Developing Analytical Inspection Criteria for Health IT Personnel with Minimum Training in Cognitive Ergonomics: A Practical Solution to EHR Improving EHR Usability

Zhen Zhang, MD, MPH, MSc, Amy Franklin, PhD, Muhammad Walji, PhD, Jiajie Zhang, PhD, Yang Gong, MD, PhD
The University of Texas School of Biomedical Informatics, Houston, TX
National Center for Cognitive Informatics and Decision Making in Healthcare

Abstract

EHR usability has been identified as a major barrier to care quality optimization. One major challenge of improving EHR usability is the lack of systematic training in usability or cognitive ergonomics for EHR designers/developers in the vendor community and EHR analysts making significant configurations in healthcare organizations. A practical solution is to provide usability inspection tools that can be easily operationalized by EHR analysts. This project is aimed at developing a set of usability tools with demonstrated validity and reliability. We present a preliminary study of a metric for cognitive transparency and an exploratory experiment testing its validity in predicting the effectiveness of action-effect mapping. Despite the pilot nature of both, we found high sensitivity and specificity of the metric and higher response accuracy within a shorter time for users to determine action-effect mappings in transparent user interface controls. We plan to expand the sample size in our empirical study.

Introduction

Electronic health record systems (EHRs) are the most critical form of IT penetration in the healthcare system. Being increasingly integrated across every aspect in healthcare delivery, EHRs are expected to be the powerful means to optimize quality of care. However, the National Research Council reported that current health IT applications provide little support for clinicians' cognitive tasks, increase the chance of error, and add to rather than reduce workload due to underutilization of human-computer interaction (HCI) principles [1]. Although usability, or human factors, has been widely recognized in the medical device industry since 1988 [2], it is currently a big challenge to improve the usability of complex EHRs.

Several differences between the medical device industry and the EHR industry shed light on practical solutions to the EHR usability challenge. First, unlike the former with human factors design process addressed in the American National Standard [3], AMIA Usability Task Force identified barriers in the EHR industry to adopt user-centered design process, such as different user types in one organization and significant process variations across different organizations [4].

Second, designing and evaluating EHRs become more complicated in organizations that are developing safer and more efficient processes. The uncertainty as to the fitness of EHRs for complex socio-technical systems coincides with the uncertainty of radical transformation in the organizations. A viewpoint from research scholars beyond the healthcare arena attempts to tackle such technology-and-institution co-evolving phenomena by engaging two fields --- IT research and organization studies --- in conversation [5].

Third, unlike ready-to-use medical devices delivered to the market, EHRs often require a significant amount of configuration made by local health IT personnel, especially for large healthcare systems. The functioning EHR that clinicians interact with is in fact the joint effort by the EHR vendor and the local IT department. Different from medical device manufacturers being held accountable for usability from early design phase [6] to post-market surveillance [7], it is yet not clear which entities should take the responsibility for improving EHR usability.

Last, unlike medical device manufacturers with established human factors programs and designated staff, healthcare organizations lack IT personnel with the expertise in usability engineering and usability evaluation. EHR analysts in healthcare organizations perform needs assessment, deliver ready-to-use EHRs, and directly support EHR use in

dynamic clinical and organizational contexts. However, only technical skills on EHR configuration and trouble-shooting in combination with a good understanding of clinical needs are currently required for this job role.

Due to the above usability challenges in the EHR world, practical tools that can be operationalized by IT personnel with minimum training in usability are in urgent demand. Healthcare organizations need to be the key players on assuring EHR quality use in dynamic contexts. Among numerous usability evaluation methods, analytical inspection is the efficient approach to screen for low-level predictable usability problems on end user interfaces (UIs). Since the scope of EHR configuration is often limited to textual labels of UI controls, cognitive ergonomics issues should be the first priority for healthcare organizations. Cognitive ergonomics focuses on the understanding of human cognitive abilities and limitations in the contexts of work in order to "improve cognitive work conditions and the overall performance of human-machine systems".[8] The goal of this research is to develop a set of reliable and valid inspection criteria for EHR analysts without systematic training in cognitive ergonomics.

Background

Usability is a multi-faceted concept in the quality model in ISO 25010 Software Quality Requirements and Evaluation (SQuaRE) and in the definitions published by research scholars [9, 10]. In spite of different opinions on usability components and measures, whether user interfaces (UIs) are easy for users to learn to use is one of the few in agreement. Clinicians highly desire EHRs that they can easily figure out how to use due to the time-pressured and interruptive work environment.

Analytical Inspection

Usability evaluation methods vary in the scope, theoretical basis, input, and output. Those that do not require user observation as the input during the evaluation are classified as analytical methods. Some are based on mathematical models such as GOMS model [11]; others are not and are generally referred to as *analytical inspection,* including heuristic evaluation [12], expert review, and cognitive walkthrough [13]. While saving a substantial amount of resources on collecting raw data from users and analyzing them, analytical inspection has limitations on the scope of usability issues being detected and evaluator effects [14].

In the effort of comparing different analytical inspection methods, a set of criteria based on the quantity of usability problems detected were proposed [15]. However, depending on the theoretical basis, these methods vary in the scope of targeted usability problems, and the concept of usability itself consists of multiple facets in very different nature. Simply using problem count as the indicator to compare inspection methods has been criticized [16].

A more helpful guidance for usability practitioners is to distinguish the scope of each method, that is, "what kinds of usability problems a method is and is not good for finding" [17]. Criteria based on the characteristics of the method, including reliability, validity, and downstream utility, are meaningful parameters for method evaluation. Reliability is the extent to which same input can yield same output regardless of the evaluator who applies the method, that is, there is very little evaluator effect [17]. Validity is "the extent to which the findings from analyses conform to those identified when the system is used in the 'real world'' [17]. Downstream utility describes the usefulness of the findings in informing redesign [17]. Downstream utility is especially important for analytical inspection methods because the ultimate goal of early-stage usability evaluation is design improvement.

Heuristic evaluation is a widely adopted inspection method. However, it relies on a limited set of loosely defined principles, which results in significant evaluator effect or low reliability across evaluators [14]. A method with low reliability means that the quality of its findings is highly inconsistent.

Cognitive walkthrough is an inspection method based on a cognitive model that guides evaluators through users' cognitive activities while performing tasks in "walk-up-and-use" applications (i.e. those that can be used with little training). In spite of the more structured evaluation procedure, usability problem detection is still subject to the evaluators' experience. Its focus on evaluating "the ease with which a user completes a task with minimal system knowledge" [18] makes it especially important to EHR usability.

The inspection criteria under development in this research are based on the same theoretical foundation as cognitive walkthrough and the goal is to improve the reliability and validity of usability inspection when the method of

cognitive walkthrough is employed. The inspection criteria are intended to provide guidance on identifying usability issues that make it difficult for users to figure out how to complete tasks, which will be further described in the following section.

Theoretical Foundation

To specify the different types of issues within this general scope and explain the theoretical basis of the proposed inspection tool, an overview of the resource model from the HCI field is provided here. The resource model explains how users figure out how to interact with UIs. The term *resource,* or task-critical information, is defined as the information that can be utilized by users to make interaction decisions [19]. The information on the UIs not only constrains what actions are possible but also helps users decide which action to take in order to accomplish their goals [20].

Users' interaction strategy varies depending on the available *resources* in the specific scenario [19]. *Plan following* strategy requires a pre-defined sequence of actions (i.e. *plan*), which is usually developed through systematic training. However, when there is adequate information on the UIs, users can figure out each action responsively without a pre-defined plan, which saves the mental cost of constructing and executing a plan. This less costly interaction strategy is named as *goal matching*. By matching the effect of a possible action with the current goal, users make interaction decisions without advance planning [19].

Table 1. Resource categories defined in the resource model [20]

Resource Category	Definition
Goal	A desired state of the world
Current state	A collection of attributes used to judge whether the state is closer to the goal
Plan	A pre-defined sequence of actions
Action possibility	Whether an action can be made or not
Action-effect mapping	Link between an action and its effect that will take place after the action is executed
Interaction history	A list of already-taken actions

The resource model reveals the association between *resource* category and interaction strategy. It provides a general guidance on how to determine what *resource* should be made readily available on the UIs in order to facilitate an interaction strategy. However, this theory fails to provide an operationalizable tool to determine whether or not task-critical information is effectively presented on user interfaces. If certain task-critical information is not effectively presented on the user interface, users have to retrieve that information from their memory, which incurs unnecessary cognitive effort.

As to EHRs, low cognitive effort of figuring out how to use is strongly desired because clinical documentation and business operations should not be an additional burden to clinicians' working memory which is already reaching its limit due to the high demand on productivity and safety in a work environment full of interruptions. According to the resource model, action-effect mapping is the key task-critical information that makes the less costly strategy *goal matching* feasible for users. Therefore, the analytical inspection criteria are developed particularly to provide guidance on how to determine whether or not action-effect mapping is effectively presented on user interfaces.

Methods

Preliminary Study

Initially, a metric for cognitive transparency was developed as a tool to guide the inspection of EHR user interfaces (UIs) for ease of use and ease of learning. As shown on Figure 1, the concept of cognitive transparency refers to clear representation of both operations and object-operation relations on UIs so that users can understand what will happen at the work domain level if they click on the UI controls. If the operation of the work that will be accomplished using the UI control is clear on the UI, the operation is considered as being externally represented according to the distributed cognition theory [21, 22]. If the operation is not clear on the UI and requires prior

knowledge in the user's mind, it is considered as being internally represented. Same applies to object-operation relation. When both of the operation and object-operation relation are externally represented on the UI, the UI control is considered as cognitively transparent. The concepts of operation, object, internal representation, and external representation will be illustrated using concrete examples given on Figure 2.

Metric for Cognitive Transparency		Representation of Object-Operation Relation	
		External (clear on the UI)	Internal (memory required)
Representation of Operation	External	Transparent	Not Transparent
	Internal	Not Transparent	Not Transparent

Figure 1. Initially proposed analytical approach: metric for cognitive transparency

On the top of Figure 2, the UI control with the textual label "Edit" is located next to a medication. Clicking "Edit" enables the user to edit that particular medication, where edit is the operation of the work that will be accomplished by clicking "Edit" and the particular medication is the object. In this case, both of the operation and object-operation relation are effectively presented on the UI, or externally represented. The UI control is cognitively transparent. On the bottom of Figure 2, the effect after clicking the UI control with the label of a medication name is not presented at all on the UI. It requires the user to have prior knowledge about the actual effect of this action. In this case, the object is the medication which seems clickable; however, the operation of the work that will be accomplished by clicking the medication name is internally represented. Thus this UI control is not cognitively transparent.
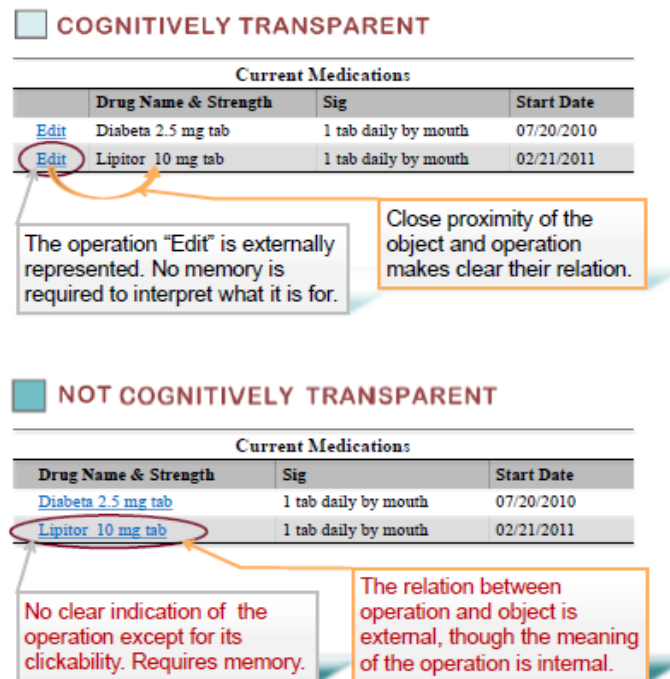
Figure 2. Examples of clickable user interface elements

The validity of this metric was tested in this preliminary study. One evaluator applied the metric to a set of clickable UI controls to predict if they are cognitively transparent or not. The UI controls were selected from E-Prescribing Use Case in three ambulatory EHRs. Three participants with general computer experience were recruited. With the original EHR screenshots presented, participants were asked to anticipate work domain effect of clicking the

selected UI controls, that is, "what do you think will happen in terms of the work that you are doing if you click this". Participants' anticipations were compared to the actual work domain effect of clicking those controls in the live EHRs. If the anticipations from all of the three participants match the actual effect, the UI control was empirically classified as cognitively transparent. If the anticipation from at least one participant does not match the actual effect, the UI control was empirically classified as not cognitively transparent. Subsequently, the predictions made using the metric were compared with the classification based on participants' anticipations. The specificity and sensitivity of the metric were analyzed.

Exploratory Experiment

Based on this initial metric, a set of inspection criteria were developed for health IT personnel with minimum training in cognitive ergonomics to identify issues of presenting action-effect mapping on UIs. According to the resources model, UIs where action-effect mapping is effectively presented are hypothesized to be easier for users to correctly match their goals with possible actions on the UIs.

The exploration for the validity of the inspection criteria on action-effect mapping began from UI controls that enable initiation of entry tasks. The proposed inspection criteria are explained in Table 2. Based on the proximity compatibility principle, the object can be directly presented on the label of the UI control or indirectly presented via immediate neighbor element(s). Published studies have demonstrated that unlabeled icons are significantly more difficult for novice users in all ages to interpret than textual labels [23]. Therefore, only texts or well-recognized symbols are considered as effective presentation.

An exploratory study was conducted to test whether user performance differs between EHR UI controls with effective presentation of action-effect mapping versus those without. Due to lack of published information on UI factors that affect human performance in relevant experiments, screenshots from several EHRs with multiple variations were used in this exploratory study.

Table 2. Proposed inspection criteria for EHR UI controls that initiate entry tasks

Effective Presentation*	Location	Form
Operation	On the same UI control	Text or symbol (e.g. "+" = add)
Object**	On the same UI control	Text
	Immediate neighbor	Text (usually as a header)
		View-only display of instances (e.g. a list of medications)

*A UI control is considered to be transparent only if both the operation and the object involved are effectively presented.
**The object is considered to be effectively presented as long as it falls in any of the three situations.

- *Stimuli*

A total of thirteen screenshots were extracted from four commercial EHRs. The work domain effects of selected UI controls were initiating prescription entry, initiating problem entry, initiating allergy entry, and initiating lab order entry. These screenshots were presented randomly to each participant via online first-click testing software Chalkmark (Optimal Product, New Zealand).

- *Participants and Procedures*

Four physicians with no prior exposure to these particular EHRs were recruited. During the experiment, they were asked to click on the presented screenshot to achieve a given goal. Prior to each screenshot being presented, one goal was indicated on the prompt, which corresponded to the work domain effect of the target UI control. An example prompt is "you would like to prescribe a medication (or order a laboratory test, etc.) to this patient. Where would you click to begin?". The location of their mouse click on each screenshot was collected together with the duration from the screenshot being presented to mouse click (i.e. response time).

- *Confounding Factors*

The variation in the quantity of textual words and pictorial icons on these screenshots was hypothesized to affect participant performance in addition to the independent variable in this exploratory study. The word count and icon count on each screenshot were analyzed with test patient data excluded.

Results

Preliminary Study

The percentages of UI controls with metric-derived prediction matching the empirical classification are presented on Figure 3. In EHR 1, among 35 target UI elements, 60% were predicted by the metric as transparent and had all of the three subjects' anticipations matching the work domain effect in the live EHR (i.e. empirically classified as transparent); 31% were predicted as not transparent and had at least one subject's anticipation not matching the work domain effect in the live EHR (i.e. empirically classified as not transparent); 6% were predicted as not transparent but empirically classified as transparent; 3% were predicted as transparent but empirically classified as not transparent. In EHR 2, among 17 target UI elements, 70% were predicted as transparent and empirically classified as transparent; 12% were predicted as not transparent and empirically classified as not transparent. In EHR 3, among 18 target UI elements, 50% were predicted as transparent and empirically classified as transparent; 39% were predicted as not transparent and empirically classified as not transparent.

EHR 1 (total 35 UI elements)		Observation from Experiment	
		Transparent (agreed by 3 subjects)	Not Transparent (at least 1 subject)
Prediction from Metric	Transparent	**60%**	3%
	Not Transparent	6%	**31%**
EHR 2 (total 17 UI elements)		**70%**	12%
		6%	**12%**
EHR 3 (total 18 UI elements)		**50%**	11%
		0%	**39%**

Figure 3. Proportions of UI controls with metric-derived prediction matching or not matching empirical classification

The specificity and sensitivity of the metric in each EHR are presented on Table 3. This preliminary study suggests that the cognitive transparency metric has high sensitivity and acceptable specificity.

Table 3. Specificity and sensitivity of metric for cognitive transparency

	Sensitivity	Specificity
EHR 1	91%	92%
EHR 2	92%	50%
EHR 3	100%	78%

Exploratory Experiment

Accuracy of response was determined by comparing the location of a participant's mouse click on the screenshot with the location of the target UI control. The small sample size (i.e. number of screenshots) in this exploratory

study makes it difficult to detect any statistical significance from the results. Basic data analysis was carried out to look for any potential difference in the two measures (i.e. accuracy of response and response time) between two groups of UI controls.

As shown in Table 4, the average accuracy of response for the group of UI controls with action-effect mapping effectively presented was 63%, whereas that for the group with action-effect mapping ineffectively presented was 33%. The average response time for the former group was 19 seconds, whereas that for the latter group was 28 seconds.

Table 4. Average accuracy of response and average response time for UI controls with effective or ineffective presentation of action-effect mapping

Presentation of Action-Effect Mapping	Number of Stimuli	Participant Performance		Average Word Count (Icon)
		Average Accuracy of Response	Average Response Time (seconds)	
Effective	10	63%	19	80 (30)
Ineffective	3	33%	28	75 (39)

UI controls with action-effect mapping effectively presented according to the proposed inspection criteria appeared to have a higher average accuracy of response with a shorter average response time. The average word count and average icon count were similar among the two groups of UI controls, suggesting that the quantity of words or icons may not be a contributing factor to the difference in participant performance.

Discussion

The EHR industry urgently needs HCI theories to be translated into inspection tools that can be operationalized by health IT personnel with minimum training in cognitive ergonomics, which empowers healthcare organizations to improve EHR usability and thereby assure EHR quality use in dynamic operational and clinical contexts.

Among different components of usability, clinicians strongly desire EHRs that are easy to figure out how to use without systematic training and memorization. Operating EHRs at the point of care should incur little unnecessary cognitive effort and clinicians' limited mental capacity should be focused on patient care. The theory-based criteria under development in this research are targeted at supporting analytical inspection of EHR UIs from this aspect.

Analytical inspection is the efficient approach to detecting low-level predictable usability problems because it does not require collection of raw data from user observation. The quality of findings relies on the theory basis of the inspection method or tool. Healthcare organizations can benefit the most from the HCI field by employing validated inspection tools in the EHR configuration process.

However, there is little agreement on how to empirically validate an inspection tool. The preliminary study and the exploratory experiment reported above are validation attempts using different research design. In the preliminary study, the target UI controls was pointed out on the screenshots presented to participants, and the raw data collected from participants were their anticipations of work domain effects after clicking those UI controls in live EHRs. In contrast, the exploratory experiment presented screenshots to participants without pointing out the target UI controls, but with specified goals that corresponded to the work domain effects of the target UI controls. The participants' choices of UI controls for the given goals were the collected empirical data. The research design in the exploratory experiment well simulated user interactions in the real world, compared to the preliminary study, thus will be adopted in future research on validating other proposed inspection tools.

Although objective performance measures, such as the location of clicking, are more robust indicators of the effectiveness of use than subjective measures, such as confidence rating, it is still important to include subjective measures in future research to understand participants' subjective perception with regard to the clarity of UIs. Other

UI factors that are different across EHRs and may affect human performance were not controlled in the exploratory experiment. A controlled experiment will be conducted in the future to establish the causal effect relationship between the effectiveness of action-effect mapping presented on UIs and participants' accuracy of response. Meanwhile, it is critical to expand the sample size so that statistical significance can be tested.

In addition to validity, reliability is another key criterion to evaluate inspection tools. Future research will test the reliability of the proposed inspection tools. Multiple evaluators with various levels of experience in usability will independently apply the inspection tools to detecting ineffective presentation of action-effect mapping on EHR UIs. Their detection results will be compared and analyzed statistically.

References

1. Stead WW, Lin HS, eds. Computational technology for effective health care: immediate steps and strategic directions. Washington, DC: National Academies Press. 2009.
2. Arnaut LY, Greenstein JS. Human factors considerations in the design and selection of computer input devices. In S. Sherr (Ed.), Input Devices 1988;71-121. San Diego, CA: Academic Press.
3. Human factors design process for medical devices. HE74:2001. ANSI/AAMI.
4. Middleton B, Bloomrosen M, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013;20:e2-e8.
5. Orlikowski WJ, Barley SR. Technology and institutions: what can research on information technology and research on organizations learn from each other? MIS Quarterly. 2001;25:145-165.
6. U.S. Food and Drug Administration. Premarket Information - Device Design and Documentation Processes. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HumanFactors/ucm119190.htm.
7. U.S. Food and Drug Administration. Postmarket Information - Device Surveillance and Reporting Processes. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HumanFactors/ucm124851.htm.
8. Hoc JM. Towards ecological validity of research in cognitive ergonomics. Theoretical issues in ergonomics science 2001;2(3):278-288.
9. Nielsen J. Usability Engineering. AP Professional. 1993. New York.
10. Zhang JJ, Walji MF. TURF: Toward a unified framework of EHR usability. J Biomedical Informatics. 2011; 44(6):1056-1067.
11. Card SK, Moran TP, Newell A. Computer text-editing: an information-processing analysis of a routine cognitive skill. Cognitive Psychology. 1980;12 (1):32-74.
12. Nielsen J. Heuristic Evaluation. In J. Nielsen & R. Mack (Eds.), Usability Inspection Methods. New York: John Wiley. 1994:25-62.
13. Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. Mack (Eds.), Usability inspection methods. New York: John Wiley. 1994:105-140.
14. Hertzum M, Jacobsen NE. The evaluator effect: a chilling fact about usability evaluation methods. Int. J. Human-computer Interaction. 2001;13(4):421-43.
15. Hartson HR, Andre TS, Williges RC. Criteria for evaluating usability evaluation methods. Int. J. Human-computer Interaction. 2003;15(1):145-181.
16. Gray WD, Salzman MC. Damaged merchandise? A review of experiments that compare usability evaluation methods. Human Computer Interaction Journal. 1998; 203-261.
17. Blandford AE, Hyde JK, Connell I, Green TRG. Scoping analytical usability evaluation methods: a case study. Human Computer Interaction Journal. 2008;23(3):278-327.

18. Mahatody T, Sagar M, Kolski C. State of the art on the cognitive walkthrough method, its variants and evolutions. Intl. J. Human-Computer Interaction. 2010; 26(8): 741-785.
19. Wright P, Fields B, Harrison M. Analyzing human-computer interaction as distributed cognition: the resource model. Human-computer Interaction. 2000;(15):1-41.
20. Fields B, Wright P, Harrison M. Designing human-system interaction using the resource model. APCHI'96: First Asia Pacific Conference on Human Computer Interaction. 1996;181-191.
21. Zhang JJ, Norman DA. Representations in distributed cognitive tasks. Cognitive Science. 1994;18:87-122.
22. Zhang JJ, Patel VL. Distributed cognition, representation, and affordance. Pragmatics & Cognition. 2006;14(2): 333-341.
23. Wiedenbeck S. The use of icons and labels in an end user application program: an empirical study of learning and retention. Behavior and Information Technology. 1999;18(2): 68-82.