Towards Observable Indicators of Learning on Search

Jacek Gwizdka
School of Information
University of Texas at Austin
sal2016@gwizdka.com

Xueshu Chen
School of Information
University of Texas at Austin
xueshu_chen@utexas.edu

ABSTRACT
On an example of a recently conducted user study, we discuss assessment of learning on search as well its correlates in search behaviors and associated eye-tracking measures. Since we are reporting on a work in progress, the study is meant to illustrate our approach and our choices of measures to inspire a discussion.

Keywords
Information Search, Search as Learning, Eye-tracking, Measurement, Learning Assessment.

1. INTRODUCTION & BACKGROUND
From its very origins information science has been concerned with ways and means of storing knowledge, organizing it, as well as helping people to find, use, and learn from it. Theorists of information science conceptually linked information interaction processes with the states of human knowledge (e.g., Belkin's ASK [2] and Dervin's sense-making [3] and information seeking was described as "a process, in which humans purposefully engage in order to change their state of knowledge." [4]. Yet in spite of the long established relationship with learning, only a few empirical studies exist that focus on search as learning (e.g [5–7]). However, with the recent special journal issue [8], and with this and two earlier workshops, we observe an increased interest in this topic. One research challenge identified at a previous workshop [9] was how to assess learning in the context of purposeful information seeking. This is where we aim to contribute through this project. This short paper presents our approach, method, and initial data analysis.

Our working definition of learning is *any* change in person's knowledge structures. We consider that learning can take place at many levels [10] and we are particularly influenced by the cognitive, skill-based, and affective theory of learning outcomes (CSALO) model [1]. This framework contains elements related to searching to learn (e.g., declarative knowledge) as well as learning to search (strategies, tactics, procedural knowledge). According to this model learning outcomes are partially reflected in changes in verbal knowledge, knowledge organization, and cognitive strategies. We are particularly interested in assessing changes in verbal knowledge.

Our prior work [11–13] has demonstrated feasibility of using eye-tracking to detect relationship between eye movement and knowledge levels. The method takes advantage of a direct relationship between eye movement patterns and cognitive processes. One goal of the project presented in this short paper is to connect eye-tracking measures and traditional IR measures (e.g., number and kind of query reformulations) with measures of learning. Since we are reporting on work in progress, the initial data analysis will serve as only a simple illustration of our approach, while our choices of measures will inspire a discussion.

Search as Learning (SAL), July 21, 2016, Pisa, Italy

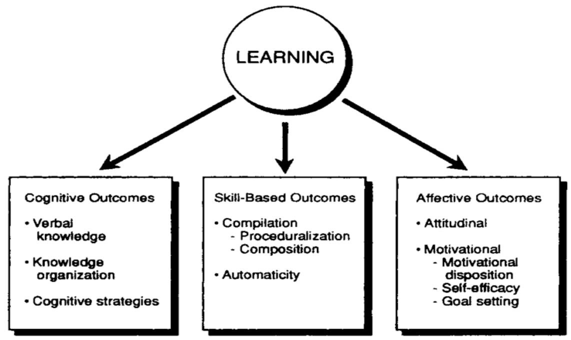

Figure 1. Cognitive, Skill-based, and Affective Theory of Learning Outcomes (CSALO) Model. Source: [1]0]

2. METHOD
A lab-based experiment was conducted in the Information eXperience lab at University of Texas at Austin (N=30). Data is reported here for 26 of these subjects (16 females; mean age of all participants 24.5). Participants who volunteered after seeing the recruitment notice posted at the university bulletin were pre-screened for their English native level, eye-sight, and topic familiarity. All participants reported daily Internet use longer than an hour and everyday Google usage. Most have been searching online for 7 years or more. The majority also considered themselves as proficient in online information search. To understand how people seek heath information using the Internet and acquire new domain knowledge, we asked each participant to perform three information search tasks (two assigned multi-faceted tasks and one self-generated) on health-related topics in counterbalanced order (six rotations), plus one training task. The assigned search tasks followed a simulated work task approach that triggers a realistic information need for participants as they were asked to find useful information for answering the task questions [14] (Table 1).

Table 1. Search tasks.

Assigned tasks
Task 1–*Vitamin A: Your teenage cousin has asked your advice in regard to taking vitamin A for health improvement purposes. You have heard conflicting reports about the effects of vitamin A, and you want to explore this topic in order to help your cousin. Specifically, you want to know: 1) What is the recommended dosage of vitamin A for underweight teenagers?* *2) What are the health benefits of taking vitamin A? Please find at least 3 benefits and 3 disadvantages of vitamin A.* *3) What are the consequences of vitamin A deficiency or excess? Please find 3 consequences of vitamin A deficiency and 3 consequences of its excess.* *4) Please find at least 3 food items that are considered as good sources of vitamin A.*

Task 2–*Hypotension: . Your friend has hypotension. You are curious about this issue and want to investigate more. Specifically, you want to know: 1) What are the causes of hypotension?*
2) What are the consequences of hypotension?
3) What are the differences between hypotension and hypertension in terms of symptoms? Please find at least 3 differences in symptoms between them.
4) What are some medical treatments for hypotension? Which solution would you recommend to your friend if he/she also has a heart condition? Why?

Example self-generated tasks

Ex.1. Chrohn's disease- I know someone who was recently diagnosed and am curious about the disease.

Ex.2. My friend has lupus. What are the symptoms for lupus? What are the long-term consequences of lupus including the life expectancy? Are there any cures? What treatments are available?

Participants searched publicly available web pages using Google and were asked to save relevant web pages with their typewritten notes and/or information copied/pasted from the source. While there was no time limit, each user session typically lasted from 1.5 to 2 hours. Each participant completed an eHEALS questionnaire, a Pre- and a Post-task Questionnaire, a Post-Search Interview on how they arrived at their solutions for one of the saved web pages per task, and an Exit Questionnaire. During search in the experiment, all of the participants' interactions with the computer system, including eye gaze, brain activity recordings (frontal area), facial expressions (web cam), were recorded. Eye tracking data was collected using a Tobii TX-300 eye-tracker. Participant brain wave levels were recorded using a wireless, consumer-level device headset (MyndWave). At the completion of a session, each participant received $25.

Both the Pre and Post-Task Questionnaires contained two parts: knowledge assessments and interest in a search topic. In knowledge assessments, participants were asked to list as many words or phrases as they can on the topic of a search task with no time limit. As we have just recently finished the study, we focus on participants' responses to the free recall test to identify knowledge gains through information seeking and relate them to basic behavioral measures on Web search, adding eye fixation durations and counts.

2.1 Measures

Our goals include measuring verbal and concept learning on the search process. We want to measure the difference in participant's knowledge of a search topic before and after each task, hence we need two measurement points. We considered a number of different possibilities of assessing participant's knowledge level on the task topics. We briefly present our deliberations. Fact-checking questions before a task were considered inappropriate, because we wanted to avoid exposing participant to the topic's content before they start the search. Since the tasks were conducted on an open web, we could not use a technique such as Sentence Verification Technique (SVT) [15], which requires creation of questions for each document. Our participants were not experts on the topics, hence concept maps and mind-mapping were deemed inappropriate as it is particularly difficult to score for non-experts.

We decided on asking participants to list words and phrases related to each task topic before and after each task. Participants were also asked to annotate relevant web pages and to create from these annotations final notes for each task. Participant entered the annotations while they were on content web pages, whereas the listed words and phrases on pre- and post-task knowledge assessment were from their memory. In addition, we collected a list of keywords and phrases on the assigned task topics from crowd workers on Amazon Mechanical Turk. We plan to use it in assessing participant knowledge by applying automated scoring and calculating semantic similarity using (e.g., using LSA).

Table 2. Dependent measures

Construct	Operationalization
Knowledge gain	difference in the number of items entered after and before each task (absolute and ratio)
Expertise gain	mean frequency of nouns after a task; normalized by the number of nouns used
	ratio of the mean frequency of nouns after to before a task; normalized by the number of nouns used
	mean frequency of new nouns used after a task; normalized by the number of nouns used
	mean rank of nouns listed after a task
	mean rank of new nouns listed after a task

The methods we used in assessing knowledge included, for example, statement counting [16], word analysis (e.g. word frequency, in particular for nouns), while we plan to use more sophisticated methods in the future (e.g., topic analysis [17] and semantic analysis). The methods aim at assessing knowledge gain and expertise gain. With increasing expertise, people use more sophisticated vocabulary. This sophistication is expressed in the use of less frequent and more specialized vocabulary, hence our use of word usage frequency (and word usage rank) as one of the dependent measures. We used word frequencies and ranks of 1/3 million of most frequent words taken from Google Web Trillion Word Corpus [18] as described by Norvig in chapter 14 in [19].

3. RESULTS

The mean frequencies and ranks of nouns entered before and after a task differed significantly (Mann-Whitney non-parametric test statistic=229728.5, p=0.0026; Figure 2).

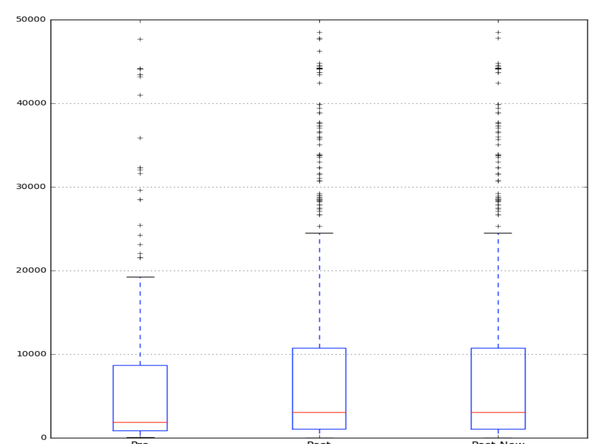

Figure 2. Mean ranks of nouns in Pre-, Post-task, and new nouns Post-task.

We performed linear regression with the independent variables presented in Table 3 and one dependent variable at a time (Table 2) – thus, we run four regressions. Three of the obtained models (except for ratio of frequencies after and before a task) were significant. However, the values of R^2 were modest and ranged from 0.24 to 0.28.

Table 3. Independent measures

Measure Category	Measure
Task level	Time on task
Query	Query count
	Query length
Content Web pages	Number of pages visited
	Time on a page
	Total fixation duration
	Count of fixations
	Proportion of reading fixations
	Proportion of durations of reading fixations
SERPs	Number of SERPs visited

The significant predictors included, 1) number of queries entered and number of SERPs visited in a model with ratio of the number of items entered after and before each task as the dependent variable, and 2) average query length in models with the mean frequency of use of nouns (or new nouns) after a task as the dependent variable.

A plausible interpretation could be that the more queries are issued the more items are entered in the post task knowledge list, and that there is a trade-off with the number of SERPs, namely, with more SERPs visited number of items entered decreases.

For the second and third one, the interpretation is less exciting as it seems to indicate that the longer the average query is the higher the normalized frequency of nouns or new nouns entered in post-task knowledge assessment.

The eye-tracking variables were not found to be significant contributors to the dependent variables of interest. This, perhaps, should not be surprising as they were obtained for all visits to content pages without further differentiation of page content of search task phase. We plan to use more specific eye-tracking measure in our future work.

4. DISCUSSION AND CONCLUSIONS

We reported on our work-in-progress, in which we seek to make a methodological contribution. The results generally indicate a feasibility of the proposed approach, which we may take as an early indication of some success. However, the relative simplicity of employed measures leaves room for improvement and, as indicated throughout the paper, we plan on using more sophisticated assessment techniques.

The broader impact of implicit detection of gains in a person's knowledge and, thus, of learning, lies in its applicability not only to the design of search systems and to improving understanding of human-information interaction but also to a wide variety of information systems, including online learning and intelligent tutoring systems.

5. ACKNOWLEDGMENTS

This project has been funded in part by IMLS Career award #RE-04-11-0062-11 and in part by a fellowship from School of Information to Jacek Gwizdka.

6. REFERENCES

[1] K. Kraiger, J.K.F. 1993. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*. 78, (1993), 311–328.

[2] Belkin, N.J. 1980. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*. 5, (1980), 133–143.

[3] Dervin, B. 1992. From the mind's eye of the user: The sense-making qualitative-quantitative methodology. *Qualitative Research in Information Management*. (1992), 61–84.

[4] Marchionini, G. 1997. *Information Seeking in Electronic Environments*. Cambridge University Press.

[5] Jansen, B.J. et al. 2009. Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management*. 45, 6 (Nov. 2009), 643–663.

[6] Wilson, M.J. and Wilson, M.L. 2013. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology*. 64, 2 (2013), 291–306.

[7] Collins-Thompson, K. et al. 2016. Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval* (New York, NY, USA, 2016), 163–172.

[8] Hansen, P. and Rieh, S.Y. 2016. Editorial Recent advances on searching as learning: An introduction to the special issue. *Journal of Information Science*. 42, 1 (Feb. 2016), 3–6.

[9] Freund, L. et al. 2013. From Searching to Learning. *Evaluation Methodologies in Information Retrieval*. M. Agosti et al., eds. 102–105.

[10] Anderson, L.W. et al. 2001. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Longman.

[11] Cole, M.J. et al. 2011. Dynamic assessment of information acquisition effort during interactive search. *Proceedings of the American Society for Information Science and Technology*. 48, 1 (2011), 1–10.

[12] Cole, M.J. et al. 2013. Inferring user knowledge level from eye movement patterns. *Information Processing & Management*. 49, 5 (Sep. 2013), 1075–1091.

[13] Cole, M.J. et al. 2011. Task and user effects on reading patterns in information search. *Interacting with Computers*. 23, 4 (Jul. 2011), 346–362.

[14] Borlund, P. 2003. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*. 8, 3 (2003), paper no. 152.

[15] Freund, L. et al. 2016. The effects of textual environment on reading comprehension: Implications for searching as learning. *Journal of Information Science*. 42, 1 (Feb. 2016), 79–93.

[16] Wilson, M.L. and schraefel, m c 2008. A Validated Framework for Measuring Interface Support for Interactive Information Seeking. (2008).

[17] Kammerer, Y. et al. 2009. Signpost from the Masses: Learning Effects in an Exploratory Social Tag Search Browser. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), 625–634.

[18] Brants, T. and Franz, A. All Our N-gram are Belong to You. *Research Blog*.

[19] Segaran, T. and Hammerbacher, J. 2009. *Beautiful Data: The Stories Behind Elegant Data Solutions*. O'Reilly Media, Inc.