MIS Quarterly

RESEARCH ARTICLE

USING EYE TRACKING TO EXPOSE COGNITIVE PROCESSES IN UNDERSTANDING CONCEPTUAL MODELS

Palash Bera
Operations and Information Technology Management, Chaifetz School of Business, Saint Louis University,
St. Louis, MO 63103 U.S.A. {palash.bera@slu.edu}

Pnina Soffer
Department of Information Systems, Faculty of Social Sciences, University of Haifa, Carmel Mount,
Haifa, ISRAEL, 31905 {spnina@is.haifa.ac.il}

Jeffrey Parsons
Faculty of Business Administration, Memorial University of Newfoundland,
St. John's, NL A1C 5S7 CANADA {jeffreyp@mun.ca}

Appendix A

Measuring Run Counts

To calculate the total run count, the total number of times the task-relevant areas were entered and exited was calculated, whereas for AOI run count, only the saccadic movements among the relevant areas were considered. Consider Figure A1 below. In it, there are nine saccadic movements where blue indicates fixations in nonrelevant areas and red indicates fixations in relevant areas. The red fixations are found in three relevant AOIs. In this figure, the total run count is five—arrows 2, 3, 5, 7, and 9. Arrows 1 and 8 are not considered as they are between fixations in nonrelevant areas. Arrows 4 and 6 are not considered as they are within the same relevant AOIs (as the saccades stay inside the AOI). The AOI run count is a subset of this total run count and in this figure the AOI run count is two (arrows 3 and 5). The arrows 2, 7, and 9 are ignored as they originate or end at nonrelevant AOIs.

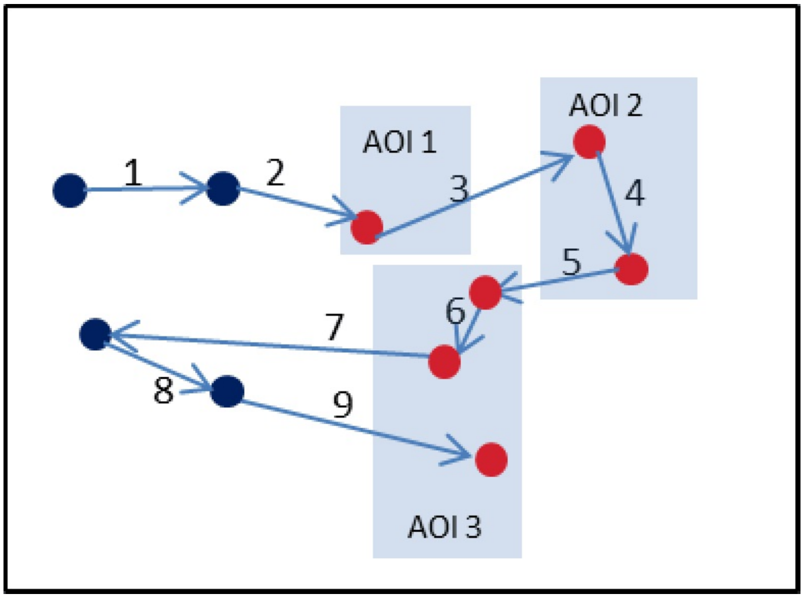


Figure A1. Example Illustrating Run Count Measurement

Appendix B

Coding of Problem Solving Task and Performance Analysis

The dependent variable for measuring domain understanding of the models is operationalized by the number of correct problem solving scores. To measure the number of correct responses, a set of possible correct responses was created by the first researcher. These responses were developed by reading domain descriptions. A university hospital nurse and an ethics manager working in a university were consulted on developing the set of correct responses. Two graduate students unaware of the objective of the study used this set as guidance to mark the participant responses. Table B1 shows the possible set of responses for the admission domain. Note that this set is not exhaustive and the coders used the table as guidance. A participant could provide multiple responses and each response was classified as correct or incorrect by the coder. The total number of correct responses was calculated for each problem solving task. It was possible for a participant to get zero if all the responses were incorrect. If more than one response was correct, then the total number of correct responses was counted. Because the responses were subjective in nature, we used two coders who independently coded the responses. The inter-rater reliability for Study 1 was 89% and 88% for the admission domain and the ethics domain respectively. Given the high inter-rater reliability, the responses coded by Coder 1 were used in the study.

Table B1. Sample Correct Responses of the Problem Solving Tasks for the Admission Domain

Problem Solving Tasks	Possible Correct Responses
After stabilizing a patient, if his/her information is not provided to the *admission department* then what problems might arise?	• The admission department will not know how critical the patient is • The admission department may not know where to send the patient for operation • Difficulty in identifying whether the patient can be discharges • Difficulty in room assignment • No patient history is created • Billing cannot be done • Patient cannot be discharged
What will happen if patients are diagnosed immediately after arrival?	• There may be a misdiagnosis because the patient's body vitals were not checked and the patient was not stabilized • The patient may be misdiagnosed • The condition of the patient may worsen • The patient may die as he/she was not stabilized • No patient record is created

Prior to the analysis of ANCOVA, an analysis was done (Table B2) to determine whether the groups differed in terms of familiarity with the domain and with modeling. No differences were found.

Table B2. Domain and Modeling Familiarity Analysis for BPM Study

	EPC	EPC-H	BPMN	t-value EPC-H vs. EPC	p-value EPC-H vs. EPC	t-value BPMN vs. EPC	p-value BPMN vs. EPC
	M (SD)	M (SD)	M (SD)				
Domain Knowledge	3.70 (0.38)	3.82 (0.35)	3.65 (0.37)	-0.87	0.19	0.36	0.36
Modeling Knowledge	4.90 (0.54)	4.80 (0.59)	4.87 (0.52)	0.48	0.32	0.17	0.43

ANCOVA is performed by aggregating the correct scores of the four problem solving questions (two for each domain) (Table 4, body of paper). Performance on the problem solving tasks is presented in Table B3. The results indicate that the effect of BPMN was stronger than the effect of EPC-H models.

Table B3. Analysis of Problem Solving Tasks for BPM Study

Treatment	PS mean (EPC)	PS SD (EPC)	PS mean (EPC-H)	PS SD (EPC-H)	PS mean (BPMN)	PS SD (BPMN)	F - value	P-value
EPC vs. EPC-H	1.50	0.42	1.76	0.31			2.17	0.08
EPC vs. BPMN	1.50	0.42			2.18	0.62	12.20	0.001
EPC-H vs. BPMN			1.76	0.31	2.18	0.62	3.13	0.01

PS mean = Average correct number of problem solving tasks; Domain knowledge and modeling knowledge were used as control variables.

Appendix C

Detailed Experimental Procedure

Study 1

Eye movements were recorded using *EyeLink 1000* eye tracking software. Participants were seated 70 cm from the display monitor (resolution of 1600 × 1200 and refresh rate of 85 Hz). A chin rest was used for head support. The EyeLink 1000 eye tracker records a minimum fixation of 4 milliseconds. The average percentage of rejected observations in the first study was 10.47%. This means that slightly more than 10% of the eye observations were not captured by the eye tracking device. After calibration, gaze-position error was less than 0.5 degree and was sampled at 1000 Hz. Once participants' eyes were calibrated, they were shown the problem solving questions one at a time and asked to read the questions carefully. Following this, they pressed a joystick to see the script (based on the group to which they were assigned) and verbalized the answers. Participants were asked to verbalize rather than write the answers as writing would have taken their eyes off the screen and their eye movements would not be captured properly. This strategy of verbally answering the questions so that the users do not need to type the answer and get distracted was used in Kagdi et al. (2007). If a participant forgot the question, then a research assistant repeated the question. When participants finished answering a question, they pressed the joy stick again to see the next problem solving question. To increase the generalizability, participants answered problem solving questions twice using scripts developed from two domains. The study took approximately 20 minutes to complete.

Study 2

The Tobii Pro X3-120 eye tracker was used in the second study. It has a sampling rate of 120 Hz and provides flexibility for participants to move during the experiment (up to 80 cm). A web-based experiment was setup. Participants answered the problem solving questions in three steps (Table C1). In the first step, participants read the question and clicked on the continue button to view the model. In the second step, while viewing the model, participants verbalized their thought process as they answered the question. The eye tracker recorded the participants' voice. The tasks performed by the participants were recorded by the eye tracker and were available in video. If participants were silent for 10 seconds, a research assistant would prompt the participant to verbalize his/her thought processes. The question and the model were not placed in the same screen to avoid having participants' attention distributed between the diagram and the question. After viewing the model, participants could go back to the question by clicking on the "back to the question" button or continue answering the question by clicking the "continue" button and typing the response at the next screen (step 3).

Fixations and durations were accumulated over visits to the model. This means if a participant visited the model two times by interacting among the three steps (e.g., clicking back to the question from the model and then again visiting the model) the total fixation count is the sum of all the fixations when participants visited the model.[1] The average percentage of rejected observations in this study was 10.35%. Participants averaged 29 minutes to complete the study.

Reference

Kagdi, J., Yusuf, H., and Maletic, J. I. 2007. "On Using Eye Tracking in Empirical Assessment of Software Visualization," in *Proceedings of the 1st ACM International Workshop on Empirical Assessment of Software Engineering Languages and Technologies,* Atlanta, pp. 21-22.

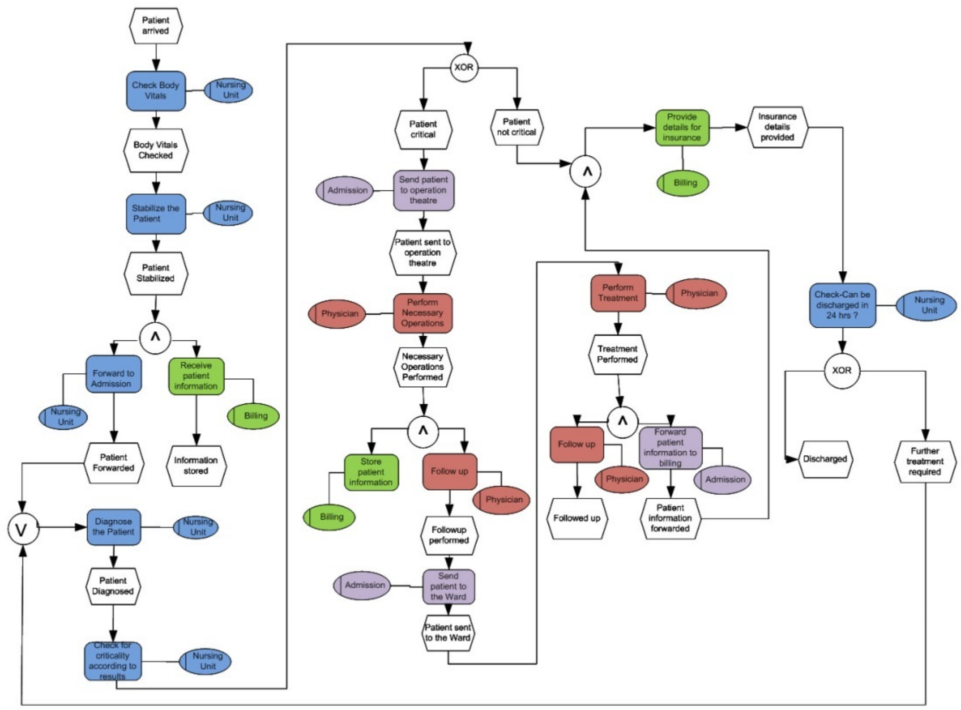
[1]In the first study, subjects could not maintain this interactivity, as the eye tracker (Eyelink 1000) did not allow subjects to go back and forth with the model and question (as this would require recalibration). Therefore, in the previous experiment, subjects read the question and then visited the model and then verbalized the answer in sequence. If subjects forgot the question, a research assistant provided a reminder.

Step 1: Participant exposed to the question

After stabilizing a patient, if his/her information is not provided to the admission department then what problems might arise?

Continue

Step 2: Participant exposed to the model



Back to the Question Continue

Step 3: Participant provides the answer in the following space

Back to the Model Save Continue

Figure C1. Steps in Answering Problem Solving Questions (Study 2)

Appendix D

Models Used in the Experiments

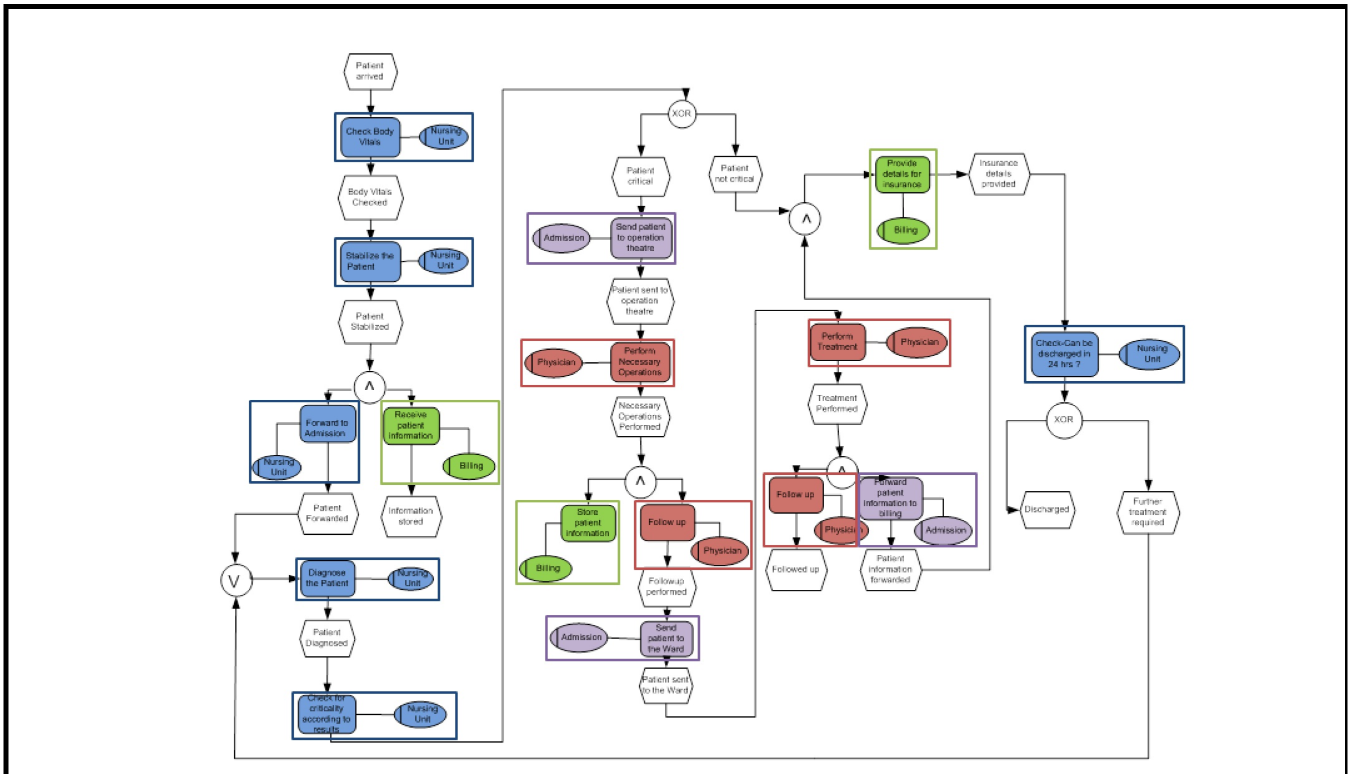


Figure D1. EPC-H Model: Hospital Treatment Domain

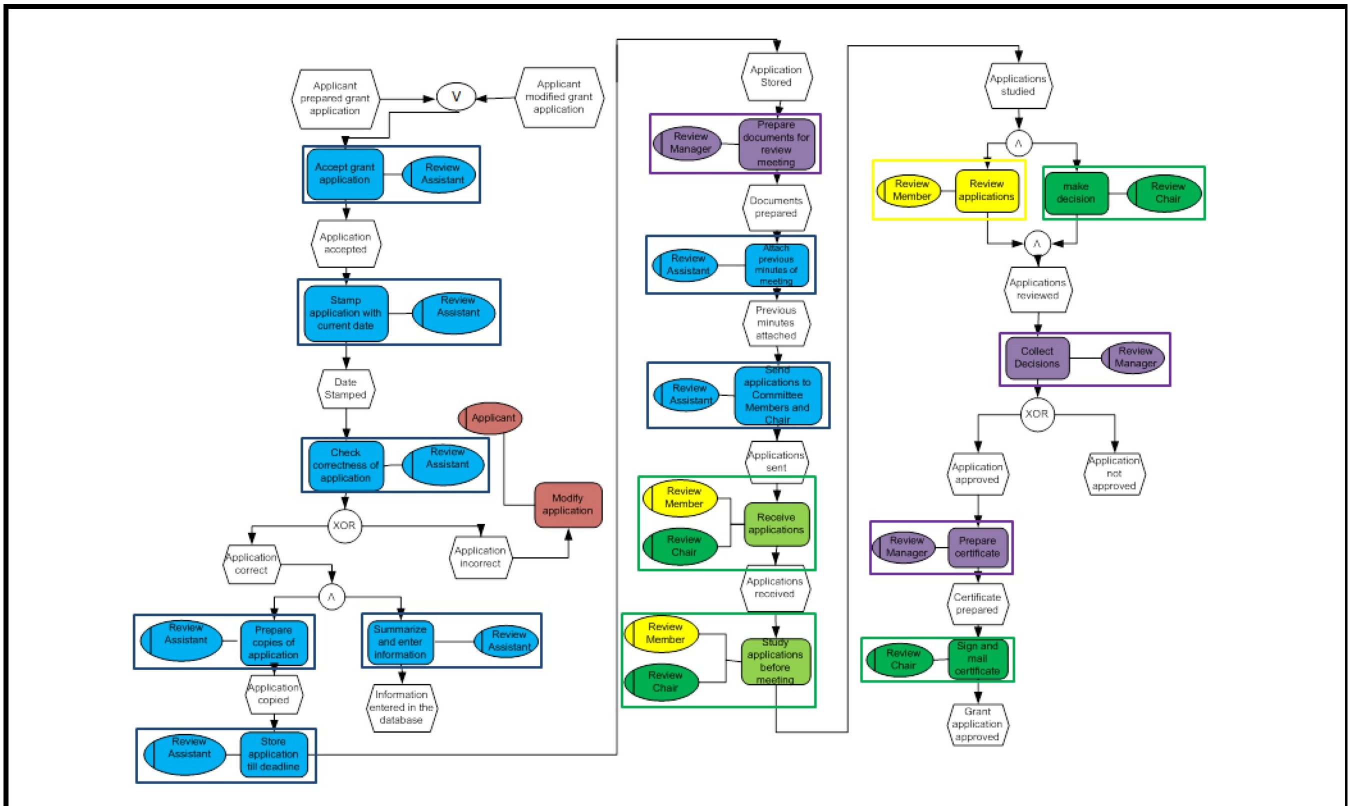


Figure D2. EPC-H Model: Grant Review Domain

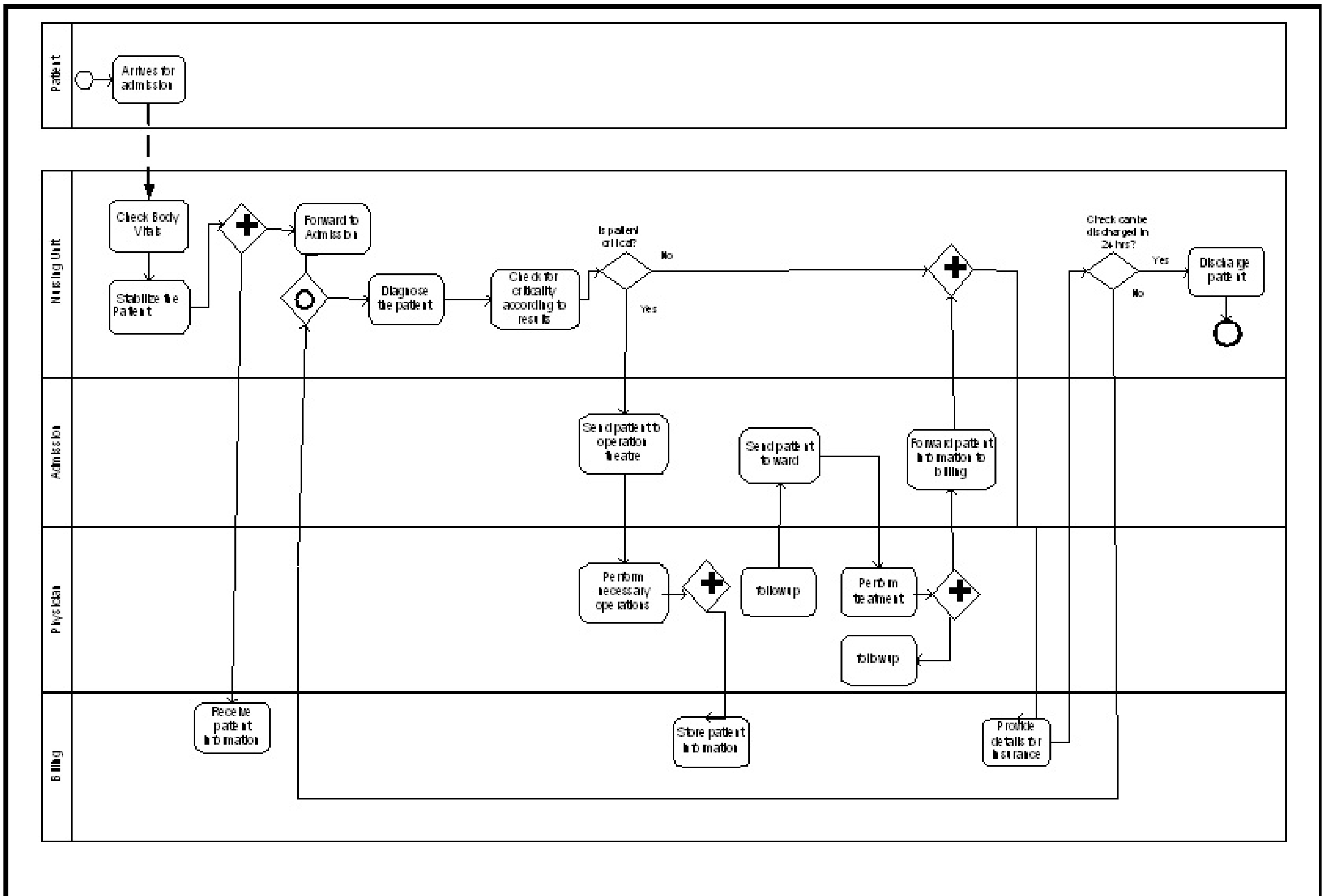


Figure D3. BPMN Model: Hospital Treatment Domain

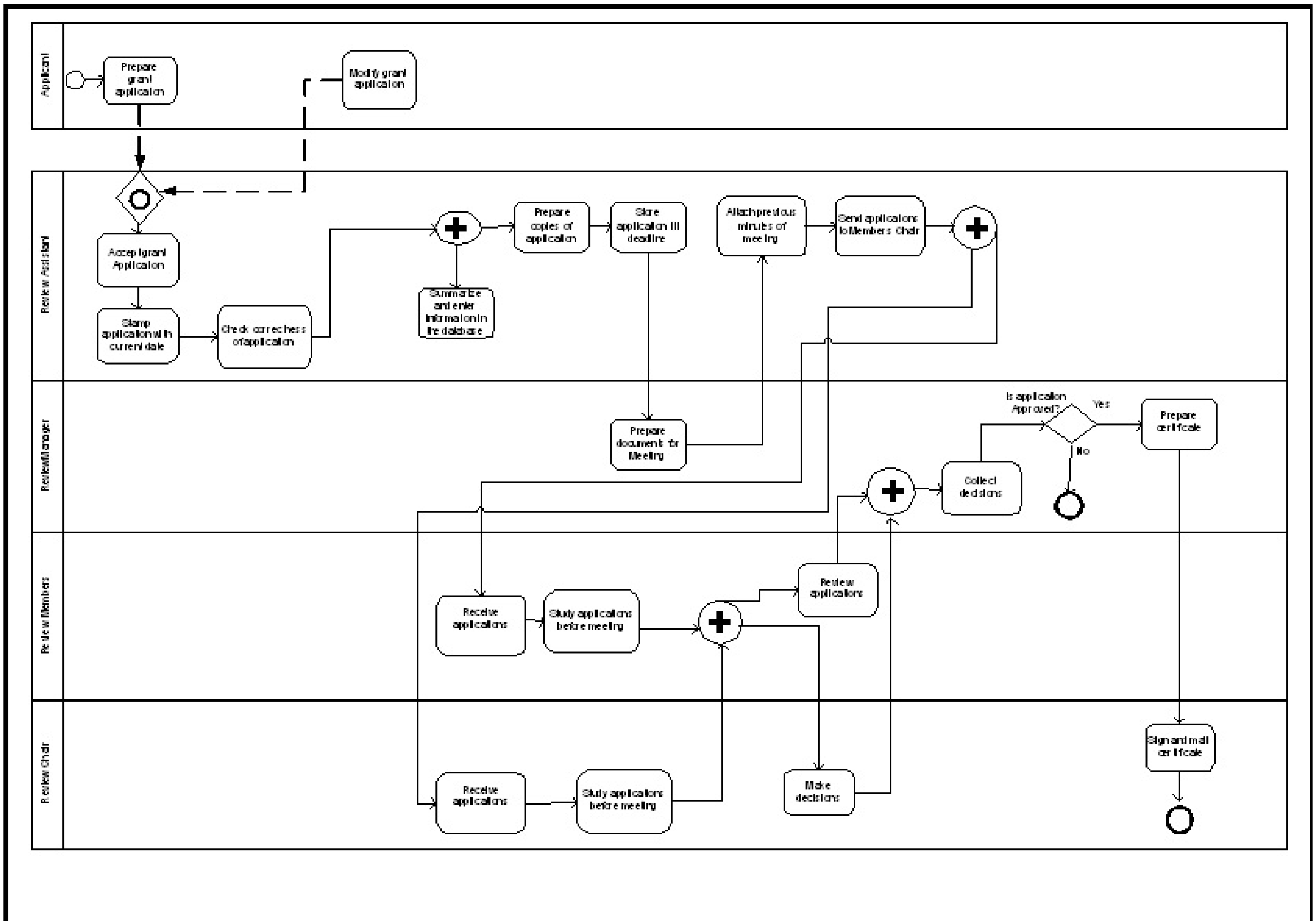


Figure D4. BPMN Model: Grant Review Domain

Appendix E

Procedures for Selecting Task-Relevant AOIs

Task-relevant AOIs depend on the problem solving tasks. A set of steps was followed to select the task-relevant AOIs. These steps and examples of the problem solving tasks are provided below. It is to be noted that the number of possible answers for each task is large and there could be many variations of these answers.

Table E1. Procedure for Selecting Task-Relevant AOIs for the Hospital Treatment Domain

Step 1	Step 2	Step 3
Develop possible correct answers for each task	Analyze the answers of the tasks to identify who performs these tasks	Set the role and the corresponding tasks as task-relevant areas
After stabilizing a patient, if his/her information is not provided to the admission department then what problems might arise?		
• The *admission department* will not know how critical the patient is • The *admission department* may not know where to send the patient for operation • *Admission department* cannot forward patient information to Billing • *Admission department* will have difficulty in assigning ward to the patient	Admission department performs these tasks	Admission department and the corresponding tasks performed by this department are the task-relevant areas.
What will happen if patients are diagnosed immediately after arrival?		
• Misdiagnosis by *nurse* because the patient's body vitals were not checked • Misdiagnosis by *nurse* because the patient was not stabilized • Misdiagnosis by *nurse* may lead to worsen patient's condition • *Nurse* does not intimate the admission department	Nursing unit perform these tasks	Nursing unit and the corresponding tasks performed by this department are the task-relevant areas.

Appendix F

Manipulation Checks

Analysis of Task Nonrelevant Areas

In this manipulation check, an analysis of eye metrics on task nonrelevant areas was performed. This was done to ensure that the difference in eye metrics is valid only on task-relevant areas and not on nonrelevant areas. Table F1 supports this for Study 1.

Table F1. Eye Metric for Task Nonrelevant Areas for Grant Review Domain of First Problem Solving Task

	EPC	EPC-H	BPMN	t-value EPC-H vs. EPC	p-value EPC-H vs. EPC	t-value BPMN vs. EPC	p-value BPMN vs. EPC
Area: Review Manager	**M (SD)**	**M (SD)**	**M (SD)**				
% of duration	9.1 (4.36)	8.89 (3.31)	10.47 (4.48)	0.15	0.44	-0.84	0.21
Run counts	9.40 (1.18)	9.46 (1.30)	10.26 (1.53)	-0.15	0.44	-1.73	0.05
Area: Review Assistant							
% of duration	27.37 (9.38)	28.32 (0.08)	30.81 (5.68)	-0.29	0.38	-1.21	0.12
Run counts	11.80 (1.86)	12.00 (2.45)	12.26 (1.98)	-0.25	0.40	-0.66	0.26

Regression analysis was performed to test whether the percent of time on nonrelevant AOI's contributed significantly to problem solving performance. For this purpose, the percentage time for nonrelevant AOI for the first problem solving on grant review domain (Study 1) was calculated and used as independent variable (Table F2).

Table F2. Regression Analysis Using Percentage of Fixation Time on Nonrelevant Areas as Independent Variable (For the First Problem Solving on Grant Review Domain)

Independent Variable	B	t	P	Adj. R square
Group = BPMN				
Constant	2.86	3.86	0.002	0.05
% of time spent on the AOI's denoting nonrelevant areas	-4.11	-1.30	0.19	
Group = EPC-Highlighted				
Constant	2.51	3.09	0.00	0.02
% of time spent on the AOI's denoting nonrelevant areas	-2.44	-1.14	0.25	

Analysis of Eye Fixations of the Central Regions of Models

It might be possible that users tend to look at the center of the models and the results of the experiment can be explained by users' tendency to focus on the center of the models rather than the task-relevant areas.

To test this proposition, we created two zones (AOIs) at the center of the models. One zone covered 5% of the entire area (pixel size 38,645) and the other covered 10% of the entire area (pixel size 77,292). These areas are shaded in blue in Figure G1. We performed fixation analysis in these zones for all the business process model types (EPC, EPC-H, and BPMN). For this analysis, from the admission domain, we selected the following question: "What will happen if patients are *diagnosed* immediately after arrival?"

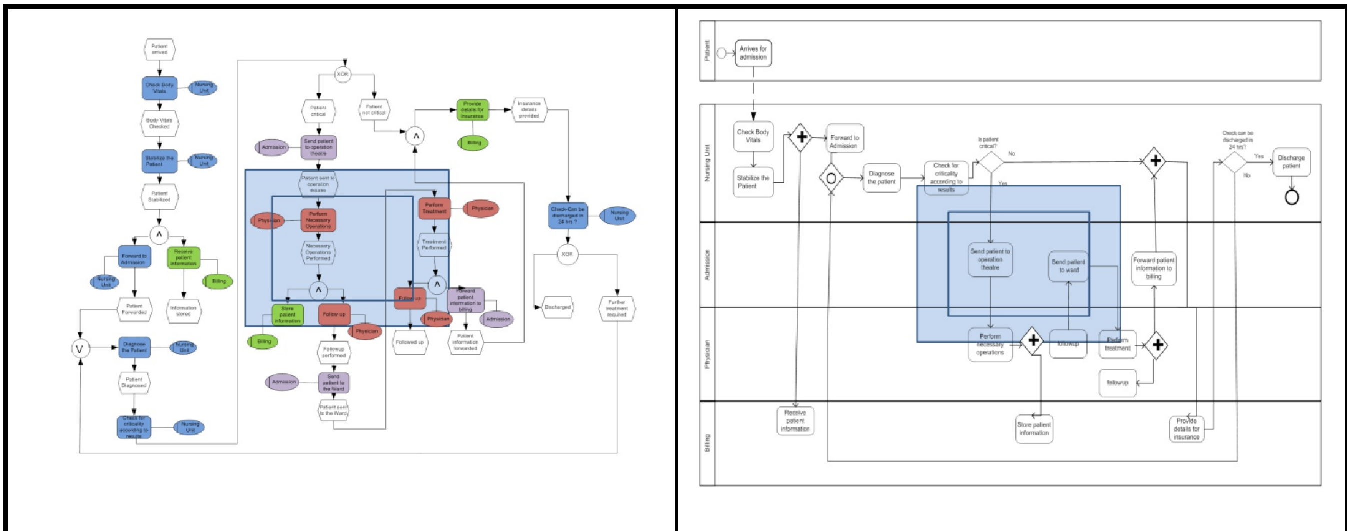


Figure F1. Center Regions of the Process Models

The analysis of the fixation numbers is shown in the table below.

Table F3. Fixation Analysis of the Center Regions of the Process Models

	EPC M (SD)	EPC-H M (SD)	BPMN M (SD)	EPC-H vs. EPC t score (p value)	BPMN vs. EPC-H t score (p value)
5% AOI	3.2 (1.2)	2.9 (0.8)	2.67 (0.6)	0.71 (0.28)	1.02 (0.16)
10% AOI	10.6 (1.2)	10.1 (0.9)	15.7 (2.3)	1.12 (0.13)	8.93 (0.00)

When the area was selected as 5% of the center of the models, there was no statistical difference in the fixations among the models. However, when the area selected was 10% of the center of the models, there was statistical difference of the fixations between BPMN and EPC-H. This can be explained as part of the task-relevant area (i.e., the nursing swimlane) was present in the BPMN when 10% of the area is considered. These relevant areas (blue activities and roles) are not present in 10% of the center of EPC-H. The above results suggest that the results in the main paper cannot be explained by considering that participants look only at the central area of the model. Rather as the main analysis indicate, the task-relevant areas are responsible for the significant differences in the results.