Evaluation of the user experience of engagement in the stages of search

Frances Johnson

Manchester Metropolitan University
f.johnson@mmu.ac.uk

Abstract. This study explores user experience of engagement in book search tasks through the factor analysis of the post search experience questionnaire. Specifically, this study aims to identify the dimensions of the participants' evaluation of their experience and to compare this across user-interaction types distinguished in this data set by their 'finding' and 'exploration' activities.

Keywords: User experience, user engagement, search evaluation

1 Introduction

This study contributes to the Social Book Search (SBS) Interactive track to explore complex book search tasks where the users' search takes them beyond a simple query based transaction typically based on objective metadata. Specifically, this study aims to identify the dimensions of the participants' evaluation of their experience and to compare this across user-interaction types distinguished, in this data set, by 'finding' and 'exploration' activities. Analysis of the post search experience questionnaire used in the SBS Lab confirms the factors of the user experience when engaged in complex tasks of book search, and identifies differences in this experience across user groups. Recommendations are made for the further development of post experience questionnaire and for its use in the development of the interface offering exploration.

Various studies have demonstrated that a search, even for the same given task can vary between individuals [1,2,3,4]. The cause of such variation is largely attributed to the knowledge state of the searcher, whether the item is already known to that person, whether they believe the subject of the quest to be findable/retrievable, what is already known about the quest and/or the subject in question as well as the searchers' previous experience in using the system selected to accomplish the task. Hence, it is important that these differences are recognised when developing and assessing novel features in the interface design. Critically, such variation may present a situation in which two users, with the same given task, may be involved in quite different information behaviours when using the same system. This may further influence their perception or assessment of their experience in using the system and so can complicate the researchers' attempt to evaluate the impact the novel interface has had on the user experience. The aim of this study is to draw on longstanding models of search behaviour to explore the extent to which the stages/steps involved in search may be used to

identify searcher types (in the book search context) through activity data, and subsequently how this might impact on the evaluation of the experience. The objectives are to:

1. Identify the types of searchers in using the SBS system to look for a book
2. Identify the factors in the user evaluation of the search experience
3. Identify the differences in the user evaluation of experience, to inform discussion regarding the further development of the user engagement/experience scales in the context of the development of novel interfaces.

2 Background

The background context to this research has two prongs. The first is to refer to the search processes and the recognition that the searcher's critical evaluation of the information retrieved has an important role in taking search beyond a transactional process. The involvement of the searcher in the cognitive activities relating to the evaluation of the information, in interacting with and in understanding the information found, leads to the forming of a personal perspective on its relevance and impacts on the searchers' current knowledge and/or information need state. Saracevic [5] reviews the nature and manifestations of relevance as the key notion in information retrieval, and studies of users' relevance behavior identify variations in the user's involvement in judging the information retrieved and in the factors influencing the interaction [6]. Typically a search involves the formation of a query (or selection of a category to browse within) and the review of items returned to find those deemed relevant or to refine the query to better reflect the information need and to continue the search. Recognition of the searcher's interaction with the information retrieved and listed on the search engine results page has given rise to research into how best to support the searcher in this critical activity. Modern search interfaces will offer functionality beyond search, such as browsing recommendations or sharing information, and, in doing so, offers more ways of exploration broadening the users' possible interactions with the information. In the SBS interface the user is presented with enriched representations of the books and, on which, they may judge or assess their interest in the books found and form an intention to read. This, in particular, assists the user in employing subjective criteria when looking for something 'interesting' or familiar rather than when searching on the more objective metadata with a query articulated for known item search. As understanding of user behaviour and interactions have developed, so too have the requirement for metrics and assessment of the interactive system, focusing on the search experience of the user and some approximation of engagement or more generally affection.

The second point to make for the background to this study is to refer to the development of instruments and tools for data collection in the evaluation of the user experience with novel search and retrieval systems. The quality of the user experience relates to the user's involvement in the experience or 'engagement' which as O'Brien & Toms [7] explain "depends on the depth of participation the user is able to achieve with respect to each experiential attribute" and which may be dependent upon (or

influenced by) a number of system factors such as aesthetics appeal and its usability. The User Engagement Scale (UES) [8] with 31-items is based on the following factors promoting engagement:

- Aesthetic appeal [ae], appeal of interface;
- Novelty [no] appeal to curiosity and unfamiliarity;
- felt involvement [fi] and focused attention [fa], users must be focused;
- perceived usability [pu] for engagement to be sustained;
- endurability [en] – people remember enjoyable and engaging experiences

Factor analysis of the engagement scale when used to study engagement has reported the components of hedonic engagement, focused attention, affective usability and cognitive effort. O'Brien [9] further explores the role of media format, such as video or audio, for example, on each of the components to further learn about users' perceived engagement measured in varying conditions. In the context of the SBS Lab, for example, we might expect to find the novel multi-stage interface promotes user engagement. Encouraging as this might be, the assessment of the system based on user experience raises the question whether such an observation holds true for different types of users with respect to their behavior or information needs or indeed whether different user types might hold different views on an engaging experience and hence the factors promoting or influencing a positive experience by which to assess the system. This study using the data collected in the SBS Lab thus aims to better understand different users' experience of engagement rather than its measurement for system evaluation.

3 SBS lab

The Social Book Search track in the CLEF 2015 initiative (Conference and Labs of the Evaluation) [10] is to investigate how searchers make use of metadata for book search and to develop a multi stage interface supporting searchers through various stages of search and offering, through social media, more ways to explore rich information about the books. The records for the books are curated from Amazon and the social metadata from Library Thing. The baseline interface represents a standard web search interface with search results shown with a search history and a book-bag resembling a shopping cart system. The item view, once selected from the results, shows the standard metadata, similar books if available and user generated context such as review, star rating and tags. The multi-stage interface supports the three stages of search, explore, focus and refine and the user is able to switch between the stages. The user is offered more ways to explore the collection with the option to filter multi-column search results and on selecting a book in the focus view the full meta data is shown along with a category filter to refine results. The refine view allows the user to refine the list of books added to the book-bag and shows the filters and results as before but with an additional similar books panel to assist the user in augmenting their list of chosen books. Each participant was randomly assigned to use either the baseline or the multi stage interface.

Participating teams from 7 European institutes took part in 2015 recruiting a total of 192 users tasked to find something of interest in the simulated scenario of finding oneself with 15 minutes or so to spare whilst waiting for a friend. The closed task involved each user to find a set number of books for a particular situation, in this case books on survival and relating to the users' hobbies to take to a desert island. All participants were asked to complete a pre task questionnaire and a post task questionnaire (twice, one for each of the tasks). The pre task questionnaire collected demographics and cultural data to characterize the participants. The post task questions asked for their assessment on a Likert scale of the interface components and the metadata parts they had used. After completing both tasks users were asked to complete O'Brien et al's [8] engagement scale.

3.1 Differences in the user types

Two types of participants were identified simply (and in first instance) by the reported usage of the metadata on the results page – that is, the descriptions, publication date, reviews and tags. An initial look at the 'bookbag' where books found were stored indicated variation in the participants' involvement in the task based on the number of books stored, their notes recorded (relating to reasons why the item appeared to be of interest), and the time spent on the tasks. These may indicate involvement, but the self -reporting in the post questionnaire on the metadata used was taken to be the strongest indicator of involvement in the user's critical evaluation of the items found. No use of the metadata was taken to be indication that the searcher had simply been engaged in the task of finding a book, possibly already known to them or as easily as possible with a requirement that may be readily satisfied, e.g,' 'anything on photography interests me and will do'. Conversely where all metadata was reported to be used and useful, indicated the searcher was fully engaged in exploring books found, thinking about their relevance, possibly assessing for 'intention to read'. In this way, two types of participants were identified, one labelled 'searcher' the other 'explorer. The 'searcher' did not use any of the metadata on the results page, or where the descriptions, reviews or tags were used they reported them to be very un-useful, or used only one view, typically the description and reported it to be un-useful or 'So-So' (scoring 2 or 3 on the scale 1-5). The 'explorer' used at least two of the views of description, reviews or tags and reported that they found these to be useful or very useful (4 or 5 on the scale). Each participant was assigned to either of these types from their responses on both the open and the closed task, with N=72 as Searchers and N=119 as Explorers. Similar patterns of behaviour were observed in the data across the tasks and this may be explained by the similarity in the underlying task – 'to find a book you would like to read', although the circumstances differed, one being in finding a bit of time to look for a book the other being to take some books to read on the desert island. Therefore usage of the metadata (the extent and evaluation of how useful) was taken as indicator of approach taken, distinguished by the participants' involvement in reviewing and choosing books. Hence the distinction was made between the two user groups in the activity of finding a book to read to an involvement in making an assessment of the interest and suitability of the book for the reader.

3.2 Analysis of the post experience

Data collected from the 191 participant responses in the post experience questionnaire were entered into IBM SPSS Statistics 22. The questionnaire comprised 31 items which had been grouped to reflect the constructs of the user experience of endurable [en], focused attention [fa], felt involved [fi], aesthetic [ae], novelty [no] and perceived usability [pu]. To verify these, in this study, factor analysis with principal component analysis (PCA) was used to analyse the relationships among the items and where these converge, to identify and extract the underlying factors in the participants' responses. The extracted factors with the grouped items are considered to measure some underlying or latent constructs which in a sense fit together as the experience of engagement as a standardized entity. The suitability of the data for this test was established. The KMO value (.915) is greater than the recommended value of 0.6. Bartlett's test was statistically significant with a value of .000 level [11]. The major principle components (with eigenvalue greater than1) were extracted as constructs and to satisfy convergent validity, that all items intended to measure a construct do reflect that construct, the factor loadings were greater than 0.5 (any below are included but shown in italics in Table 1).

These derived factors are presented in Table 1 with the Item Reliability (IR) score shown in column 1. The principal components extracted suggest a good fit of 31 items to five constructs accounting for 62 % of total variance. Specifically, factor 1 explains 32.2% of the total variance, factor 2 (13.6%), factor 3 (7.8%), factor 4 (4.9%) and factor 5 (3.3%). For consistency these were labelled as close to the original labels of endurability [en], focused attention [fa], aesthetic appeal [ae], and perceived usability [pu] (with novelty [no] and felt involved [fi] items integrating in with endurability),

Table 1. Factor analysis of the user experience

Factors	Items	IR	Mean
Endurable *(Enjoyable Rewarding)*	en2 I consider my experience a success	.778	3.31
	no3 I felt interested in my exploration task	.741	**3.35**
	en4 My exploration experience was rewarding	.681	3.04
	no2 The content of the website incited my curiosity	.631	3.12
	en1 Exploring this website was worthwhile	.624	3.03
	fi3 This exploration experience was fun	.547	2.83
	fi2 I felt involved in this exploration task	.519	3.41
	pu7 I felt in control of my exploration experience	.505	3.23
	no1 I continued to explore this website out of curiosity	*.409*	2.40
Perceived Usability	pu8 I could not do some of the things I needed to do	.736	**3.13**
	pu1 I felt frustrated while exploring this website	.724	2.86
	en3 This experience did not work out as I had planned	.685	2.76
	pu2 I found this website confusing to use	.652	2.51
	pu3- I felt annoyed while visiting this website	.573	2.73
	*pu4 found this website **confusing** to use ?? pu4*	*.446*	2.53
Aesthetic	ae2 This website was aesthetically appealing	.879	2.75
	ae5 The screen layout of this website was visually ple	.844	2.92
	ae4 This website appealed to my visual senses	.831	2.70
	ae3 I liked the graphics and images used on this webs	.801	**3.04**
	ae1 This website is attractive	.798	2.68
	en5 I would recommend exploring this website to my friends and famil	*.414*	2.92
Focused Attention	fa1 I lost myself in this experience	.886	2.16
	fa2 I was so involved in this experience I lost track of time	.822	2.34
	fa5 The time I spent exploring just slipped away	.744	2.70
	fa4- When exploring, I lost track of the world around me	.731	2.07
	fa7 During this experience I let myself go	.654	2.73
	fa6 I was absorbed in exploring	*.491*	**3.06**
	*fi1 I was really **drawn into** my exploration task*	*.455*	2.69
	fa3 I blocked out things around me when I was exploring this website	*.377*	2.85
Demand-ing	pu6 this experience was demanding	.876	2.61
	pu5 Using this website was mentally taxing	.856	2.67

The findings from this analysis aligns with previous research using the UES [7,8,9] suggesting that that users' experience is determined on a range of factors and that the

evaluation of Endurability (or reward and enjoyment) is a pivotal component. Less significant, in terms of accounting for variation in the evaluation of the experience, but nonetheless a factor, is the feeling of Focused Attention. Flow and the feeling of losing track of time is considered to be a key indicator of a person's successful use when deeply engaged in the process of searching. In this data set it appears that the factor Perceived Usability is relatively more important as a factor in the evaluation of the experience. The third factor is Aesthetic Appeal and the items that load onto this suggest that the participants in this study have taken the look and appeal of the interface into account.

3.3 Differences in the experience by searcher type

The SBS interface is designed to support users when looking for an 'interesting' book to read with the enriched metadata enabling user assessment of the books found. Two types, the searcher (N=72) and the explorer (N=119) were identified in the reporting of the use of the metadata. Further analysis of the post experience questionnaire was carried out to determine the extent to which the searcher types hold different experience of the system in terms of what is important or what are the components of the experience. The means were taken on each item in Table 1 and compared by the two groups of users to obtain some hint of the differences in the evaluation of the experience across the two groups. In the Endurability factor, the most influencing in evaluating experience, with Searchers rating was pu7 *I felt in control of my exploration experience* (3.19) whereas Explorers rating was on fi2 *I felt involved in this exploration task* (3.64).

To investigate further, each of the two datasets, for searchers and explorers respectively, were subjected to Principal Component Analysis (PCA) to calculate the contribution of each factor to experience formation. The KMO Measure of Sampling Adequacy was 0.695 for the Searchers and 0.891 for the Explorers verified the suitability of both the Searcher and Explorer datasets. Item Reliability (IR) mostly exceeded the acceptable value of 0.5. Together these indices show that both models had an appropriate level of reliability but that the searcher set only just reached the level. There is an issue regarding the number of respondents to the 31 items in the questionnaire. The general rule is to work at the ratio of 5:1 for factor analysis although studies have reported their results on far lower ratio of 2:1 [12]. In doing so it is important to acknowledge that the findings can only be interpreted in respect to this group of study participants and a generalised model cannot be assumed.

Table 2. Explorers' experience

Explorers		
FACTOR 1 Endurable	IR	Mean
pu7 I felt in control of my exploration experience	.681	3.27

en2 I consider my experience a success	.668	3.55
en4 My exploration experience was rewarding	.654	3.21
no3 I felt interested in my exploration task	.615	3.61
fi2 I felt involved in this exploration task	.585	**3.64**
en1 Exploring this website was worthwhile	.558	3.21
fi3 This exploration experience was fun	*.414*	3.07
FACTOR Focused Attention		
fa5 The time I spent exploring just slipped away	.806	2.76
fa7 During this experience I let myself go	.770	2.98
fa4 When I was exploring, I lost track of the world around me	.699	2.21
fa6 I was absorbed in exploring	.633	**3.27**
fi1 I was really drawn into my exploration task	.577	2.87
FACTOR Aesthetics		
ae4 This website appealed to my visual senses	.909	2.85
ae2 This website was aesthetically appealing	.906	2.83
ae5 The screen layout of this website was visually pleasing	.856	3.02
ae1 This website is attractive	.833	2.75
ae3 I liked the graphics and images used on this website	.819	3.12
en5 I would recommend exploring this website to my friends and family	*.429*	2.81
no2- The content of the website incited my curiosity	*.408*	**3.30**
FACTOR Demanding		
pu6 This experience was demanding	.891	2.70
pu5 Using this website was mentally taxing	.848	2.71
FACTOR Perceived Usability		
pu2 I found this website confusing to use	.759	2.54
pu1 I felt frustrated while exploring this website	.694	2.71
pu3 I felt annoyed while visiting this website	.685	2.55
pu8 I could not do some of the things I needed to do on this website	.594	**3.13**
pu2 -- I found this website confusing to use (pu4	.541	2.33
en3 This experience did not work out as I had planned	*.498*	2.65
[Focused attention 2]		
fa3 I blocked out things around me when I was exploring this website	.692	3.05
no1– *I continued to explore this website out of curiosity*	*.338*	2.55

Table 3. Searchers' experience

SEARCHERS	IR	Mean
FACTOR Aesthetics		
ae3 I liked the graphics and images used on this we	.832	2.92
ae2 This website was aesthetically appealing	.776	2.63
ae1 This website is attractive	.749	2.60
ae4 This website appealed to my visual senses	.726	2.47
ae5 The screen layout of this website was visually pleasing	.715	2.78
en5 I would recommend exploring this website to my friends &family	.623	2.32
en1 -Exploring this website was worthwhile	*.445*	2.72

FACTOR Perceived Usability		
pu5 Using this website was mentally taxing	.817	2.60
pu7 I felt in control of my exploration experience	.789	3.19
pu6 This experience was demanding	.700	2.44
pu2 I found this website confusing to use pu4)	.623	2.85
pu2 -I found this website confusing to use	.510	2.47
pu1 -- I felt frustrated while exploring this website	*.398*	3.10
FACTOR Focused Attention		
fa7 During this experience I let myself go	.681	2.33
fa6 I was absorbed in exploring (n	.640	2.69
pu3 I felt annoyed while visiting this website	.560	3.06
en4 My exploration experience was rewarding	.520	2.76
fi2- I felt involved in this exploration task	*.429*	2.99
fi3 This exploration experience was fun	*.404*	2.43
FACTOR focused attention (2)		
fa5 The time I spent exploring just slipped away	.889	2.61
fa1 I lost myself in this experience	.789	2.06
fa2 I was so involved in this experience that I lost track	.511	2.17
fi1 I was really drawn into my exploration task	*.483*	2.42
FACTOR focused attention (3)		
fa3 I blocked out things around me when I was exploring this website	.802	2.51
fa4 When I was exploring, I lost track of the world around me	.576	1.88
no3- I felt interested in my exploration task	*.374*	2.93
FACTOR Novelty		
no2 The content of the website incited my curiosity	.773	2.81
no1 I continued to explore this website out of curiosity	.759	2.15
en2 I consider my experience a success	*.397*	2.92
FACTOR Endurable		
pu8 I could not do some of the things I needed to do on this	.818	3.15
en3 This experience did not work out as I had planned	.567	2.96

For Explorers the evaluation of Endurability (reward and enjoyment) is a pivotal component (accounting for 33.1% of variation) along with Focused Attention (13.8% of variation). Less significant, but nonetheless factors are Aesthetic Appeal, Demand and Perceived Usability. For Searchers the experience appears quite different based on Aesthetic Appeal (as the pivotal factor with 27.1% of variation) and Perceived Usability (15.2% of variation). The three smaller factors of Focused Attention are of less importance and with a negative loading indicate users' assessment of feeling 'Not Focused'. Finally the weakest factors for the Searchers are Novelty and Endurability based on only a small number of negative items. For example, the Endurability factor for the Searcher was formed with two items pu8 'I could not do some of the things I needed to do with this' and en3 'This experience did not work out as I had planned'.

4 Discussion and perspectives for future work

This study based on the factor analysis of the post experience of two types of participants in the SBS interactive task suggests that users may well have a different experience. Our analysis suggests that the experience of engagement may be measured on the components previously identified, in particular Endurability, Focused Attention, Aesthetics, Perceived Usability. However further analysis of the extent of the users' interaction to find a book indicates type of searcher/explorer as a determinant of that experience. The Explorers interacting with the enriched metadata describing the items found were influenced by Endurability (the enjoyment or reward) and Focused Attention in assessment of experience. The Searchers with little or no such interaction assessed the experience on the factors of Aesthetic appeal and Perceived Usability of the website.

Further analysis of the data may be possible to further distinguish types of searchers by activity data and according to search stages. Further analysis is needed of the relationship between the type of searcher, the type of interface used and their assessment on the components of the experience to help better understand interactive search and its assessment. The cross tabulation table in the Appendix suggests that the interface used, baseline or multi stage, may impact on the search behavior when distinguished in this way, that is - by responses to use of the metadata. Where the codes 1 and 3 identify Searchers and 4 and 5 Explorers, tests of significance could indicate whether fewer Searchers came about from use of the multi stage interface. This might be expected and could indicate success of the multi stage interface with respect to its goals to encourage exploration. The results of this study suggest that the user experience of modern features supporting interactive search activity, where the user is involved in the later stages of the critical evaluation of the information found, will be assessed on the hedonic scales of engagement. Where these features are not used, the experience may be assessed on the more objective and perhaps pragmatic usability, ease of use and appeal of the interface.

Factor analysis of the post experience has helped to better understand the searcher and their behavior and interactions with the system and the information found. The evaluation of the novel search interface requires that the users' involvement in the critical assessment of the information found on criteria of relevance, usefulness or interestingness has been reached and is assessed by engagement. As a perspective for future work, this study contributes to the paradigm for IR evaluation based on user experience of the search as a process rather than the evaluation of the output, the search results. However it further contributes in making the distinction of the users' involvement in the critical evaluation of the found items as identifying two types of users of the same system and seemingly the same task. It appears that this interaction with the information distinguishes users' activity of search to such an extent that their entire evaluative experience of a search on a system will be different.

References

1. Ford, N., Miller, D., Moss, N. The role of individual differences in Internet searching: An empirical study. J. Am. Soc. Inf. Sci., 52, 1049–1066 (2001)
2. Vibert, N., Ros, C., Bigot, L. L., Ramond, M., Gatefin, J., Rouet, J.-F. Effects of domain knowledge on reference search with the PubMed database: An experimental study. J. Am. Soc. Inf. Sci., 60, 1423–1447 (2009)
3. Kules, B., Capra, R. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. J. Am. Soc. Inf. Sci., 63,114–138 (2012)
4. Xianjin Zha, Jinchao Zhang, Yalan Yan. Exploring the effect of individual differences on user perceptions of print and electronic resources,Library Hi Tech, 32 .346 - 367 (2014)
5. Saracevic, T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. Journal of the American Society for Information Science and Technology 58, 2126-2144 (2007):
6. Johnson, F., Rowley, J, Sbaffi, L., Exploring information interactions in the context of Google. Journal of the Association for Information Science and Technology. (2015) First published online doi: 10.1002/asi.23443
7. O'Brien, H. L., Toms, E. G. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci., 59, 938–955 (2008).
8. O'Brien, H. L.,Toms, E. G. The development and evaluation of a survey to measure user engagement. J. Am. Soc. Inf. Sci., 61, 50–69 (2010)
9. O'Brien, H. L., Lebow, M., Mixed-methods approach to measuring user experience in online news interactions. J. Am. Soc. Inf. Sci., 64,1543–1556 (2013)
10. Cappellato, L., Ferro, N., Jones, G., San Juan, E. (editors) CLEF 2015 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/ Vol-1391/ (2015).
11. Hair, J.F, Black, W.C. Babin,, B.J., Anderson, R.E, Tatahm, R.L Multivariate data analysis, Prentice Hall 2006
12. Costello, Anna B. & Jason Osborne Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis Practical Assessment, Research & Evaluation, 10(7) ISSN 1531-7714 (2005)

Appendix

interface * VAR00005 Crosstabulation

Count

		VAR00005				
		1.00	3.00	4.00	5.00	Total
interface	baseline	8	34	35	19	96
	multi-stage	3	27	42	23	95
Total		11	61	77	42	191