Xie Dong. Assessing the Relationships between Task demand, Task difficulty and Performance during Physicians' Interaction with EHR systems. A Master's Paper for the M.S. in IS degree. April, 2018. 32 pages. Advisor: Lukasz Mazur

The study's aim was to relate task demand, task difficulty with performance during physicians' interaction with electronic health records (EHR) system. The results indicated that there was a significant effect of task demands on task difficulty and performance; task difficulty was also related to performance. Practically, the results suggest that EHR designers might be able to positively affect physicians' performance by enhancing usability of interfaces aimed at directing physician' EHR-related interaction strategies.

Headings:

Task Demand

Task Difficulty

EHR system

Strategy

Interface Redesign

TASK DEMAND, TASK DIFFICULTY, EHR SYSTEM, STRATEGY, INTERFACE REDESIGN

by
Xie Dong

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2018

Approved by

____

Lukasz Mazur

Table of Contents

1 INTRODUCTION

Despite of many advantages, suboptimal functionality, usability, and use of electronic health record system (EHR) can "create new hazards in the already complex delivery of health care" (Lin & Stead, 2009; Vincente, 1999; Peute & Jasper, 2007; Russ et al., 2014; Elliott, Young, Aguiar & Kolm, 2014). Based on a survey of the memberships of the American Society for Healthcare Risk Management and the American Health Lawyers Association, the most common EHR-related safety concerns included incorrect patient identification, failure to read a computer-generated warning or alert, failure to identify, find or use the most recent patient data and incorrect orders (Sittig & Singh, 2013). Moreover, evidence suggests that physicians and nurses who have difficulty in using EHR systems make more errors (Kellogg & Fairbanks, 2017).

Automated alerts about potential errors (e,g., drug/allergies interactions) in EHRs can add a significant value to patient safety, however, they can also create alert fatigue and thus performance degradation (Zagorski, 2017; Kaipio & Lääveri, 2017). For example, clinicians frequently fail to acknowledge and/or follow-up on critical abnormal test results, despite being alerted by EHRs (Singh et al., 2007; Callen et al., 2012). On the other hand, when supported by 'optimized' usability, automated alerts in EHRs were shown to improve providers' acknowledgment and/or follow-up of abnormal test results (Moore et al., 2008; Lin & Moore, 2011). Therefore, there is evidence that properly designed and implemented EHRs can improve care quality and patient safety, while reducing ever increasing healthcare cost (Noblin & Cortelyou-Ward, 2013; Tanner & Gans, 2015).

Overall, there is a need to better understand providers' interaction with EHRs. Therefore in this exploratory research, drawing on the theories related to mental capacity and attention (Kahneman 1973), the specific aims are to understand following relationships during providers' interaction with EHRs:

- Aim #1: To assess the relationship between task demand and task difficulty.
- Aim #2: To assess the relationship of task demand and task difficulty and providers' performance.
- Aim #3: To assess the relationship between task difficulty and performance.

1.1 Background

According to basic theories of attention, the effect on performance is both due to workload due to task demands and task difficulty (i.e., how the task is experienced), and it is dependent on the context, state, capacity, and strategy of allocation of mental resources (De Waard, 1996; Kantowitz, 1987; Parasuraman & Hancock, 2001), which seems relevant during physicians' interaction with EHRs.

1.1.1 Task Demand

Task demand is determined by the goal that must be attained by means of task performance and it is *independent* of the individual (Kahneman, 1973; Parsuraman & Hancock, 2001; Robinson, 2001). In other words, variation in task demand can be achieved when task characteristics are changed (Campbell, 1988; Hancock and Williams, 1995). For example, in studies conducted in driving simulators, task demands were manipulated by increasing number of vehicles on the road (De Waard, 1996). In healthcare, Mosaly and colleagues (2017) varied task demands based on number of sub-tasks involved in each scenario. In human-computer interaction (HCI) domain, task demand was quantified using external features of the task, especially interface design elements like color, shapes, textures, and pictorial graphics (Bedny & Karwowski, 2012).

1.1.2 Task Difficulty

Task difficulty is the regulated behavior or adaptable strategy used to cope with increased task demand to mitigate mental effort required to perform the task (Kantowiz, 1987; Parsuraman & Hancock, 2001). Task difficulty is defined as interaction between the task and individual, and changes in task demands have usually significant effect on task difficulty and performance (De Waard, 1996; Merat, Antilla & Luoma, 2005; Moray, Dessouky, Kijowski, & Adapathya, 1991; Parsuraman & Hancock, 2001; Szalma, 2002). For example, in airplane flying tasks under high task demands conditions (i.e., low visibility), pilots' strategies to safely land the aircraft were correlated to potential unsafe levels of steep descendent (Boehm-Davis & Casali, 2007). In human-computer interaction, computer mouse 'events' were frequently used to represent different behaviors correlating to various levels of performance. For example, task difficulty was measured via computer mouse clicks, click speed, and scrolls (Macaulay, 2004; Lin & Imamiya, 2008; Arguello, 2014). Further, task-flows were also used as a measure of task difficulty. For example, De Alwis & Murphy (2008) measured task difficulty in programming tasks via task-flows described as connectedness and sequence of solution program elements that depended on the mechanism of objective oriented programming.

Further, in computer-based tasks, task difficulty was also measured via task-flow patterns, eye gaze, and eye scan patterns (Bedny & Karwowski, 2012). During interaction with EHRs, Mosaly and colleagues (2017) quantified task difficulty based on deviation from the instructed task-flow and click patterns (i.e., combination and sequence of navigation, input and decision clicks) and found that specific click patterns were related with task demands and performance.

1.1.3 Performance

Performance is the most immediate upstream surrogate for outcome, and is therefore a commonly considered metric (Vidulich & Wickens 1986; Parasuraman 1993; Karsh & Holden 2006; Vozenilek & Gordon 2008; Carayon & Gurses 2008). Performance is measured based on the context of the task via errors and reaction times, specifically in laboratory settings (De Waard, 1996). For human-computer interaction settings, task completion time and error rates are the most commonly used measures of performance (Leporini & Paternò, 2008; Wycislik & Warchal, 2014; Goncalves & Sarsenbayeva, 2016). For example, Mazur and colleagues (2016) used task completion time, errors, and severity of errors as measures of performance during physician-EHR interactions. Mosaly and colleagues (2017) used omission ad commission error as a measure of performance.

2 METHODS

2.1 Participants and Procedure

This study was performed in a simulated environment as part of an institutional review board (IRB) approved study. The focus of the study was on the follow-up of abnormal test results in an Epic-based EHR environment. The Epic Playground was used and was designed to recreate the ‘real’ clinical environment.

Invitations to participate in the research study were sent to all residents and fellows in the school of medicine at one large academic institution, while clearly stating the need for experience with Epic as related to our simulated scenarios. All participants were incentivized to participate with a $100 gift card. Final selections were made based on subjects’ availability to participate in the study during designated weeks for data collection. We enrolled a total of 38 residents from Internal Medicine, Family Medicine, Pediatrics, gynecology oncology, psychiatry, and surgery (PGY: 1 to 5) departments to participate in the study (see Table 1 for details).

Table 1 Composition of participants within each experiment.

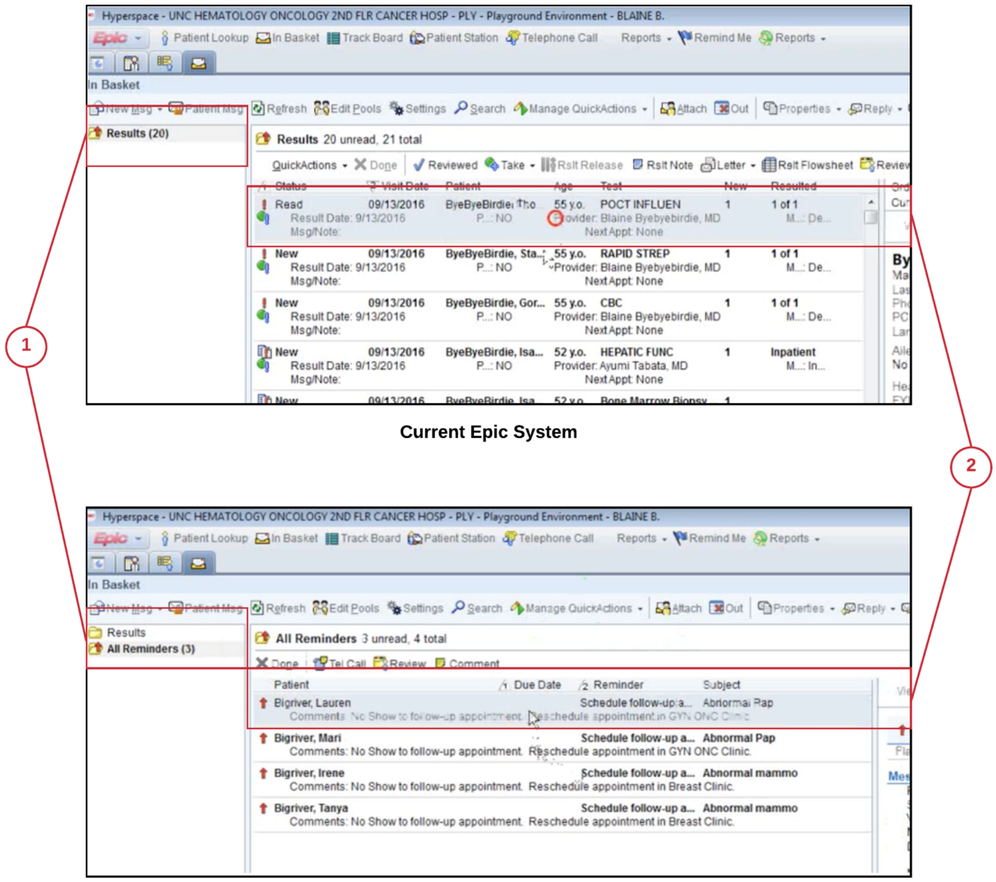
Specialty	**N**	**Post Graduate Year (PGY) PGY:** ***count***	**Gender F: Female; M: Male**
Internal Medicine	14	1:*4* 2:*2* 3:*5* 4:*3*	F:9 M:5
Family Medicine	4	1:*1* 2:*1* 3:*1* 4:*1*	F:2 M:2
Pediatrics	9	1:*3* 2:*2* 3:*4* 4:*0*	F:7 M:2
Surgery (general, neuro, ortho, head & neck)	5	1:*1* 2:*2* 3:*0* 4:*1* 5:*1*	F:3 M:2
Other (cardiology, psychiatry, critical care, ob/gyn)	6	1:*1* 2:*1* 3:*1* 4:*2* 5:*1*	F:3 M:3
Total	**38**	**1:*10*** **2:*08*** **3:*11*** **4:*06*** **5:*03***	**F:24** **M:14**

2.2 Simulated Sessions

Participants were asked to review test results in the "In-basket" consisting of a list of normal results and abnormal results, which including abnormal with *no-show* status patients results. Participants performed two simulated sessions, a baseline session (session #1) and a test session (session #2). Participants were randomized to get high-volume (13 abnormal with 5 *no-show* status) vs. low-volume (8 abnormal with 4 *no-shows* status) of patient results in the "In-basket".

2.3 Interface Design

Two EHR interface designs were used for assessing enhancement on *no-show* status patients. The 'enhanced' interface was made by creating a separate folder called 'All Reminders' and moving all abnormal results with *no-show* status into this folder. Results in "All Reminder" specifically display no-show status of the patient in the snapshot, thus minimizing clutter in the 'In-basket' which consisted on normal and abnormal results only. In contrary, the 'current' interface consisted of all patient results (normal, abnormal, and abnormal with *no-show*) all presented in 'In-basket' (Figure 1), and does not any information/highlight indicating *no-show* status. All 38 participants performed tasks on abnormal results in current interface design, but only 18 of them were randomly assigned to perform task on abnormal results with *no-show* status in enhanced interface design.



Current Epic System

Enhanced Epic System

Figure 1 Differences in Interface Design. Difference ①: In current Epic design, all abnormal test results were put in one folder ("Results"). In enhanced Epic design, no-show abnormal results were put into a separated folder ("All Reminders").Difference ②: In current Epic design, no information of record status was showed in record list. In enhanced Epic design, a one-sentence summary indicating no-show status was showed in record list of no-show folder.

Table 2 Number of patient results interacted by the study participants in two epic-EHR design for both low- and high-volume conditions.

Epic Interface Design	Low Volume N=10 for Current Epic N=9 for Enhanced Epic	High Volume N=10 for Current Epic N=9 for Enhanced Epic
Current-Epic *# of patient results*	74	118
Enhanced-Epic *# of patient results*	71	112

2.4 Procedure

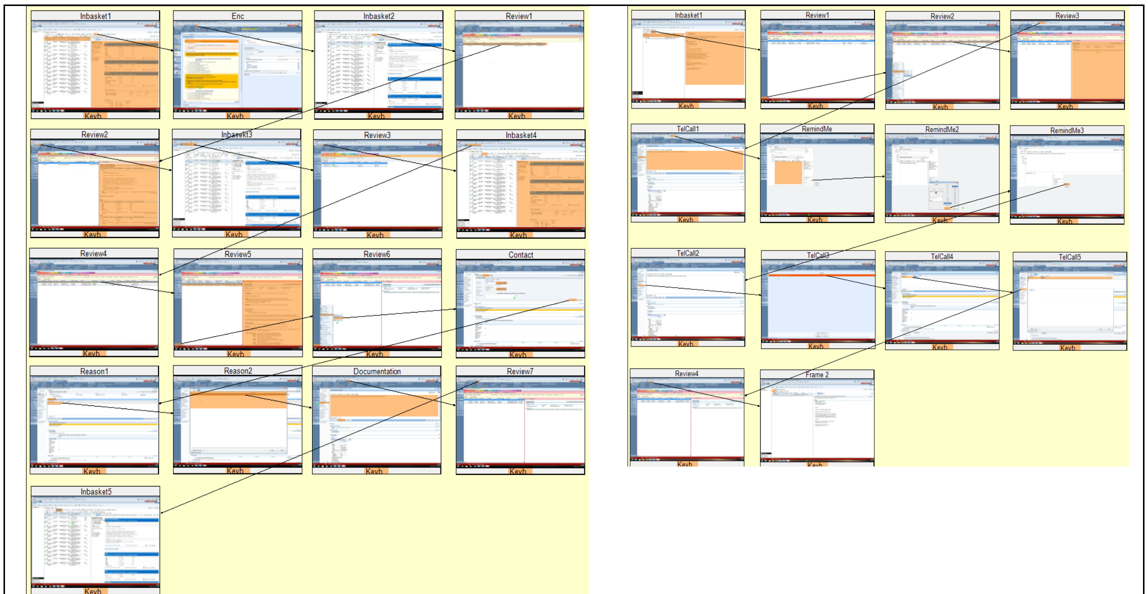
After arriving to the lab, participants were given a brief introduction about the study and asked them to sign the consent form. They were then asked to sit on a chair in front of a computer screen (24-inch monitor with 1024 x 1900 resolution) and adjust to their comfort. First, they were asked to read the instructions as displayed on the computer and open Epic-EHR to start the experiment session. Tobii X-60 eye tracker and Eyeworks Inc software was used to collect screen video recordings and mouse clicks. Recorded video were watched and analyzed by 2 researchers independently. Any discrepancies were resolved during weekly research meetings to reach to consensus.

2.5 Data Collection

2.5.1 Task demand

Based on experiment design, task demand was determined via Epic EHR interface design and volume of patient results. Current Epic and high volume of patient results were considered as *highest* task demands, whereas enhanced Epic and low-volume of patient results were considered as *lowest* task demands. We used cognitive predictive model

(CogTool) to validate highest and lowest levels of task demand (Anderson & Corbett, 1995; Sherry & Medina, 2008; Xian & Jin, 2014; Suzuki & Nakao, 2009; Harris & John, 2010). In line with our expectation, the model predicted longest completion times for current Epic design with high patient volume (115 seconds) when compared to enhanced Epic design with low patient volume condition (97 seconds) (see Figure 2 for reconstructed workflows in CogTool).



(current Epic Design) (enhanced Epic Design)

Figure 2 Screenshot of Cognitive Modeling Procedure with CogTool on two selected tasks for current vs. enhanced epic interface.

2.5.2 Task difficulty

Task difficulty was quantified based on interaction task-flows and click behaviors. Task-flows, where significant comprehension and decision-making were involved, were coded into six categories for both current and enhanced Epic interfaces (Table 3). User behaviors under same interface were coherent since each interface provided limited functional modules and it was usually designed to guide users following most common task-flow (comprehension→decision-making). The comprehension interfaces included:

(1) 'In basket/All reminder', when selected a patient result, these interfaces ('In basket' for current Epic and 'All Reminder' for enhanced Epic) provide a snapshot of patient condition and lab/test results.

(2) 'Chart Review' interface provides comprehensive information about previous note, labs, images, encounters, etc. about the patient.

The decision-making interfaces included:

(1) 'Telephone Call' interface enables participants to select patient's mode of contact → input reason for call → select diagnosis/medication & orders → associate diagnosis with orders → complete documentation → sign encounter.

(2) 'Encounter' interface enables participants to create a plan by providing some 'best practice advisories' and choose/search/select from problem list and visit diagnosis → select medication & orders → associate diagnosis with orders → sign visit.

(3) 'Orders only' interface enables participants to choose/search/select from problem list → access smart sets → write progress notes → select diagnosis and order and associate them → sign visit.

(4) 'Letter (documentation only)' interface enable participants to leave a note (either to patient or concerned clinical staff for follow-up), without involving any access to orders or medication lists. Participants' mouse clicks were obtained from video analysis. Each click action was coded into three categories: (1) navigation clicks (e.g., moving from one tab to another for navigational purposes); (2) decision clicks (e.g., selecting a test or medication to order or cancelling a selected order or search result); and (3) Input clicks

(e.g., placing the mouse cursor into the search box to type search terms) using the similar procedures used by Mazur et al (2016). The number of total clicks and proportions of each category of clicks type were used for data analysis.

2.5.3 Performance

Performance was quantified using three measures, (1) errors (2) efficiency measurements based on number of revisits to patient information and number of searches; (3) task completion time. For each patient result, errors were coded into commission errors (where participant identify abnormal results, but failed to take appropriate actions, for example, duplicating existing order/referral, failing to follow-up via clinical note or reminder to patients or appropriate clinical staff). To assess efficiency during tasks, we measured (1) revisited (more than once) patient information; (2) number of searches quantified as 'regular' number of searches (1 to 3 searches during task) or 'increased' number of searches (more than 3 searches during task). Time taken to complete each patient result was obtained from the video analysis.

2.6 Data Analysis

Before data analysis, we completed tests for normality and equal variance for all study variables using Shapiro-Wilk's and Bartlett test respectively. Results indicated that assumptions were satisfied (normality: all $p > 0.05$; equal variance: all $p > 0.05$).

Aim #1: To assess the relationship between task demand and task difficulty.

A nominal logistic regression was used to assess the relationship between task demand and task-flow. A multiple linear regression was used to assess the relationship between task demand and total clicks.

Aim #2: To assess the relationship of task demand and providers' performance.

A nominal logistic regression was used to assess the relationship between i) task demand and errors (no errors vs. commission errors), ii) task demand and number of searches, and iii) task demand and number of revisits to patient information. A multiple linear regression was used to assess the relationship between task demand and task completion time.

Aim #3: To assess the relationship between task difficulty and performance.

A nominal logistic regression was used to assess the relationship between i) task-flow and errors, ii) mouse clicks and errors, iii) task-flow and number of searches; iv) mouse clicks and number of searches, v) task-flow and number of revisits to patient information; and vi) mouse clicks and number of revisits to patient information. A multiple linear regression will be used to assess the relationship between i) task-flow and task completion time, and ii) mouse clicks and task completion time. An alpha level of 0.05 was set for significance testing. All analyses were performed using SAS 94.

3 RESULTS

Table 3 presents the descriptive statistics of task demand, task difficulty and performance quantifications.

3.1 Relationship between Task Demand and Task Difficulty

3.1.1 Relationship between Task Demand and Task-flow

Results indicated a significant relationship between task demand and task-flow ($\chi^2_{(10, n=375)}$ =87, p<.001), indicating that participants in lower task demand were more likely to follow *In-basket/All-Reminder→Chart Review→Telephone Call*, whereas the ones in high task demand were more likely to follow *In-basket/All Reminders→Telephone Call* ($\chi^2_{(6, n=375)}$ =24, p<.01) (Table 3).

Table 3 Descriptive Statistics for Task Demand, Task Difficulty and Performance

Measure			**Task Demand**			
			Current Epic (N=210)		Enhanced Epic (N=189)	
			Low-volume (n=74)	High-volume (n=118)	Low-volume (n=71)	High-volume (n=112)
Task Difficulty	Task-flow (counts)	In-basket*/All Reminders** → Letter (Documentation)	0	0	7	0
		In-basket*/All Reminders** → Encounters/Orders Only	7	13	9	5
		In-basket*/All Reminders** → Telephone Call	10	43	21	51
		In-basket*/All-Reminder**→Chart Review → Letter (Documentation)	2	0	3	3
		In-basket*/All-Reminder**→Chart Review → Encounters/Orders Only	21	12	19	28
		In-basket*/All-Reminder**→Chart Review → Telephone Call	34	50	12	25
	Mouse Click *Average (sd)*	Total clicks	29 (10)	32 (12)	25 (11)	26 (12)
		Navigation clicks (%)	50 (12)	53 (12)	54 (12)	55 (13)
		Input clicks (%)	14 (4)	14 (5)	16 (5)	15 (6)
		Decision clicks (%)	36 (11)	32 (10)	30 (10)	29 (10)
Performance	Time *Average (sd)*	Time to complete (in sec) per patient result	131 (51)	149 (66)	123 (58)	113 (69)
	Errors (counts)	No errors vs. commission	36 vs. 38	58 vs. 60	59 vs. 12	84 vs. 28
	Efficiency (counts)	Revisits to patient information *No vs. Yes*	53 vs. 21	85 vs. 33	55 vs. 17	92 vs. 20
		Number of searches *Regular searches (3 or less) vs. Increased searches (>3)*	36 vs. 38	58 vs. 60	59 vs. 12	84 vs. 28

*In-basket – corresponds to Current Epic
**All Reminder – corresponds to Enhanced Epic

3.1.2 Relationship between Task Demand and Computer Mouse Clicks

Results indicated that there was a significant relationship between tasks demand and computer mouse clicks (R^2=.3, $F_{(1,375)}$=10.4, p<.01) indicating that high task demand had increased number of total clicks compared to low task demands (on average: 31 vs. 25;

Table 3). Additionally, high task demand had increased proportion of decision clicks (R^2=.32, $F_{(1,375)}$=6, p=.01) compared to low task demands (35% vs. 29%; Table 3).

3.2 Relationship between Task Demands and Performance

3.2.1 Relationship between Task Demand and Errors

Results indicated a significant relationship between task demands and errors ($\chi^2_{(2,\ n=375)}$ =45, p<.01), indicating that high task demands significantly increased the odds of errors by 4 (p<.01), when compared to low task demands (Table 3).

3.2.2 Relationship between Task Demand and Efficiency

There was a significant relationship between tasks demands and number of searches ($\chi^2_{(2,\ n=375)}$ =35, p<.01), indicating that high task demand significantly increased the number of searches by 4 times when compared to low task demand. There was no significant relationship between task demand and number of revisits to patient information (Table 3).

3.2.3 Relationship between Task Demand and Task Completion Time

There was a significant relationship between task demand and task completion time (R^2=0.32, $F_{(2,\ 375)}$=3, p=0.05), indicating that high task demand significantly increased task completion time per patient results, when compared to low task demands (on average: 139 vs. 117; Table 3).

3.3 Relationship between Task Difficulty and Performance

3.3.1 Relationship between Task-flow and Errors

There was a significant relationship between task flow and errors ($\chi^2_{(5,\, n=375)} = 10$, $p=.05$), indicating that the odds of making an error was 2 times more likely in task-flow *In-basket/All Reminders→Encounters/Orders Only* compared to *Chart Review→Telephone Call* (Table 4).

3.3.2 Relationship between Task-flow and Efficiency

There was a significant relationship between task-flows and revisits to patient ($\chi 2_{(5,\, n=375)} = 10$, $p=.05$). Task flow *In-basket/All Reminders→Telephone Call* had 2 times more number of revisits compared to task-flow *Chart Review→Encounters/Orders Only*.

There was no significant relationship between task-flows and number of searches (Table 4).

3.3.3 Relationship between Task-flow and Task Completion Time

There was a significant relationship between task-flow and task completion time ($R^2=.43$, $F_{(5,375)}=9$, $p<.01$), indicating that task-flow *Chart Review→Telephone Call* (mean=152 (sd=6) seconds) and *Chart Review→Encounters/Orders Only* (mean=147 (sd=7) seconds) had significantly longer task completion time compared to *Chart Review→Letter (documentation only)* (mean=45 (sd=22) seconds) (Table 4).

Table 4 Descriptive Statistics for Task Difficulty (Task-flows) and Performance

Task Difficulty		Performance			
		Errors *No-error vs. Commission* [Count]	# of searches *Regular vs. Increased searches* [Count]	# of Revisits to patient information *No vs. yes* [Count]	Task Completion Time Mean (sd)
Task-flow	In-basket*/All Reminders** → Letter (Documentation)	7 vs. 0	0 vs. 24	7 vs.0	61 (32)
	In-basket*/All Reminders** → Encounters/Orders Only	17 vs.17	7 vs. 0	25 vs. 9	111 (43)
	In-basket*/All Reminders** → Telephone Call	80 vs. 45	17 vs. 17	88 vs. 37	121 (60)
	In-basket*/All-Reminder**→ Chart Review → Letter (Documentation)	6 vs. 2	80 vs. 45	7 vs. 1	99 (43)
	In-basket*/All-Reminder**→ Chart Review → Encounters/Orders Only	48 vs. 32	6 vs. 2	68 vs. 12	129 (68)
	In-basket*/All-Reminder**→ Chart Review → Telephone Call	79 vs. 42	48 vs. 32	89 vs. 32	151 (68)

3.3.4 Relationship between Computer Mouse Clicks and Errors

There was a significant relationship between computer mouse clicks and errors ($\chi2_{(1,375)}$=13, p<.001), indicating that participants with errors had most number of clicks when compared to participants with no-errors (on average: 33 vs. 27, p<.01; Table 3). Furthermore, there was a significant relationship between proportion of navigation clicks ($\chi2_{(1,375)}$=28, p<.01), and proportion of decision clicks ($\chi2_{(1,375)}$=35, p<.001) and errors, indicating that participants with errors had increased proportion of input clicks, but decreased proportion of decision clicks (see Table 5).

3.3.5 Relationship between Computer Mouse Clicks and Efficiency

There was a significant relationship between total computer mouse clicks and number of searches ($\chi2_{(1,375)}$=13, p<.01), indicating that increased total number of computer mouse clicks had a negative relationship with number if searches (odds=0.2; Table 5). There was also a significant relationship between proportion of navigation clicks ($\chi2_{(1,375)}$=28, p<.01) and proportion of decision clicks ($\chi2_{(1,375)}$=36, p<.01) with number of searches; that is, increased number of searches significantly increased the proportion of navigation clicks, but reduced the proportion of decision clicks. There was no relationship between proportions of input clicks with number of searches (p>0.05).

There was no significant relationship between proportion of clicks and revisit to patient information (p>0.05).

3.3.6 Relationship between Computer Mouse Clicks and Task Completion Time

There was a significant relationship between computer mouse clicks and task completion time (R^2=.63, $F_{(1,375)}$=53, p<.0001), indicating that increased number of computer mouse clicks increased task completion time, mostly due to increased proportion of input clicks (R^2=.33, $F_{(1,375)}$=6, p=.01; see Table 5).

Table 5 *Descriptive Statistics for Task Difficulty (Clicks) and Performance*

Task Difficulty		**Performance**			
		Errors *No-error vs. Commission*	# of searches *Regular vs. Increased searches*	# of Revisits to patient information *No vs. yes*	Task Completion Time
Clicks *Mean(sd)*	Total clicks	27(11) vs. 33 (11)	28(11) vs. 29(13)	25(11) vs. 35(11)	125(65)
	Navigation clicks (%)	55(11) vs. 47 (9)	56(11) vs. 54(17)	55(14) vs. 55(11)	
	Input clicks (%)	15(5) vs. 16(5)	15(5) vs. 13(7)	14(6) vs. 15(5)	
	Decision clicks (%)	30(9 vs. 37(9)	30(9) vs. 33(12)	31(11) vs. 30(9)	

4 DISCUSSION

4.1 Aim#1: Relationship between task demand and task difficulty

High task demand, as represented by current interface design and high volume of patient results, significantly increased task difficulty. That is, participants subjected to high task demand had more computer mouse clicks (on average by 10 clicks) while using 'abbreviated' workflows (skipping *Chart Review*; an interface that provides detailed information about the patient [e.g., previous notes, encounters, labs, images etc.; information that is not provided in other comprehension interfaces]; 30% less utilization overall), when compared to participants in low task demands. These results are in line with previous findings by Mazur et al. (2016) and Mosaly et al (2017) who also found a similar relationship between task demand and task difficulty.

4.2 Aim #2: Relationship between Task Demand and Performance.

Increased task demand decreased performance. Specifically, under high task demand, we found that participants made more commission errors while being less efficient, and took longer time. Sub-optimal interface design have found to increase errors despite of experience and training (Sittig & Singh, 2013; Kellogg & Fairbanks, 2017). We found that in high task demands, participants were unable to access required information (e.g., no-show status) and proceeded to address the patient's condition as new patient, thus placing new referrals and orders. While signing the orders, they frequently encountered with error message indicating duplicate orders, which were overridden repeatedly by physicians. This behavior also led to more clicks, specifically increased decision clicks. Previous studies also found that task demand were positively related to increased number of errors and severity of errors (Mazur et al 2016, Mosaly 2017).

4.3 Aim #3: Relationship between Task Difficulty and Performance

Task difficulty significantly affected all measures of performance. Specifically, under high task demand, we found the task-flows that did not include the *Chart Review* interface, which included patient status information, to be more likely to generate errors, while being less efficient. In low task demand, the patient status was displayed in the snapshot (All Reminders) and thus participants who did not access the *Chart Review* interface were still able to maintain good performance, while being more efficient.

4.4 Limitation

There are several limitations in this study, which could provide suggestions for further study. First, since this study was implemented on 38 participants (resident physicians and medical students) from the same institution, they were more likely to taking similar actions on particular clinical tests results based on their education background and training history. Within the experiment scope, the reliability of study could be improved by recruiting participants from several institutions. Second, there was no exact standard to scoring performance. Physician's performance for each task was measured in errors, which was decided by if physician correctly recognized abnormal result and if took appropriate actions. Although experts decided on the final code of errors for each task, there was still a possibility of misunderstanding physician's behavior. For example, when handling patients with no-show status, we decided that duplicating previous orders was an inappropriate action, thus coded it as error. Third, although we defined different level of task demand by different task characteristic in interface design and volume of patient results, we were not able to scale them to understanding how much they differed. The same problem of scaling also existed in quantification of task difficulty. Further studies could work on scaling task demand and task difficulty levels.

5 CONCLUSION

The study's aim was to relate task demand, task difficulty with performance during physicians' interaction with EHR system. The results indicated that there was a significant effect of task demand on task difficulty and performance; task difficulty was also related to performance.

Considering theoretical implications, the results suggest that performance is affected by both task demands and task difficulty, and task difficulty could be determined the strategies and behavior of the physicians (or providers) while interacting with EHRs. Considering practical implication, the results suggest that EHR designers might not be able to positively affect physicians' performance by enhancing usability of interfaces aimed at directing physician' interaction strategies.

Bibliography

National Research Council. (2009). *Computational technology for effective health care: immediate steps and strategic directions*. National Academies Press.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC Press.

Peute, L. W., & Jaspers, M. W. (2007). The significance of a usability evaluation of an emerging laboratory order entry system. *International journal of medical informatics, 76(2),* 157-168.

Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., ... & Hawsey, J. M. (2014). Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. *Journal of the American Medical Informatics Association, 21(e2),* e287-e296. Elliott, D. J., Young, R. S., Brice, J., Aguiar, R., & Kolm, P. (2014).

Effect of hospitalist workload on the quality and efficiency of care. *JAMA internal medicine, 174(5),* 786-793.

Tanner, C., Gans, D., White, J., Nath, R., & Pohl, J. (2015). Electronic Health Records and Patient Safety. *Applied clinical informatics*, *6*(01), 136-147.

Noblin, A., Cortelyou-Ward, K., Cantiello, J., Breyer, T., Oliveira, L., Dangiolo, M., ... & Berman, S. (2013). EHR implementation in a new clinic: a case study of clinician perceptions. *Journal of medical systems*, *37*(4), 9955.

Sittig, D. F., & Singh, H. (2013). A red-flag-based approach to risk management of EHR-related safety concerns. *Journal of Healthcare Risk Management*, *33*(2), 21-26.

Kellogg, K. M., Fairbanks, R. J., & Ratwani, R. M. (2017). EHR Usability: Get It Right from the Start. *Biomedical instrumentation & technology*, *51*(3), 197-199.

Zagorski, N. (2017, July 19). Are Today's EHR Alerts Putting Clinicians, Patients at Risk? Retrieved from https://psychnews.psychiatryonline.org/doi/full/10.1176/appi.pn.2017.5b23

Kaipio, J., Lääveri, T., Hyppönen, H., Vainiomäki, S., Reponen, J., Kushniruk, A., ... & Vänskä, J. (2017). Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. *International journal of medical informatics*, *97*, 266-281.

Clarke, M. A., Belden, J. L., & Kim, M. S. (2014). Determining differences in user performance between expert and novice primary care doctors when using an electronic health record (EHR). *Journal of evaluation in clinical practice*, *20*(6), 1153-1161.

Sheffer, J., Hettinger, Z., Lowry, L., Mander, A., Suarez, W., Weinger, M. B., & Wiklund, M. (2017). A Roundtable Discussion: Usability, Human Factors, and Health IT: Providing Effective Guidance While Nurturing Innovation. *Biomedical instrumentation & technology*, *51*(3), 252-259.

Van der Veer, G. C. (2008). Cognitive Ergonomics in Interface Design-Discussion of a Moving Science. *J. UCS*, *14*(16), 2614-2629.

da Silva, F. P. (2014). Mental Workload, Task Demand and Driving Performance: What Relation?. *Procedia-Social and Behavioral Sciences*, *162*, 310-319.

Hancock, P. A., Williams, G., & Manning, C. M. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, *5*(1), 63-86.

Tabari, M. A., & Ivey, T. A. (2015). Cognitive task complexity effects on L2 writing performance: An application of mixed-methods approaches1. *Latin American Journal of Content & Language Integrated Learning*, *8*(1), 55.

Liljegren, E., & Osvalder, A. L. (2004). Cognitive engineering methods as usability evaluation tools for medical equipment. *International Journal of Industrial Ergonomics*, *34*(1), 49-62.

Mount, W. M., Tuček, D. C., & Abbass, H. A. (2012, November). Psychophysiological evaluation of task complexity and cognitive performance in a human computer interface experiment. In *International Conference on Neural Information Processing* (pp. 600-607). Springer, Berlin, Heidelberg.

Lee, S., Baik, Y., Nam, K., Ahn, J., Lee, Y., Oh, S., & Kim, K. (2014). Developing a cognitive evaluation method for serious game engineers. *Cluster computing*, *17*(3), 757-766.

Reychav, I., & Wu, D. (2016). The interplay between cognitive task complexity and user interaction in mobile collaborative training. *Computers in Human Behavior*, *62*, 333-345.

Mosaly, P. R., Mazur, L. M., Yu, F., Guo, H., Derek, M., Laidlaw, D. H., ... & Mostafa, J. (2017). Relating task demand, mental effort and task difficulty with physicians' performance during interactions With Electronic Health Records (EHRs). *International Journal of Human–Computer Interaction*, 1-9.

Mazur, L. M., Mosaly, P. R., Moore, C., Comitz, E., Yu, F., Falchook, A. D., ... & Marks, L. B. (2016). Toward a better understanding of task demands, workload, and performance during physician-computer interactions. *Journal of the American Medical Informatics Association*, *23*(6), 1113-1120.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, *22*(1), 27-57.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of management review*, *13*(1), 40-52.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Netherlands: Groningen University, Traffic Research Center.

Bedny, G. Z., Karwowski, W., & Bedny, I. S. (2012). Complexity evaluation of computer-based tasks. *International Journal of Human-Computer Interaction*, *28*(4), 236-257.

Kantowitz, B. H. (1987). 3. Mental workload. In *Advances in psychology* (Vol. 47, pp. 81-121). North-Holland.

Boehm-Davis, D. A., Casali, J. G., Kleiner, B. M., Lancaster, J. A., Saleem, J. J., & Wochinger, K. (2007). Pilot performance, strategy, and workload while executing approaches at steep angles and with lower landing minima. *Human factors*, *49*(5), 759-772.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language learning*, *51*(3), 401-436.

Macaulay, M. (2004). The speed of mouse-click as a measure of anxiety during human-computer interaction. *Behaviour & Information Technology*, *23*(6), 427-433.

Lin, T., Imamiya, A., & Mao, X. (2008). Using multiple data sources to get closer insights into user cost and task performance. *Interacting with Computers*, *20*(3), 364-374.

Arguello, J. (2014, April). Predicting search task difficulty. In *European Conference on Information Retrieval* (pp. 88-99). Springer, Cham.

De Alwis, B., Murphy, G. C., & Minto, S. (2008, May). Creating a cognitive metric of programming task difficulty. In *Proceedings of the 2008 international workshop on Cooperative and human aspects of software engineering* (pp. 29-32). ACM.

Leporini, B., & Paternò, F. (2008). Applying web usability criteria for vision-impaired users: does it really improve task performance?. *Intl. Journal of Human–Computer Interaction*, *24*(1), 17-47.

Wycislik, L., & Warchal, L. (2014). A performance comparison of several common computation tasks used in social network analysis performed on graph and relational databases. In *Man-Machine Interactions 3* (pp. 651-659). Springer, Cham.

Goncalves, J., Sarsenbayeva, Z., van Berkel, N., Luo, C., Hosio, S., Risanen, S., ... & Kostakos, V. (2016). Tapping task performance on smartphones in cold temperature. *Interacting with Computers*, *29*(3), 355-367.

Huber, V. L. (1985). Effects of task difficulty, goal setting, and strategy on performance of a heuristic task. *Journal of Applied Psychology*, *70*(3), 492.

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types?. *International Journal of Psychophysiology*, 83(3), 269-275.

Scasserra, D. (2008). The influence of perceived task difficulty on task performance.

Bellamy, R., John, B., & Kogan, S. (2011, May). Deploying CogTool: integrating quantitative usability assessment into real-world software development. In *Proceedings of the 33rd International Conference on Software Engineering*(pp. 691-700). ACM.

Ferré, X., Juristo, N., Windl, H., & Constantine, L. (2001). Usability basics for software developers. *IEEE software*, *18*(1), 22-29.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. The journal of the learning sciences, 4(2), 167-207.

Deng, L., Wang, G., & Yu, S. (2016). Layout design of human-machine interaction interface of cabin based on cognitive ergonomics and GA-ACA. Computational intelligence and neuroscience, 2016, 2.

Coury, B. G., & Strauss, R. A. (1998, October). Cognitive models in user interface design. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 42, No. 3, pp. 325-329). Sage CA: Los Angeles, CA: SAGE Publications.

Niemann, J., Presse, V., Reissland, J., & Naumann, A. (2010). Developing a user-centered mobile service interface based on a cognitive model of attention allocation. Human-Computer Interaction, 50-57.

Harris, B. N., John, B. E., & Brezin, J. (2010, April). Human performance modeling for all: importing UI prototypes into CogTool. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3481-3486). ACM.

Sherry, L., Medina, M., Feary, M., & Otiker, J. (2008, May). Automated tool for task analysis of NextGen automation. In *Integrated Communications, Navigation and Surveillance Conference, 2008. ICNS 2008* (pp. 1-9). IEEE.

Xian, H., Jin, L., Hou, H., Niu, Q., & Lv, H. (2014). Analyzing effects of pressing radio button on driver's visual cognition. In *Foundations and Practical Applications of Cognitive Systems and Information Processing* (pp. 69-78). Springer, Berlin, Heidelberg.

Suzuki, S., Nakao, Y., Asahi, T., Bellotti, V., Yee, N., & Fukuzumi, S. I. (2009, July). Empirical comparison of task completion time between mobile phone models with matched interaction sequences. In *International Conference on Human-Computer Interaction* (pp. 114-122). Springer, Berlin, Heidelberg.

Bannon, L. J. (1995). From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In *Readings in Human–Computer Interaction* (pp. 205-214).

Card, S. (1983). The psychology of human-computer interaction. *Lea*.