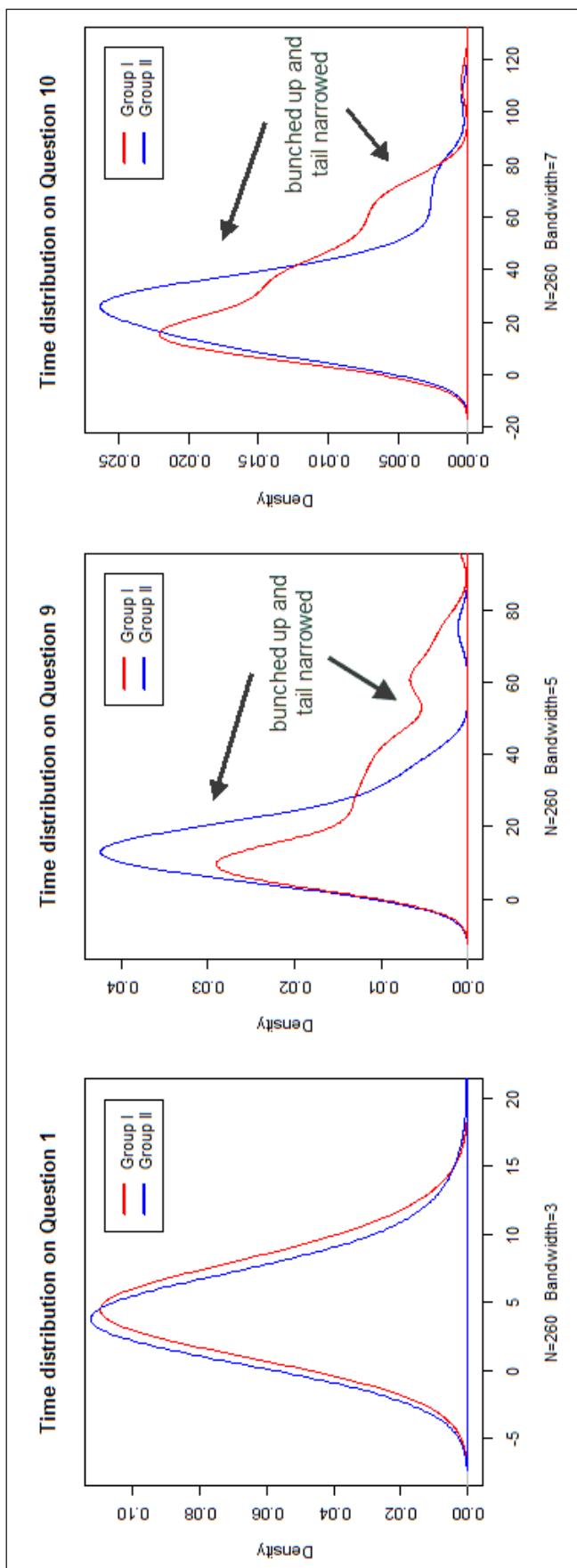


Fig. 2 Time distribution on question 1, 9 and 10 between control (I) and treatment (II) groups

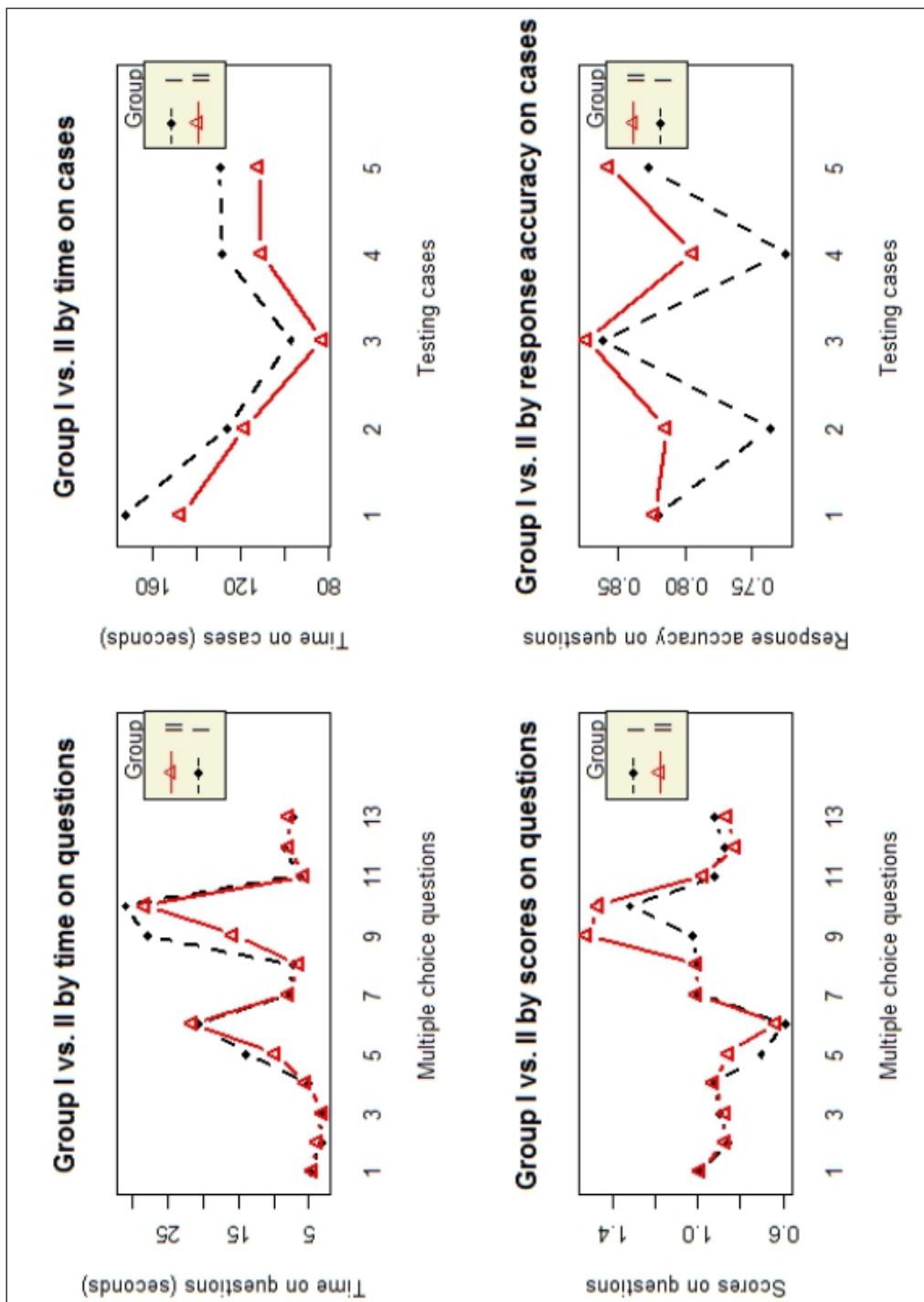


Fig. 3 Time and response accuracy on questions/cases between control (I) and treatment (II) groups

Table 1 Dependent measures with data sources, evaluation dimensions, and methods

Measures	Data Sources	Evaluating Dimensions	Methods
Response accuracy	Participant's responses on questions	Single score on question (S_n) and overall accuracy in percentage (A_s)	Expert review and descriptive statistics
Time on question	Logged operations with timestamps	Mean of time values at the millisecond level across reports	Descriptive statistics
Prediction list active frequencies	Logged mouse clicks associated with text prediction list	Denominator: the times of the question answered. Numerator: the times of the attached list activated.	Probability
Keystroke savings	Logged keystroke operations	Mean difference of the count of keystroke between groups	Descriptive statistics

Table 2 Participants' performance on MCQs between the control and treatment group

List of Questions (Appendix 1)	Options	Time (Seconds)			Score and Accuracy (%)		
		Ctrl. (N = 125)	Trt. (N = 135)	p-value	Ctrl. (N = 125)	Trt. (N = 135)	p-value
1. Assisted	3	4.9±2.2	4.5±2.9	0.235	0.99	0.99	0.563
2. Observed	3	3.2±2.9	3.6±2.9	0.299	0.86	0.88	0.714
3. Witness	2	3.22±.7	3.0±2.0	0.744	0.90	0.87	0.573
4. Injured	3	5.2±3.7	5.3±4.6	0.678	0.92	0.93	0.826
5. Sustained injuries* ● (Prediction list active frequency 90.5%)	5	14.1±8.7	9.9±7.1	0.000	0.70	0.84	0.015
6. Prior activity* ● (Prediction list active frequency 32.8%)	11	20.8±15.6	21.9±14.9	0.678	0.59	0.64	0.518
7. Risk assessment	3	7.7±5.3	7.7±5.0	0.849	1.00	1.00	N/A
8. At risk	3	7.4±4.2	6.5±4.3	0.305	1.00	1.00	N/A
9. Risk factors*§ ● (Prediction list active frequency 70.4%)	6	28.0±23.1	16.7±11.3	0.000	1.02	1.50	0.000
10. Preventive protocols*§ ● (Prediction list active frequency 44.0%)	16	31.2±20.8	28.7±17.6	0.234	1.31	1.48	0.139
11. Affected by medication	3	6.3±4.1	6.3±4.0	0.988	0.92	0.97	0.115
12. Risk increased by meds	3	8.5±6.8	7.6±5.6	0.644	0.86	0.81	0.560
13. Affected by physical device	3	7.2±6.6	7.8±5.4	0.416	0.92	0.87	0.155
Summary		131.0±50.0	114.0±41.7	0.004	79.4±10.1%	83.2±11.0%	0.005

*indicates the question with a commentary field; §indicates a multiple response question

References

1. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. *Journal of the American Medical Informatics Association* 1997; 4(3): 213–221.
2. Kaplan B. Reducing barriers to physician data entry for computer-based patient records. *Top Health Inf Manage* 1994; 15(1): 24–34. PubMed PMID: 10135720. Epub 1994/07/07.
3. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. *BMJ* 2004; 328(7449): 1184–1187.
4. Structured data capture initiative 2013. Available from: <http://wiki.siframework.org/Structured+Data+Capture+Initiative>.
5. AHRQ. Users' guide AHRQ Common formats for patient safety organizations : Agency for healthcare Research and Quality. 2008.
6. Beukelman D, Mirenda P. Augmentative and Alternative Communication. ISBN 1-55766-333-5: Brookes; 2005.
7. Mackenzie SI, Soukoreff WR. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction* 2002; 17(2 & 3): 147–198.
8. Eng J, Eisner JM. Informatics in radiology (infoRAD): Radiology report entry with automatic phrase completion driven by language modeling. *Radiographics* 2004; 24(5): 1493–1501.
9. Koester HH, Levine SP. Learning and performance of able-bodied individuals using scanning systems with and without word prediction. *Assistive technology: the official journal of RESNA* 1994; 6(1): 42–53. PubMed PMID: 10147209. Epub 1993/12/09.
10. Millar DC, Light JC, Schlosser RW. The Impact of Augmentative and Alternative Communication Intervention on the Speech Production of Individuals With Developmental Disabilities: A Research Review. *J Speech Lang Hear Res* 2006; 49(2): 248–264.
11. Gevarter C, O'Reilly MF, Rojeski L, Sammarco N, Lang R, Lancioni GE, et al. Comparisons of intervention components within augmentative and alternative communication systems for individuals with developmental disabilities: A review of the literature. *Research in Developmental Disabilities* 2013; 34(12): 4404–4414.
12. Kim J, Bates DW. Results of a survey on medical error reporting systems in Korean hospitals. *Int J Med Inform* 2006; 75(2): 148–155. PubMed PMID: 16095963. Epub 2005/08/13.
13. Gong Y. Data consistency in a voluntary medical incident reporting system. *J Med Syst* 2009; 35(4): 609–615. PubMed PMID: 20703528. Epub 2010/08/13.
14. Gong Y. Terminology in a voluntary medical incident reporting system: a human-centered perspective. *Proceedings of the 1st ACM International Health Informatics Symposium*; Arlington, Virginia, USA. 1882996: ACM; 2010. p. 2–7.
15. Rosenthal J, Takach M. 2007 guide to state adverse event reporting systems. National Academy for State Health Policy 2007.
16. Kohn LT, Corrigan, J.M., Donaldson, M.S. To err is human: building a safer health system. Report of Committee on Quality of Healthcare in America. Institute of Medicine, National Academy of Science. 1999.
17. Hua L, Gong Y. Developing a user-centered voluntary medical incident reporting system. *Studies in health technology and informatics* 2010; 160(Pt 1): 203–207.
18. Hua L, Gong Y. Usability evaluation of a voluntary patient safety reporting system: Understanding the difference between predicted and observed time values by retrospective think-aloud protocols. In: Kurosu M, editor. *Human-Computer Interaction Applications and Services*. Lecture Notes in Computer Science. 8005: Springer Berlin Heidelberg; 2013. p. 94–100.
19. Hua L, Gong Y. Design of a user-centered voluntary patient safety reporting system: Understanding the time and response variances by retrospective think-aloud protocols. *Stud Health Technol Inform* 2013; 192: 729–733. PubMed PMID: 23920653. Epub 2013/08/08.
20. Hua L, Gong Y. Information gaps in reporting patient falls: the challenges and technical solutions. *Stud Health Technol Inform* 2013; 194: 113–118. PubMed PMID: 23941941. Epub 2013/08/15.
21. AHRQ. Common Formats 2011. Available from: https://www.psoppc.org/web/patientsafety/version-1.2_documents
22. JQuery. Available from: <http://jquery.com/>.
23. SlidesJS. Available from: <http://www.slidesjs.com/>.
24. Ehlke A, Challand S, Schmidt T, Carneiro L. Tag-it. Available from: <http://aehlke.github.io/tag-it/>.
25. Higginbotham DJ, Bisantz AM, Sunm M, Adams K, Yik F. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication* 2009; 25(1): 19–31.

26. Hunnicutt S, Carlberger J. Improving word prediction using markov models and heuristic methods. *Augmentative & Alternative Communication* 2001; 17(4): 255–264.
27. AHRQ. Web M&M morbidity and mortality rounds on the web: Cases & commentaries [cited 2011 12/4]. Available from: <http://www.webmm.ahrq.gov/home.aspx>.
28. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials* 1988; 9(4): 327–344.
29. Allan J, Englebright J. Patient-centered documentation: an effective and efficient use of clinical information systems. *Journal of Nursing Administration* 2000; 30(2): 90–95.
30. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: A systematic review. *Journal of the American Medical Informatics Association* 2005; 12(5): 505–516.
31. Wachter R. Hospital Incident Reporting Systems: Time to Slay the Beast. San Francisco2009 [cited 2011 August 25th]. Available from: <http://community.the-hospitalist.org/2009/09/20/hospital-incident-reporting-systems-time-to-slay-the-monster/>.
32. James JT. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety* 2013; 9(3): 122–128.
33. Tuttle MS, Olson NE, Keck KD, Cole WG, Erlbaum MS, Sherertz DD, et al. Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises. *Methods of information in medicine* 1998; 37(4–5): 373–383.
34. Light J, Lindsay P, Siegel L, Parnes P. The effects of message encoding techniques on recall by literate adults using AAC systems. *Augmentative & Alternative Communication* 1990; 6(3): 184–201.
35. Koester HH, Levine S. Effect of a word prediction feature on user performance. *Augmentative & Alternative Communication* 1996; 12(3): 155–168.
36. Goodenough-Trepagnier C, Rosen M. Predictive assessment for communication aid prescription: Motor-determined maximum communication rate. *The vocally impaired: Clinical practice and research* 1988; 167–185.
37. Wand Y, Wang RY. Anchoring data quality dimensions in ontological foundations. *Commun ACM* 1996; 39(11): 86–95.
38. Holden RJ, Karsh BT. A review of medical error reporting system design considerations and a proposed cross-level systems research framework. *Human Factors* 2007; 49(2): 257–276. PubMed PMID: 17447667. Epub 2007/04/24.

Appendix 1 The MCQs used in the study

Page No.	Question and response options in detail
One	<p>1. Was the fall unassisted or assisted? CHECK ONE: a. Unassisted b. Assisted c. Unknown</p> <p>2. Was the fall observed? CHECK ONE: a. Yes b. No c. Unknown</p> <p>3. Who observed the fall? CHECK FIRST APPLICABLE: a. Staff b. Visitor, family, or another patient, but not staff</p>
Two	<p>4. Did the patient sustain a physical injury as a result of the fall? CHECK ONE: a. Yes b. No c. Unknown</p> <p>5. What type of injury was sustained? CHECK ONE; IF MORE THAN ONE, CHECK MOST SEVERE: a. Dislocation b. Fracture c. Intracranial injury d. Laceration requiring sutures e. Other: PLEASE SPECIFY _____</p>
Three	<p>6. Prior to the fall, what was the patient doing or trying to do? CHECK ONE: a. Ambulating without assistance and without an assistive device or medical equipment b. Ambulating with assistance and/or with an assistive device or medical equipment c. Changing position (e.g., in bed, chair) d. Dressing or undressing e. Navigating bedrails f. Reaching for an item g. Showering or bathing h. Toileting i. Transferring to or from bed, chair, wheelchair, etc. j. Undergoing a diagnostic or therapeutic procedure k. Unknown l. Other: PLEASE SPECIFY _____</p>

Appendix 1 Continued

Four	<p>7. Prior to the fall, was a fall risk assessment documented? CHECK ONE:</p> <ul style="list-style-type: none"> a. Yes b. No c. Unknown <p>8. Was the patient determined to be at increased risk for a fall? CHECK ONE:</p> <ul style="list-style-type: none"> a. Yes b. No c. Unknown <p>9. At the time of the fall, were any of the following risk factors present? CHECK ALL THAT APPLY:</p> <ul style="list-style-type: none"> a. History of previous fall b. Prosthesis or specialty/prescription shoe c. Sensory impairment (vision, hearing, balance, etc.) d. None e. Unknown f. Other: PLEASE SPECIFY _____
Five	<p>10. Which of the following were in place and being used to prevent falls for this patient? CHECK ALL THAT APPLY:</p> <ul style="list-style-type: none"> a. Assistive device (e.g., wheelchair, commode, cane, crutches, scooter, walker) b. Bed or chair alarm c. Bed in low position d. Call light/personal items within reach e. Change in medication (e.g., timing or dosing of current medication) f. Non-slip floor mats g. Hip and/or joint protectors h. Non-slip footwear i. Patient and family education j. Patient sitting close to the nurses' station k. Physical/occupational therapy, includes exercise or mobility program l. Sitter m. Supplemental environmental or area lighting (when usual facility lighting is considered insufficient) n. Toileting regimen o. Visible identification of patient as being at risk for fall (e.g., Falling Star) p. None q. Unknown r. Other: PLEASE SPECIFY _____
Six	<p>11. At time of the fall, was the patient on medication known to increase the risk of fall? CHECK ONE:</p> <ul style="list-style-type: none"> a. Yes b. No c. Unknown <p>12. Was the medication considered to have contributed to the fall? CHECK ONE:</p> <ul style="list-style-type: none"> a. Yes b. No c. Unknown <p>13. Did restraints, bedrails, or other physical device contribute to the fall (includes tripping over device electrical power cords)? CHECK ONE:</p> <ul style="list-style-type: none"> a. Yes b. No c. Unknown