Usability Evaluation of an EHR's Clinical Notes Interface from the Perspective of Attending and Resident Physicians: An Exploratory Study

Rubina F. Rizvi[a], Jenna L. Marquard[b], Marcus A. Seywerd[c], Terrence J. Adam[a, d], Jed T. Elison[e]
Gretchen M. Hultman[a], Kathleen A. Harder[a, f], Genevieve B. Melton[a, g]

[a] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[b] *Department of Industrial Engineering, University of Massachusetts, Amherst, MA, USA*
[c] *EPIC Innovations, Fairview Health Services, University of Minnesota, Minneapolis, MN, USA*
[d] *College of Pharmacy, University of Minnesota, Minneapolis, MN, USA*
[e] *Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, USA*
[f] *Center for Design in Health, College of Design, University of Minnesota, Minneapolis, MN, USA*
[g] *Department of Surgery, University of Minnesota, Minneapolis, MN, USA*

Abstract

Usability gaps between current and future improved Electronic Health Record (EHR) system designs exist due to insufficient incorporation of User-Centered Design (UCD) principles during System Development Life Cycle (SDLC). Usability of a commercial, inpatient EHR clinical notes documentation interface was analyzed from standpoints of two provider groups employing two standardized patient cases. Both objective and subjective data were collected from attending (n=6) and resident physicians (n=8) through usability testing employing a mixed method approach. The study results suggested that (i) EHR usability and desirability is influenced by user characteristics, (ii) workloads associated with H&P and progress notes writing are perceived differently between two groups, (iii) repeated task performance improves user efficiency and (iv) user performance is correlated to their subjective system assessments. Understanding usability of clinical documentation interface from perspectives of two different user groups, provides interface designers with an opportunity to develop an EHR system centered on UCD principles.

Keywords:

Electronic Health Record (EHR); User-Computer Interface; Documentation

Introduction

While Electronic Health Record (EHR) systems have been widely adopted with the ultimate goal of improved health care delivery [1], substantial gaps exist between the current state of EHRs and their potential usefulness [2]. Poor EHR usability appears to be a major factor for this discrepancy [2]. To facilitate optimal end product usability, it is critical to understand end users' "usage behavior", considered a core feature of a User-Centered Design (UCD) approach [3,4]. The UCD philosophy is that "the final product should suit the users, rather than making the users suit the product" [5]. According to the International Organization for Standardization (ISO)-framework used in this research study, usability is defined as the, "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [6]. Similarly, in EHR design, user involvement throughout the System Development Life Cycle (SDLC) can facilitate the development of systems that are easy to learn and remember, efficient, minimize errors and improve user satisfaction [7], which could improve EHR adoption and better patient outcomes [8].

Despite the critical role of the Human Computer Interaction (HCI) in the SDLC process [9], it is often neglected during EHR interface design. Usability studies on EHRs' clinical decision support system and user interfaces for medical equipment have been done in the past [10,11], but there are not many studies focusing on clinical notes documentation within an EHR interface [12-15], with only few studies done on usability evaluation and prototyping of clinical notes user interfaces in the medical domain [12-17]. Similarly, usability of a system could vary with vendor types and user profiles (e.g. clinical experience, EHR training, age, gender, technology skills etc.) However, few research studies incorporate usability comparisons from diverse user perspectives (e.g., expert users vs. novice users; physician vs. patients; users vs. usability experts) [18-20].

Usability testing is accepted as the most effective usability methodology with greatest strategic impact [21]. It is an "activity that focuses on observing users working with a product and performing tasks that are real and meaningful to them" [9]. The purpose of this study is to quantify EHR usability around inpatient notes usage focusing on the clinical note documentation and clinical note viewing interface, an area that poses tremendous challenges to physicians and other clinicians working under time limitations [13]. Both objective and subjective data on users' task performance were collected from two user groups (i.e., attendings and residents) and analyzed via usability metrics as defined by ISO (i.e., effectiveness, efficiency and satisfaction) [6]. Supplementary data were also analyzed for subjective workload using the NASA-TLX instrument [22] and system desirability with Product Reaction Cards (PRC) [23]. The insight gained through this research provides an opportunity to better understand EHR usability around clinical documentation from the standpoints of two provider groups and identify usability gaps to benchmark future EHR design.

Methods

This research study evaluated the usability of an enterprise EHR (Epic Systems Corporation) system at Fairview Health

Services, University of Minnesota Medical Center (UMMC). The study specifically focused on clinical documentation tasks (e.g., H&P and progress note writing). Scenario-based usability testing was conducted on two high fidelity simulated test patient charts [24] in an Epic test environment replicating the real work environment, both in design and functionality. Testing was done at the usability laboratory.

Study sample

Physician participants (n=14) were from two user groups: attendings (n=6) and residents, excluding interns (n=8). Participants were in all cases either trained in Internal medicine or Family medicine with past and/or current inpatient experience with the Epic Fairview EHR. Detailed user characteristics categorized by user group, are summarized in Table 1.

Table 1- users charactersistics

	Age (M/F)	Clinical Exp. (Yrs.)	Technology Exp.	Epic Exp. (Yrs.)	Epic proficiency
A	31 (F)	≤5	Somewhat	5-10	Average
A	43 (F)	> 10	Less	5-10	Average
A	xx (F)	≤5	Somewhat	5-10	Average
A	43 (F)	> 10	Very	> 10	Proficient
A	36 (M)	≤5	Somewhat	< 5	Proficient
A	39 (M)	≤ 5	Somewhat	5-10	Average
R	30 (F)	≤ 5	Somewhat	< 5	Average
R	xx (F)	≤ 5	Somewhat	< 5	Average
R	xx (M)	≤ 5	Somewhat	< 5	Average
R	30 (F)	≤ 5	Somewhat	< 5	Average
R	29 (M)	≤ 5	Somewhat	< 5	Proficient
R	26 (M)	≤ 5	Very	< 5	Proficient
R	29 (M)	≤ 5	Somewhat	< 5	Proficient
R	29 (F)	≤ 5	Somewhat	< 5	Average

A=Attendings; R=Residents; Clinical Exp.: Clinical Experience (Residency training and later); Epic Exp.: Total years of experince with using Epic

Participation was voluntary and participants received $50/hour. Each session was 2.5-3 hours long and each physician was at least 24 hours off night call on their day of data collection. The study protocol was approved by the Institutional Board Review.

Data collection and analysis

Two simulated, high fidelity test patient charts with rich, realistic clinical data were created in an Epic test environment to provide scenario-based EHR usability testing [24]. Patients were built from representative cases after extensive discussion among five experts: the lead EHR physician trainer (MS) and four physician informaticists (RR, TA, GMM & EA). Patient cases with similar complexities were selected using a Charlson weighted comorbidity index and number of prior admissions, clinic visits, and clinical notes. In both clinical scenarios, patients with a history of Chronic Obstructive Pulmonary Disease and Congestive Heart Failure presented in the emergency department with sudden onset shortness of breath. Each participant was assigned two patient cases in a random order employing an online randomization tool [25]. A Randomized blocked design approach was used to create balanced distribution of test patients across two groups. Each participant performed the same tasks of entering a H&P and a day 1 progress note, on each test patient's chart.

Raw data was extracted employing Tobii studio version 3.4.5 and was evaluated in three ways: (a) user satisfaction, via the System Usability Scale (SUS) questionnaire [26,27] (b) efficiency, via time on tasks, key presses, & mouse clicks and (c) effectiveness, via note quality using the Physician Documentation Quality Instrument-9 (PDQI-9) [28] and overall Gestalt judgment [29]. Data from each user group was also analyzed for subjective workload index using the NASA-TLX questionnaire [22] and system desirability via Product Reaction Cards (PRC) listing 118 words [23]. All participants were asked to circle their top 5 choices, which were later compiled as a word cloud and Venn diagram to visualize total and unique word selection by each user group.

Note quality assessment was performed by two co-authors/physicians (RR and TA) using standardized metrics as previously reported with the Physician Documentation Quality Instrument-9 (PDQI-9) [28] and overall Gestalt judgment [29]. Pretesting of these instruments for note quality assessment was conducted on a set of unrelated notes to ensure that both reviewers shared a common understanding of item scoring. Once consensus was achieved, both evaluators reviewed and assessed approximately 14% of notes (8 of 56 notes). The consistency in quality assessment was checked by inter-coder agreement with final mean agreement for PDQI-9 of 81% (kappa=0.69) and Gestalt scoring of 87.5% (kappa=0.71). We report summative statistics using SAS enterprise guide 5.1 and StatPlus LE 6.0.3 (a statistical software plugin for Macintosh) with means and standard deviation (sd).

Results

While not statistically different, user satisfaction with respect to overall usability of clinical note entry was perceived worse by attendings (mean SUS = 60.8 ± 15.6 (i.e., marginal usability) compared to residents (mean SUS = 73.4 ±13.5, (i.e., acceptable usability)), despite longer average Epic experience among attendings (≥ 5 years, n=5/6) compared to residents (< 5 years, n=8/8). The SUS and their interpretation [27] are illustrated in Fig 1.
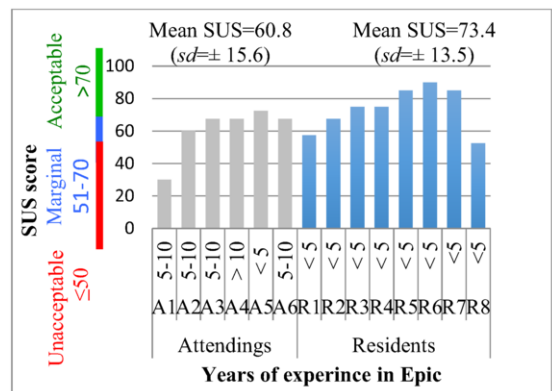

Figure 1-SUS based on users' characteristics

Efficiency was quantified based on time on task, key presses, and mouse clicks. H&P writing was more time-intensive than progress notes for both attendings (26.2 ± 9.7 vs. 14.0 ± 6.4 minutes) and residents (24.2 ± 7.7 vs. 12.3 ± 4.5 minutes). Residents took slightly less time than attendings writing both

H&P (24.2 ± 7.7 vs. 26.2 ± 9.7 minutes) and progress notes (12.3 ± 4.5 vs. 14.0 ± 6.4 minutes). Time on task decreased from the 1st to 2nd patient, except for progress note writing among residents (Fig 2).
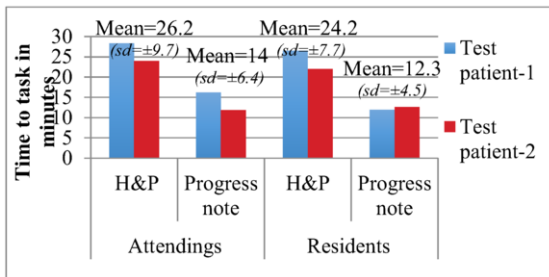

Figure 2-Time to tasks comparison between two user groups

More key presses (KP) and mouse clicks (MC) were observed with H&P as compared to progress note writing for both attendings (KP=2,644 ± 1535 vs. 1,433 ± 682, MC=201 ± 83 vs. 126 ± 60) and residents (KP=3,468 ± 1,199 vs. 1,758 ± 689 MC=214 ± 82 vs. 112 ± 46) with residents generally performing more key presses and mouse clicks compared to attendings with exception of progress notes where attendings had more mouse clicks. The number of key presses and mouse clicks decreased from the 1st to 2nd patient, except for residents' number of mouse clicks during progress note writing (Fig 3, 4).
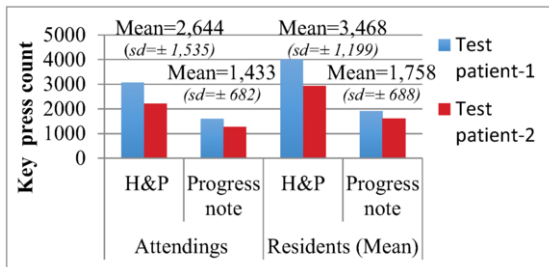

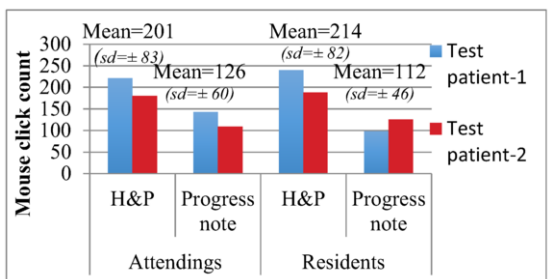*Figure 3-Number of key presses comparison between two user groups*

Figure 4-Number of mouse clicks comparison between two user groups

Effectiveness, as measured through PDQI-9 scores on note quality showed no quality differences between H&P and progress notes by attendings (34.9 ± 3.8 vs. 34.8 ± 4.8), though resident progress notes were slightly higher quality than H&P notes (35.5 ± 6.3 vs. 33.8 ± 4.0). Attendings' H&P notes (34.9 ± 3.8 vs. 33.8 ± 4.0), and residents' progress notes (35.5 ± 6.3 vs. 34.8 ± 4.8) showed only minimal quality differences. No noticeable differences in note quality between attending and residents were detected through Gestalt scoring both for H&P (3.7 ± 0.7 vs. 3.8 ± 0.8) and progress notes (3.9 ± 0.9 vs. 4.0 ± 1.0). PDQI-9 scores increased from the 1st to 2nd patient, except for residents' progress notes (Fig 5).
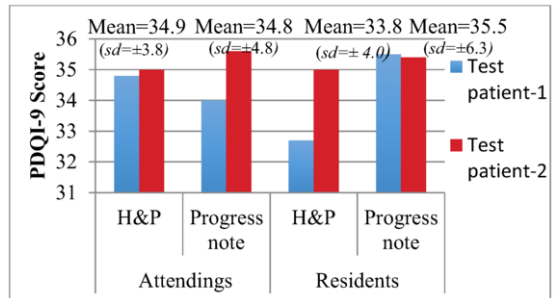

Figure 5-PDQI-9 scores comparison between two user groups

The NASA-RTLX questionnaire revealed that H&P note writing had higher overall workload (OW) than progress note writing among both attendings (27.8 ± 11.4 vs. 27.2 ± 16.0) and residents (33.6 ± 16.7 vs. 22.5 ± 10.2). Residents also had considerably higher subjective OW for H&P note writing (33.6 ± 16.7 vs. 27.8 ± 11.4), while attendings had higher subjective OW for progress note writing (27.2 ± 16.0 vs. 22.5 ± 10.2). There was no effect of patient order on perceived workload (Fig 6)
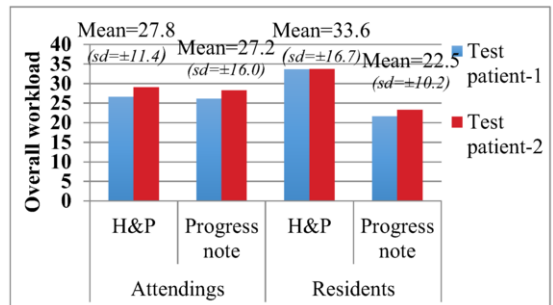

Figure 6-NASA-RTLX scores comparison between two user groups

Pearson correlation coefficient calculations were performed on the data after visually inspecting the distributions for normality. The results showed some correlation between metrics: NASA & SUS (-0.79 vs. -0.55 i.e., strong negative), NASA & Gestalt (-0.27 vs. -0.26 i.e., fair degree of negative), PDQI-9 & Gestalt (0.82 vs. 0.70 i.e., strong positive), Gestalt & SUS (0.39 vs. 0.30 i.e., fair degree of positive). Other metrics showed weak or no correlation: PDQI-9 & SUS (0.13 vs. 0.14), time on task & PDQI-9 (-0.18 vs. 0.10), and time on task & Gestalt (0.23 vs. 0.20).

System desirability analysis compared the proportion of positive vs. negative terms from a comprehensive list of 118 words (23). A higher percentage of positive as compared to negative words were selected both by attendings (63% vs. 37%) and residents (73% vs. 28%). Attendings selected a higher percentage of negative words (37% vs. 28%) while residents selected a higher percentage of positive words (73% vs. 63%) words respectively as depicted in the word cloud images (Fig 7). Similar results were seen for unique word selection as shown in the Venn diagram.
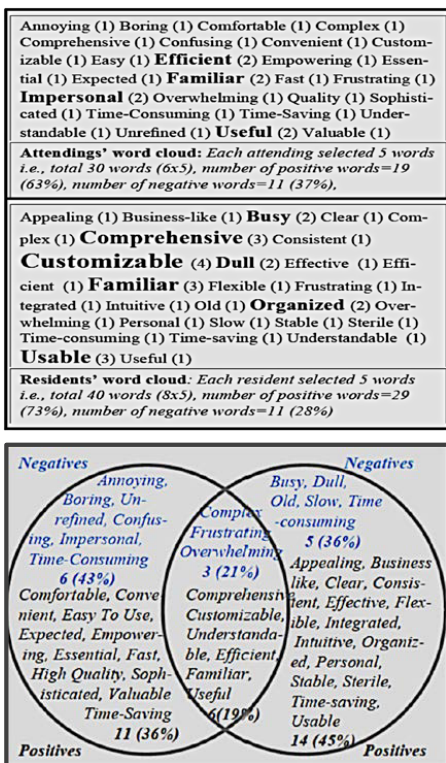
Annoying (1) Boring (1) Comfortable (1) Complex (1) Comprehensive (1) Confusing (1) Convenient (1) Customizable (1) Easy (1) **Efficient** (2) Empowering (1) Essential (1) Expected (1) **Familiar** (2) Fast (1) Frustrating (1) **Impersonal** (2) Overwhelming (1) Quality (1) Sophisticated (1) Time-Consuming (1) Time-Saving (1) Understandable (1) Unrefined (1) **Useful** (2) Valuable (1)

Attendings' word cloud: *Each attending selected 5 words i.e., total 30 words (6x5), number of positive words=19 (63%), number of negative words=11 (37%),*

Appealing (1) Business-like (1) **Busy** (2) Clear (1) Complex (1) **Comprehensive** (3) Consistent (1) **Customizable** (4) **Dull** (2) Effective (1) Efficient (1) **Familiar** (3) Flexible (1) Frustrating (1) Integrated (1) Intuitive (1) Old (1) **Organized** (2) Overwhelming (1) Personal (1) Slow (1) Stable (1) Sterile (1) Time-consuming (1) Time-saving (1) Understandable (1) **Usable** (3) Useful (1)

Residents' word cloud: *Each resident selected 5 words i.e., total 40 words (8x5), number of positive words=29 (73%), number of negative words=11 (28%)*

Figure 7-Product Reaction Cards- Attendings and residents

Discussion

This research study is an important initial step towards understanding the usability of EHR clinical notes documentation from attending and resident physician perspectives. EHR usability, as quantified through objective measures of user performance and their subjective perceptions, varied with each group, note type, and repeated tasks. Varying degrees of correlation were also discovered between variables, suggesting that user performance is related to their subjective system assessments. The insight gained through this research provides an opportunity to better understand EHR clinical documentation usability, identify and address existing usability gaps, and establish benchmarks for future EHRs.

Based on the SUS, residents perceived the system to have "acceptable usability" while attendings perceived the system to have "marginal usability", despite attendings having more experience with Epic. Similarly, system desirability was considered better among residents compared to attendings, with a higher percentage of positive words used to describe the system. Since residents are generally exposed to EHRs early in their medical training and tend to have little exposure to traditional paper charting, this may explain more favorable responses to EHR usability and desirability. Additionally, resident, predominantly males, tended to be younger, and rated themselves as having more technical experience, leading to easier technology adoption. Thus, user characteristics appear to be a critical factor for EHR usability.

In terms of efficiency, as quantified by time on task, key presses and mouse clicks, attendings and residents both took significantly more effort with H&P compared to progress note writing. Residents perceived less subjective workload associated with progress notes suggesting that residents were more at

ease in writing progress notes. A potential reason for this is the nature of progress note writing task itself, which is more repetitive and most likely to be influenced by a system's usability (e.g., copying and pasting, auto population, multiple screen panel functionalities, etc.) In comparison, attendings showed less subjective workload with H&P writing suggesting that they are better skilled in writing H&P notes, a cognitively demanding task which involves providing a reason for admission and providing initial patient management direction. Thus, targeted note documentation training of physicians where there is a lack of proficiency (e.g., H&P among residents and progress note in attendings), would be a reasonable approach to consider. No noticeable difference in note quality between attending and residents was detected through Gestalt scoring.

Generally, efficiency improved as users performed the same note writing tasks on the 2nd patient with the exception of progress note writing among residents. The plausible explanation of the observed differences may be due to user familiarity with the system and faster cognitive processing as a result of repeated task performance, as well as specifics around the second patient case. No effect of patient order was observed on perceived workload while there was some indication of improvement in note quality, especially progress notes, among attendings and H&P writing among residents.

We discovered that increases in subjective workload (NASA) were associated with decreases in user satisfaction (SUS) and note quality (per Gestalt). Higher satisfaction was associated with better quality notes (per Gestalt). We found a strong positive correlation between PDQI-9 & Gestalt, but no correlation was detected between PDQI-9 & SUS or with time on task and note quality for both PDQI-9 & Gestalt.

There are some limitations associated with this study, including a small sample size lacking significant inferential statistical results. Generalizability is limited due to the inclusion of physicians (MDs) with training in either Internal medicine or Family medicine and testing of inpatient EHR interfaces only. Future studies are mandated with larger sample size, more diverse group of participants (e.g., medical students, nurses, ER physicians etc.) and on different EHR systems (e.g., EHRs used in ambulatory settings, emergency rooms and as well as specialty specific EHRs). Additionally, the impact of other user characteristics (e.g., age, gender, etc.) needs to be explored further. There are also some limitations associated with usability testing itself (e.g., relevance of scenarios/ tasks being tested, system speed and connectivity etc.). In addition, understanding physicians' EHR usage behaviors around clinical note documentation, the focus of our next study, is an essential area that needs to be further explored.

Conclusions

We discovered that EHR usability measures of satisfaction, efficiency, and effectiveness vary with users' characteristics, specific note types, and from repeated performance of the same task on consecutive patients' charts. This study provides preliminary, yet essential information on objective measures of user performance and their perceptions of EHR usability around clinical notes usage. These measures can serve as initial guidance to build EHR interfaces grounded on a "User-Centered Design" approach.

Acknowledgements

References

[1] Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health Aff (Millwood) 2005 Sep-Oct;24(5):1103-1117.

[2] Zhang J, Walji MF. TURF: Toward a unified framework of EHR usability. J Biomed Inform 2011 12;44(6):1056-1067.

[3] Norman DA, Draper SW. User centered system design; new perspectives on human-computer interaction: L. Erlbaum Associates Inc.; 1986.

[4] Zheng K, Padman R, Johnson M, Diamond H. An interface-driven analysis of user interactions with an electronic health records system. J Am Med Inform Assoc 2009 Mar-Apr;16 (2):228-237.

[5] Baxter K. Understanding your users: a practical guide to user research methods. Second edition. Waltham, MA : Morgan Kaufmann; 2015.

[6] Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Software Quality Journal 2003;11(4):325-338.

[7] Nielsen J. Usability engineering. : Elsevier; 1994.

[8] Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005 2;38(1):75-87.

[9] Barnum CM, Barnum CM. Usability testing essentials ready, set-- test! Burlington, Mas. : Morgan Kaufmann; 2010.

[10] Carroll C, Marsden P, Soden P, Naylor E, New J, Dornan T. Involving users in the design and usability evaluation of a clinical decision support system. Comput Methods Programs Biomed 2002;69(2):123-135.

[11] Liu Y, Osvalder A. Usability evaluation of a GUI prototype for a ventilator machine. J Clin Monit Comput 2004;18(5-6):365-372.

[12] Sharda P, Das AK, Cohen TA, Patel V. Customizing clinical narratives for the electronic medical record interface using cognitive methods. Int J Med Inf 2006;75(5):346-368.

[13] Rizvi RF, Harder KA, Hultman GM, Adam TJ, Kim M, Pakhomov SV, et al. A comparative observational study of inpatient clinical note-entry and reading/retrieval styles adopted by physicians. Int J Med Inf 2016;90:1-11.

[14] Brown P, Marquard J, Amster B, Romoser M, Friderici J, Goff S, et al. What Do Physicians Read (and Ignore) in Electronic Progress Notes? Appl Clin Inform 2014;5(2):430-444.

[15] Using High-Fidelity Simulation and Eye Tracking to Characterize EHR Workflow Patterns among Hospital Physicians. AMIA Annual Symposium Proceedings: 2015.

[16] 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 May-Jun;13(3):277-288.

[17] Brown SH, Hardenbrook S, Herrick L, St Onge J, Bailey K, Elkin PL. Usability evaluation of the progress note construction set. Proc AMIA Symp 2001:76-80.

[18] Clarke MA, Belden JL, Kim MS. Determining differences in user performance between expert and novice primary care doctors when using an electronic health record (EHR). J Eval Clin Pract 2014;20(6):1153-1161.

[19] Ozok AA, Wu H, Garrido M, Pronovost PJ, Gurses AP. Usability and perceived usefulness of personal health records for preventive health care: A case study focusing on patients' and primary care providers' perspectives. Appl Ergon 2014;45(3):613-628.

[20] Corrao NJ, Robinson AG, Swiernik MA, Naeim A. Importance of testing for usability when selecting and implementing an electronic health or medical record system. J Oncol Pract 2010 May;6(3):120-124.

[21] A toolkit for strategic usability: results from workshops, panels, and surveys. Proceedings of the SIGCHI conference on Human Factors in Computing Systems: ACM; 2000.

[22] NASA-task load index [NASA-TLX]; 20 years later. Proceedings of the human factors and ergonomics society annual meeting: Sage Publications; 2006.

[23] Benedek J, Miner T. Measuring Desirability: New methods for evaluating desirability in a usability lab setting. Proceedings of Usability Professionals Association 2002;2003:8-12.

[24] Leveraging a Proprietary Electronic Health Record System to Create High-Fidelity, Smart Test Patients. iHealth 2016 Clinical Informatics Conference; 2016; ; 2016.

[25] Urbaniak GC, Plous S. Research Randomizer (Version 4.0) [Computer software].

[26] Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry 1996;189(194):4-7.

[27] Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies 2009;4(3):114-123.

[28] Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). Appl Clin Inform 2012;3(2):164-174.

[29] Penaloza A, Verschuren F, Meyer G, Quentin-G S, Soulie C, Thys F, et al. Comparison of the unstructured clinician gestalt, the wells score, and the revised Geneva score to estimate pretest probability for suspected pulmonary embolism. Ann Emerg Med 2013;62(2):117-124.

Address for correspondence

Genevieve B. Melton, MD, PhD
Institute for Health Informatics and Department of Surgery
420 Delaware Street SE, Mayo Mail Code 450
Minneapolis, MN 55455
Tel: + 612-625-7992, E-mail: gmelton@umn.edu