Study Design for "**User Interface Evaluation**"

By

Jay Joshi (200440993) jjm376@uregina.ca

Of

Faculty of Graduate Studies and Research

University of Regina

Context / Research Motivation: Task complexity is one of the characteristics which have received significant attention in many pieces of literature over the years. An important area of Interface design evaluation is to examine how task characteristics affect user interaction and outcomes. One way of defining task complexity is based on prior-determinability of tasks- Task determinability can be defined based on degree of uncertainty associated with outcomes of task and the search process involved in completing the tasks. A prior determinable task has lower degree of uncertainty about outcomes of tasks and search process involved in finding relevant information. A user interface is responsible for supporting two-way communication between the IR system and its user. For example, a user communicates with an IR system interface by specifying query terms related to information need in the search box. An IR system communicates the relevancy of a system suggested documents to the user. Representation of "relevance" of the system suggested documents have received a significant amount of concern in the last two decades. Most of the times, a user on the way of the information search process is not aware of the documents required to complete the tasks. As a result, this imposes a challenge for the IR system interface design to assist the user in completing an uncertain task especially if the user has to traverse multiple pages to seek the relevant information. Hence, evaluation of alternative arrangements of interface design for a determinable and indeterminable task is important criteria for research. Therefore, an experimental interface (Klink Search interface) is derived from baseline interface (IEEE search interface) with modifications in interface design which supports system-user communication in different way. The main purpose of this research is to evaluate how much an experimental interface in assist users in completing uncertain tasks compared to baseline interface. Hence, we evaluate differences in user interaction, user experience by studying participants on both experimental and baseline interface.

Research Questions

RQ1: What is the effect of task determinability on user interaction with experimental interface compared to baseline interface?

Hypothesis 1:

- 1a: If a task is specified (S) task, study proposes null hypothesis. No major differences will be observed in users interacting with experimental interface and baseline interface because pre-determinable tasks are easy to achieve and users will be able to complete such tasks easily on both the interfaces in similar manner.

- 1b: If a task is unspecified (U) task, study proposes alternative hypothesis i.e. User will interact more with baseline interface than experimental interface. Since experimental interface supports evaluation of search results with additional features like (list of keywords), users will need to spend less effort and time to complete open-ended tasks on experimental interface than baseline interface. So, users will have to spend more effort, might have to traverse multiple pages and hence must interact more with baseline interface in order to find relevant information.

RQ2: What is the effect of task determinability on user experience with experimental interface compared to baseline interface?

Hypothesis 2:

- 2a: If a task is specified (S) task, study proposes null hypothesis. Since these tasks are easy to achieve and hence independent of interface design, users will not much invest their cognitive and emotional effort in completing the search tasks. Therefore, study hypothesize that users will not experience significant differences while completing information specified (S) search task on experimental interface compared to baseline interface. Hence, user rating of each sub-scale of user experience for experimental interface will be identical to baseline interface.

- 2b: If a task is unspecified (U) task, study proposes alternative hypothesis i.e., user will rate each sub scale of user experience higher for experimental interface compared to baseline interface. As experimental interface supports features by which open-ended tasks can be completed effectively, user will rate better experience with experimental interface compared to baseline interface.

Independent Variables:

1. Task Determinability

Bystorm and Hasen [1] have defined tasks based on three levels of granularity: work tasks, information seeking tasks and information search tasks. Information seeking tasks can be completed with context of information search tasks and both of this can be completed with context of work tasks [1]. For this study design, I have considered tasks at information search level (most granular level).

The main goal of this research is to manipulate task determinability of information search task. Task determinability can be defined based on degree of uncertainty associated with outcomes of task and the search process involved in completing the tasks. We manipulate task determinability by narrowing and broadening the scope of search task, respectively. A task with narrowed scope has items specified to look for and can be easily completed. Hence, a task with narrowed scope is less uncertain and user has pre-determined knowledge of documents required and search process involved in completing the tasks. Our task with wide open scope will not mention any items to consider. Hence open ended and tasks are difficult in nature and user may need to traverse multiple pages to complete such tasks. This study design only includes one independent variable which is **task determinability** which has following levels

Unspecified Items (U): For example, what artificial intelligence algorithms can be used to solve traffic management problems? What are major differences between them?
In the above search tasks, no explicit items are mentioned for user to search for. Such tasks are called open-source and are not pre-determined. A user on a way to search will have little knowledge of objects required and search process involved in order to complete the tasks successfully.

Specified Items (S): For example, what is difference between linear regression and logistic regression?
In the above example two machine learning algorithms are explicitly mentioned to the user. Hence, user has prior knowledge of documents required and hence such tasks can be completed easily.

Dépendent Variables

1. User Interaction

User interaction is defined as activities and processes that subject undergoes in order to complete the search tasks.

User interaction is operationalized using following five measures:

- Total queries: Total number of queries issued in order to complete tasks.
- Total time: It is defined as total time spent to complete the information search tasks. It is measured once user start initiating its first query and is measured till user completes the entire search tasks.
- Number of links visited: Total number of search results clicked.
- Number of SERP visited: Total number of search result pages visited in order to complete the tasks.
- Total number of saved documents: Total number of relevant documents saved in order to complete the search tasks

2. User Experience

Task determinability has a significant impact on user experience with interface. Since the main purpose of this research is to evaluate that how much an IR system experimental design contributes to the user to complete the unknown search tasks. Therefore, it is mandatory to evaluate IR system design with the user. User experience can be conceptualized as user's evaluation of a system interface after completing an information search task. User engagement is the qualitative measure of user experience [2]. Thus, for operationalization, O'Brien's based User Engagement Scale (UES) forms basis to measure participant's evaluation of system interface [reference]. O'Brien's user engagement scale is a 31-item scale with 6 sub-factors of engagement i.e., Focused Attention, Perceived Usability, Felt Involvement, Endurability, Novelty and Aesthetic Appeal. A slight modification in O'Brien's UES is done to measure user experience [2]. A 5 item sub-factor "Search effectiveness" is added to the scale which evaluates interface and quality of retrieved documents [2].

- Focused Attention (FA) (7): Focused Attention, feeling of completely occupied in the interaction and losing sense of time.
- Perceived Usability (PU) (7): Perceived usability, negative experience because of interaction with interface and total effort invested [reference]
- Felt Involvement (3): The emotion and feeling of being "drawn in" and lost completely.
- Endurability (5): User's feeling to use or recommend system to other folks in future. Also includes overall success of interaction with the system.
- Search Effectiveness (5): Evaluating interface features, quality of search results.

For all the sub-scales users will indicate their measure of agreement by selecting from the items were (1= Strongly Disagree to 5= Strongly Agree). The above version of O'Brien's User engagement scale was developed by researchers to serve the purpose of evaluation in Human Computer Interaction (HCI)

Experimental Design:

- 2 X 2 Factorial Design

- Task Determinability (between-subject) design
  Task determinability is chosen as between subject design for this study. Each participant will experience same level of tasks in each interface. There are couple of reasons for task determinability as a between subject design. Firstly, if each participant experiences each task level for each interface there is possibility that interaction and experience for one task level at both interfaces might influence the user interaction and experience for another task level. Secondly, we also keep in mind the learning effect of one task might influence and user might perform better for another task level. Also an average user experience for unspecified task at experimental or baseline interface might affect their perception of the interface and can affect their performance for even simple tasks.

- Interface (within-subject) design

Each participant will be exposed to both experimental and baseline interface to complete only one same level of independent variable on each interface, respectively. This allows to compare both the interfaces for one task level and measure outcomes like user interaction, user experience.

- Latin Square design will be used in order to rotate and counterbalance the effect of two interfaces.

I1	I2
I2	I1

Subjects

- For this experiment, we will focus on subjects which have low search experience. Before participating in experiment, users are asked to fill out questionnaires which can assess subjects' previous experience with search interfaces. Subjects which will be found to have low search experience will be only selected to participate in study design. The subjects with high search experience are not tested because they will anticipate past learning experience into evaluation of experimental interface design which will not match exactly with our problem statement. Each subject will be asked to complete one same level tasks on each of interfaces respectively. Four participants will be required to complete the experiment.

- The below table shows activity associated with each participant.
  I1: Experimental Interface
  I2: Baseline Interface
  S: Specified Tasks
  U: Unspecified Tasks

Subjects	Time 1	Time 2
S1	I1: S	I2: S
S2	I2: S	I1: S
S3	I1: U	I2: U
S4	I2: U	I1: U

Tasks

- Tutorials
  Participant will be asked to complete one video tutorial before conducting search tasks. Since our study design has subject with low search experience in past, they need some instruction on how to use system before initiating search tasks. Hence, few minutes video will be recorded showcasing instructions of how to use both the interfaces. Video tutorial also explains user about unique features of each interface and how they can be used to complete the search task. This video was scripted first and then it was recorded. The video only explains about interface features and will not introduce any bias by teaching subjects how to interact with the system.

- Search tasks and User Experience Questionnaire
  Each subject will be asked to either complete specified tasks or unspecified tasks on both the interfaces. Note: Since, users will complete same level of task at both the interfaces, but tasks descriptions will be different for each interface. Example of specified and unspecified tasks is

provided earlier. After completing task at each interface, user will be asked to completer user experience questionnaire. User experience sub scale is defined earlier in this study.

- Exit Questionnaire
  After completing search tasks on both the interface, user will be asked few questions regarding the system interface design and preferences. Firstly, subjects will be asked if they notice any difference between experimental and baseline interface. If yes, then they will be asked to list out potential differences between both the systems. If No, they will be asked detail explanation for that. Next, they will be asked about usefulness of experimental features in completing the search tasks. At last user will be asked about his or her system preference for completing a particular search task level.

Study Procedures Manuals

Pre-Study tasks

1. Each participant should ensure that they have reliable internet connection and speed which will not be interrupted for next 1 hour.
2. Each participant should ensure that they are in quite place and will not be disturbed for next 1 hr while completing search tasks.
3. Participants should have basic knowledge of zoom features like how to share screen, how to on your video etc.

Study Procedures:

1. The study will be conducted online via Zoom meeting. Participants will be asked to activate their video and zoom screen sharing. Zoom session will be recorded in order to monitor your interaction with the interface.
2. User will be provided with consent form and participant will be asked to carefully read and review it. Participants will be asked to digitally sign the form shared with them and share the signed consent form with us back again. Consent form emphasizes that participant would denote approximately 60 minutes to complete information search task on both the interfaces. Participants will be given opportunity to ask questions before moving on to next step.
3. A video tutorial will be played which will allow participants to learn how to use both the interfaces to and to ensure that participants use all the features on experimental and baseline interface to complete the search task according to requirements. You will be given opportunity to make yourself comfortable with both the interfaces before actual start of information search task.
4. We will proceed with the information search task once you are comfortable and confident about the both the interfaces.
5. One by one participant will be asked to complete the search tasks of same levels on both the interfaces. Search task description and instructions will look like

"Please carefully follow the instructions before initiating search task. Your goal is to find the information that is mentioned in the task descriptions. You will need to save documents that you find relevant to information search task. Use the interface feature to save the documents and in case if interface do not provide feature to save document, just bookmark those documents that you found relevant. Once you complete your task just type "complete" in zoom chat box."

6. After completing search task at each interface, you will be asked to complete a user experience questionnaire based on your experience with system interface for given search task. This usually take around 10 minutes. You will be given a short 5-minute break between each tasks.
7. At the end, after completing task and user experience questionnaire for both the interfaces, you will be moved on to last stage to fill out Exit Questionnaire form in which you will be asked about system preference.
8. At the end of exit questionnaires subjects will be requested to provide feedback of the session.

Possible problems and risks

- Since, we are working with computers and technologies there are few unavoidable risks like losing internet connectivity which can interrupt our testing. If we encounter any interruption, we will start the process again were it had abandoned with participants consent. If a search process on an interface is abandoned it will be carried out again with different task descriptions.

References:

[1] Katriina Byström and Preben Hansen. 2005. Conceptual framework for tasks in information studies. JASIST 56, 10 (2005), 1050–1061.

[2] Arguello, Jaime & Wu, Wan-Ching & Kelly, Diane & Edwards, Ashlee. (2012). Task complexity, vertical display and user interaction in aggregated search. SIGIR'12 - Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. 10.1145/2348283.2348343.