Eye Tracking Experiments in Business Process Modeling: Agenda Setting and Proof of Concept

Frank Hogrebe[1], Nick Gehrke[2], Markus Nüttgens[3]

[1]Hessische Hochschule für Polizei und Verwaltung,
Fachbereich Verwaltung, D- 65197 Wiesbaden

[2]Nordakademie, Hochschule der Wirtschaft,
Köllner Chaussee 11, D-25337 Elmshor

[3]Universität Hamburg, Lehrstuhl für Wirtschaftsinformatik,
Von-Melle-Park 5, D-20146 Hamburg

Abstract. For almost all applications there is a need to judge business process models by user´s perspective. This paper presents first facts and findings how the eye tracking method can contribute to a deeper empirical understanding and evaluation of user satisfaction in business process modeling. The method of eye tracking is used in order to check subjective perceptions of users through objective measurement. The experimental investigation is done using two variants of the widespread business process modelling notation "Event-driven Process Chain (EPC)".

Keywords: eye tracking, user satisfaction, event-driven process chain (EPC)

1 Introduction

User satisfaction is one of the most widely researched topics in the information systems field [ANC08], [A03]. There are strong relationships between user satisfaction and intended use or actual use of information systems [ANC08, p.44]. The paper focuses on user-based requirements in business process modeling. In order to be able to judge subjective appraisals of test users and real-life users through objective measurements, the method of eye tracking is used in our experiment. For the evaluation of requirements of user satisfaction in business process modeling we focused on, a research framework is developed and used as part of an experiment. Due to the fact, that event-driven process chains (EPC) are still establish as a widespread notation in business process modeling, we used two variants of the EPC to show the prospects and limitations of our approach.

2 Eye Tracking Technique

Eye tracking is a technique with which the course of a person's vision when viewing an object, model or an application (investigation objects) can be measured [RP07]. In the taxonomy of research methods, eye tracking is assigned to the user-oriented

183

methods [Y03, p.125]. In the combination of eye tracking and surveys, survey results are supplemented through quantitative eye tracking data. The test person comments on his or her activities, feelings and thoughts in the use of the object of investigation (qualitative aspect). This methodological approach is to be transferred in this investigation to the measurement and analysis of the user satisfaction in business process modeling.

3 User Satisfaction Eye Tracking Experiment

3.1 Research Framework and Hypotheses

The research framework has the aim, to create a framework for designing experiments measuring the user satisfaction by combining surveys and eye tracking methods. Our aim in this experiment is to compare the user satisfaction regarding two semiformal modeling notations as the test case. We developed the following basic hypothesis (H1 – H3) for this comparison:

H1: Seven user-related requirements deducted from our literature review are acknowledged by users (editors and recipients).

H2: Seven requirements are appropriate criteria to check subjective assessments of users with objective measurements of experiments.

H3: Eye tracking is an appropriate method for measuring and judging user-related requirements related to the user satisfaction in business process modeling.

Table 1. Research Framework for the empirical study

Requirements	*Subjective Perception (Assessment of users in the survey)*	*Measurement (Experiment with users)*
Understanding	Users have been asked: the variant of the modeling language is easy to understand?	A graphical representation of a process is shown to the users and specific questions are asked. Number of false answers is counted.
Completeness	Users have been asked: the variant of the modeling language is complete?	Users have to model a pre-defined process. It is noted if users are capable to model the process (Editors)
Easy to use	Users have been asked: the variant of the modeling language is easy to use?	Users have to model a pre-defined process. Number of modeling errors is counted (editors).
Usefulness	Users have been asked: the variant of the modeling language is useful?	In a given graphical representation of a process users need to find a pre-defined object. The following is measured: *Numbers of fixations with eye tracking methods *Length of fixation with eye tracking methods. The more and the longer fixations are, the more the effort and the worst the balance between effort and utility.
Timeliness	Users have been asked: the variant of the modeling language is time saving?	Users have to model a pre-defined process. The time needed is recorded (editors).
Flexibility	Users have been asked: the variant of the modeling language is flexible?	No measurement since users considered this requirement to be less important (rank 6 of 7).
Accuracy	Users have been asked: the variant of the modeling language is accurate?	No measurement since users considered this requirement to be less important (rank 7 of 7).

Due to the fact, that event-driven process chains (EPC) are still establish as a widespread notation in business process modeling, we used two variants of the EPC to show the prospects and limitations of our approach. In our experiment two variants of the event-driven process Chain (EPC) are used for validating the seven requirements:

184

the extended EPC (eEPC) and the object oriented EPC (oEPC) [SNZ97]. Our investigation is based on 24 users (12 editors and 12 recipients of models). In each group 6 users have been students and 6 have been employees of a public administration. We use as our test case graphical representation of public administration processes, because there are good experiences about the notations in this area (see survey, [anonymous]). The users have been asked to subjectively assess the modeling languages regarding the seven requirements. Furthermore, we did experiments with the users like measuring time to model a process (for editing users) or counting the number of errors a user was able to find in a model (for recipient users). Another example is measuring the length of fixations during solving a specific task by using eye tracking (editors and recipients). All requirements have been mapped in order to measure how the requirements are achieved by the two variants of the modeling language. After that we compare the results of the subjective assessment and the objective experiment for both, requirements and variants. As a result of our investigation, the most important requirements were 'understanding' (editors: 8, recipients: 11), 'completeness' (editors: 7, recipients: 6) and 'easy to use' (editors: 5, recipients: 7). In this context, we asked if a user missed an important requirement but none of the 24 users answered 'yes'. Thus, we conclude that the seven requirements used in our investigation are sufficient. Hypothesis 1 can be considered to be confirmed in the context of our experiment. Table 4 depicts the research framework for our empirical study of the user-related requirements. The column 'measurement' describes the experiment used for the specific requirement.

3.2 Statistical Analysis

In the following section we analyze the data gathered during our survey regarding the requirements. For all requirements interviewees could answer with the following options: I fully agree (=4), I agree partially (=3), I doubt to agree (=2), I do not agree at all (=1) for each variant of the modeling language (eEPC and oEPC). Due to missing answers the sample size could be less than 12. Table 2 depicts basic statistic ratios regarding the assessment of the requirements by the interviewees during the survey. Obviously, all requirements – except the requirement 'accuracy'– got a better average assessment regarding the oEPC. The requirement 'accuracy' has been assessed equally. The correlation of the assessments between eEPC and oEPC (right table) only reveal a significant (negative) relationship [> (-) 0.5] for the requirement 'time saving'. That means interviewees assessing one modeling language as time saving tend to assess the other language as less time saving. In the following we want to verify if differences in the assessments of the interviewees are significant regarding each requirement. For this reason we conducted a t-Test testing for different means in dependent samples. Table 3 shows the results of the statistic test.

The column 'Sig. (2-tailed)' shows the significance level of the difference of means regarding the assessment of the interviewees for each requirement (meaning H0: eEPC=oEPC; H1: oEPC≠eEPC). But we want to focus on the hypothesis that one modeling language is better than the other one. Thus, we also test H0: eEPC=oEPC; H1: oEPC>eEPC to validate if the oEPC language dominates the eEPC variant (see

185

column ‘Sig. (1-tailed)’). Since we test H0: eEPC=oEPC; H1: oEPC>eEPC the confidence level is always half the confidence level of column ‘Sig. (2-tailed)’. As a result, we can see that the requirement ‘flexibility’ has been considered to be better in respect to the oEPC language on a 95% confidence level whereas the requirements ‘easy to use’ and ‘usefulness’ are significant on a 90% level. In the same manner we analyzed the data of the survey we now want to consider the results of the experiments as depicted in table 4. Please note that Pair 4 and 5 both are linked to the requirement ‘usefulness’ since the data represents on the one hand the number of fixations and on the other hand the length of the fixations.

Table 2. Statistic ratios regarding the assessment of requirements by interviewees

Paired Samples Statistics

	Mean	N	Std De iation	Std Error Mean
air 1 E C easy-to-use	00	12	000	000
oE C easy-to-use		12	8	22
air 2 E C Flexi ility	2	11	8	2
oE C Flexi ility		11		20
air E C accuracy	2	11	8	2
oE C accuracy	2	11		19
air E C understandin	2	12	1	1 9
oE C understandin		12	92	1 2
air E C completeness	1	12	8	2 1
oE C completeness		12	8	22
air E C timeliness	2	11	1 009	0
oE C timeliness	18	11	1	22
air E C usefulness	2 82	11	982	29
oE C usefulness		11		20

Paired Samples Correlations

	N	Corr	Si
air 1 E C easy-to-use oE C easy-to-use	12		
air 2 E C flexi ility oE C flexi ility	11	9	2 0
air E C accuracy oE C accuracy	11	0	91
air E C understandin oE C understandin	12	- 120	11
air E C completeness oE C completeness	12		12
air E C timeliness oE C timeliness	11	- 88	0
air E C usefulness oE C usefulness	11	2 1	8

Table 3. t-tests of the assessments of the interviewees for each requirement (different means)

Paired Samples Test

	aired Differences									
				90						
	Mean	Std De iation	Std Error Mean	o er	pper	t	df	Si (2-tailed)	Si (1-tailed)	inner
air 1 E easy-to-use - oE easy-to-use	-	8	22	-	0 0	-1 8	11	1	08	oE
air 2 E flexi ility - oE flexi ility	-	809	2	-1 0 8	- 19	-2 09	10	02	01	oE
air E accuracy - oE accuracy	000	1 000	02	-		000	10	1 000	00	N A
air E understandin - oE understandin	- 2 0		218	- 1	1 1	-1 1 9	11	2	1	oE
air E completeness - oE completeness	- 1	8	2 1	- 99	2	- 92	11	0	2 2	oE
air E timeliness - oE timeliness	-	1 2		-1 1	0	- 9 9	10	0	180	oE
air E usefulness - oE usefulness	-	1 0	12	-1 111	021	-1	10	111	0	oE

Table 4. Statistical ratios of the experiments regarding the requirements

Paired Samples Statistics

	Mean	N	Std De iation	Std Error Mean
air 1 E understandin errors	2 0	12	1 1	
oE understandin errors	2 92	12	1 8	
air 2 E easy-to-use errors	10 08	12	12	
oE easy-to-use errors	2 8	12	09	89
air E timeliness time	9 00	12	2 89	8
oE timeliness time	00	12	2 09	0
air E usefulness Num erOfFixations	29 1	12		18 29
oE usefulness Num erOfFixations	2 2	12	2 8	18 1
air E usefulness en thOfFixations	1 8 92	12	1	10 2
oE usefulness en thOfFixations	1 1 92	12	0 9	1 1

Paired Samples Correlations

	N	Corr	Si
air 1 E understandin errors oE understandin errors	12	- 11	
air 2 E easy-to-use errors oE easy-to-use errors	12	22	0
air E timeliness time	12	2	2
air E usefulness Num erOfFixations oE usefulness Num erOfFixations	12	- 0	8
air E usefulness en thOfFixations oE usefulness en thOfFixations	12	21	2

Regarding table 4 please note that – compared to the assessments of the interviewees – small values are better than big ones (e.g. less errors are better than more errors or less time consuming is better than more time consuming). Obviously, regarding the

186

means the oEPC language performs better for all requirements than the eEPC variant except for the requirement ‘understanding’. Focusing on ‘understanding’ one can see that users answered with fewer errors considering the eEPC. In this context, we reason that eye tracking is an appropriate method to measure and assess requirements for user satisfaction in business process modeling in our experiment. Therefore, Hypothesis 3 can be considered to be confirmed. T-tests regarding the differences of the means of the measurements reveal that the oEPC variant dominates significantly the eEPC variant (Table 5). Only the requirement ‘understanding’ cannot be considered significant. The requirements ‘easy to use’ and ‘timeliness’ are significant on a 95% level (see column Sig. (2-tailed)). The two measurements for ‘usefulness’ are significant only on a 90% level. Table 8 does not contain the requirement ‘completeness’ since all users where able to model the predefined process in the appropriate experiment with the eEPC and the oEPC variant. Thus, the requirement ‘completeness” remains undecided.

Table 5. t-tests of the experiments / measurements for top 5 requirements

Paired Samples Test

		aired Differences									
					90 Confidence Inter al						
		Mean	Std De iation	Std Error Mean	o er	pper	t	df	Si (2-tailed)	Si (1-tailed)	inner
air 1	E understandin errors - oE understandin errors	- 2	2 2		-1 0		-	11		2 9	E
air 2	E easy-to-use errors - oE easy-to-use errors	0	12		1 09	1 91	2 10	11	0	0 0	oE
air	E timeliness time - oE timeliness time	00	2 98	8	1		8	11	01	00	oE
air	E usefulness Num erOfFixations - oE usefulness Num erOfFixations	92	91	2	-	91 8	1	11	12	0 2	oE
air	E usefulness en thOfFixations - oE usefulness en thOfFixations	2 00		1 2	-2 1	1	1	11	1	0	oE

After having analyzed the results of the survey and the experiments separately, we now want to focus on a comparison of both as an overall result. Table 6 shows the results of this comparison.

Table 6. Comparison of survey and results, * 90% significance, ** 95% significance

Requirement	"Winner" survey	"Winner“ experiment	Comment
Understanding	oEPC	eEPC	no significance
Completeness	oEPC	N/A	no significance
Easy to use	oEPC*	oEPC**	significant for oEPC
Usefulness	oEPC*	oEPC*	significant for oEPC
Timeliness	oEPC	oEPC**	Experiment significant for oEPC, Survey not significant for oEPC
Flexibility	oEPC**	No measurement	significant for oEPC, no experiment
Accuracy	N/A	No measurement	no significance

Obviously, a significant result can be detected regarding the requirements ‘easy to use’ and ‘usefulness’. Only significant in one area, but on an average the oEPC dominates regarding the requirement ‘timeliness’. Considering the requirement ‘flexibility’, the oEPC has been significantly assessed better than the eEPC, but we did not conduct an experiment for that requirement. For the requirement ‘accuracy’ we cannot deduct a winner. Furthermore, for the requirements ‘understanding’ and ‘com-

187

pleteness' there is no significant result. To sum up the oEPC seems to dominate most requirements (often significant). The eEPC variant succumbs or is equal to the oEPC variant. As a result, the literature based common user related requirements are appropriate to check subjective assessments of users in a survey with objective experiments. Therefore, hypothesis 2 can be considered to be confirmed regarding our experiment.

4 Summary of the results

The eye tracking method was used for first time in the measurement and assessment of the user satisfaction in business process modeling. In the application case, the eye tracking method was used to visualise objective measurement and progression data, in addition to the survey techniques. The graphical depictions from the eye tracking (heatmaps) were able to visually support the investigation results. The research framework was used on the basis of extended and object-oriented event-driven process chains (eEPC and oEPC). We used always the same modeling tool to minimize external effects when testing our framework (bflow Toolbox [BT10]). To what extent investigations on the basis of other semi-formal modelling languages lead to other results remains reserved for other investigations. The hypotheses (H1 to H3) could be confirmed on the basis of two variants of the event-driven process chain (EPC). To what extent investigations on the basis of other semi-formal modelling languages in the field of business process modeling lead to other results remains to be validated in follow-up investigations. The initial results of this investigation are interesting, in particular with regard to the options to visualise measurement results in the representation of heatmaps. Here, however, further investigations should follow in order to be able to validate the relevance of the eye tracking method for investigations in the research area of information systems moreover.

References

[A03] Aladwani, A.M.: A Deeper Look at the Attitude-Behavior Consistency Assumption in Information Systems Satisfaction Research. In: The Journal of Computer Information Systems (44:1), (2003), pp. 57--63

[ANC08] Au, N., Ngai, E.W.T., Cheng, T.C.E.: Extending the Understanding of End User Information Systems Satisfaction Formation: An Equitable Needs Fulfillment Model Approach. In: MIS Quarterly Vol. 32 (2008), 1, pp.43—66

[BT10] Bflow Toolbox. www.bflow.org. Accessed 2011-07-31

[RP07] Rosen, D.E., Purinton, E.: Website Design: Viewing the Web as a Cognitive Landscape. In: Journal of Business Research, 57 (2007) 7, pp.787--794

[SNZ97] Scheer, A.W., Nüttgens, M., Zimmermann, V.: Objektorientierte Ereignisgesteuerte Prozeßketten (oEPK) – Methode und Anwendung (in german). In: working paper vol. 141, institute of information systems, (1997), Saarbrücken (germany)

[Y03] Yom, M.: Web usability von Online-Shops. Better-Solutions-Verlag Gierspeck, Göttingen (2003)

188