Draft January 24, 2023

The Dissertation Committee for Nilavra Bhattacharya
certifies that this is the approved version of the following Dissertation:

LongSAL: A Longitudinal Search as Learning Study With University Students

Committee:

Jacek Gwizdka, Supervisor

Soo-Young Rieh

Matthew Lease

Robert Capra

Draft January 24, 2023

LongSAL: A Longitudinal Search as Learning Study With University Students

by

Nilavra Bhattacharya

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
May 2023

Acknowledgements

This section will be fleshed out in more detail after the initial committee-submission on Feb 27, 2023. For now, I wish to thank the following people :

- Committee Members: Jacek Gwizdka, Soo Young Rieh, Matt Lease, Rob Capra

Abstract

LongSAL: A Longitudinal Search as Learning Study With University Students

Nilavra Bhattacharya , PhD TBD
The University of Texas at Austin, 2023

Supervisor: Jacek Gwizdka

Learning today is about navigation, discernment, induction, and synthesis of the wide body of information on the Internet present ubiquitously at every student's fingertips. Learning, or addressing a gap in one's knowledge, has been well established as an important motivator behind information-seeking activities. The Search as Learning research community advocates that online information search systems should be reconfigured to become educational platforms to foster learning and sensemaking. Modern search systems have yet to adapt to support this function. An important step to foster learning during online search is to identify behavioural patterns that distinguish searchers gaining more vs. less knowledge during search. Previous efforts have primarily studied searchers in the short term, typically during a single lab session. Many researchers have expressed their concern over this ephemeral approach, as learning takes place over time, and is not fleeting. We propose an exploratory longitudinal study to analyze the long-term searching behaviour of students enrolled in a university course, over the span of a university semester. Our research aims are to identify if and how students' searching behaviour changes over time, as they gain new knowledge on a subject; and how do processes like motivation, metacognition, self-regulation, and other individual differences moderate their 'searching as learning' behaviour. Findings

from this exploratory longitudinal study will help to build improved search systems that foster human learning and sensemaking, and are more equitable in the face of learner diversity.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Searching as Learning: Overview

Searching for information is a fundamental human activity. In the modern world, it is frequently conducted by users interacting with online search systems (e.g., web search engines), or more formally, **Information Retrieval** (IR) systems. As early as in 1980, Bertam Brookes, in his 'fundamental equation' of information and knowledge, had stated that an information searcher's current state of knowledge is changed to a new knowledge structure by exposure to information (Brookes, 1980, p. 131). This indicates that searchers acquire new knowledge in the search process, and the same information will have different effects on different searchers' knowledge states. Fifteen years later, (Marchionini, 1995) described information seeking as "a process, in which humans purposefully engage in order to change their state of knowledge". Thus, we have known for quite a while that search is driven by higher-level human needs, and IR systems are a means to an end, and not the end in itself. **Interactive information retrieval** (IIR), a.k.a. human-computer information retrieval (HCIR) (Marchionini, 2006) refers to the study and evaluation of users' interaction with IR systems and

users' satisfaction with the retrieved information (Borlund, 2013).

Despite their technological marvels, modern IR systems falls short in several aspects of fully satisfying the higher level human need for information. In essence, IR systems are software that take, as input, some query, and return as output some ranked list of resources.

> *Within the context of information seeking, (search engines and IR systems)* ***feel*** *like they play a prominent role in our lives, when in actuality, they only play a small role: the* ***retrieval*** *part of information . . .*
>
> - *Search engines* ***don't help us identify what we need*** *– that's up to us; search engines don't question what we ask for, though they do recommend queries that use similar words.*
> - *Search engines* ***don't help us choose a source*** *– though they are themselves a source, and a heavily marketed one, so we are certainly compelled to choose search engines over other sources, even when other sources might have better information.*
> - *Search engines* ***don't help us express our query*** *accurately or precisely – though they will help with minor spelling corrections.*
> - *Search engines do help retrieve information—this is the primary part that they automate.*
> - *Search engines* ***don't help us evaluate the answers we retrieve*** *– it's up to us to decide whether the results are relevant, credible, true; Google doesn't view those as their responsibility.*
> - *Search engines* ***don't help us sensemake*** *– we have to use our minds to integrate what we've found into our knowledge.*
>
> — (Ko, 2021)

In recent years, the IIR research community has been actively promoting the **Search as Learning** (SAL) research direction. This fast-growing community of researchers propose that search environments should be augmented and reconfigured to foster learning, sensemaking, and long-term knowledge-gain. Various workshops and seminars have been organized to develop research agendas at the interaction of IIR and the Learning Sciences (Agosti et al., 2014; Allan et al., 2012; Collins-Thompson

et al., 2017; Freund et al., 2013, 2014; Gwizdka et al., 2016) Additionally, special issues on Search as Learning have also been published in the *Journal of Information Science* (Hansen & Rieh, 2016) and in the *Information Retrieval Journal* (Eickhoff et al., 2017). Articles in these special issued presented landmark literature reviews (Rieh et al., 2016; Vakkari, 2016), research agendas, and ideas in this direction. Overall, these works generally advocate that future research in this domain should aim to:

- understand the contexts in which people search to learn
- understand factors that can influence learning outcomes
- understand how search behaviours can predict learning outcomes
- develop search systems to better support learning and sensemaking
- help searchers be more critical consumers of information
- understand the cognitive biases fostered by existing search systems
- develop search engine ranking algorithms and interface tools that foster long term knowledge gain

Parallelly, the Educational Science and the Learning Science research communities have also been organizing workshops and formulating research agendas to conceptualize forms of 'new learning' (Cope & Kalantzis, 2013; Kalantzis & Cope, 2012; New London Group, 1996) that are afforded by innovations in digital technologies and e-learning ecologies (Cope & Kalantzis, 2017). Higher education researchers have been increasingly studying how students' information search and information use behaviour affect and support their learning (Weber et al., 2019, 2018; Zlatkin-Troitschanskaia et al., 2021). Efforts are underway to conceptualize a theoretical framework around new forms of e-Learning that is aided and afforded by digital technologies (Amina, 2017; Cope & Kalantzis, 2017). In the community's own words:

"learning today is more about navigation, discernment, induction, and synthesis" of the wide body of information present ubiquitously at every student's fingertips (Amina, 2017). Therefore, "knowing the source, finding the source, and using the information aptly is important to learn and know now more than ever before" (Cope & Kalantzis, 2013). All of these interests in the intersection of searching and learning goes to emphasize that understanding learning during search is critical to improve human-information interaction.

1.2 Problem Statement

A major limitation in the area of Search as Learning, Interactive IR (IIR), and more broadly, in Human-Computer Interaction (HCI) research is that, the user is examined in the short-term, typically over the course of a single experimental session in a lab (Karapanos et al., 2021; Kelly et al., 2009; Koeman, 2020; Zlatkin-Troitschanskaia et al., 2021). Very few studies exist in the search-as-learning domain that have observed *the same participant* over a longer period of time than a single search session (Kelly, 2006a, 2006b; Kuhlthau, 2004; Vakkari, 2001; White et al., 2009; Wildemuth, 2004). This ephemeral approach has acute implications in any domain where learning is involved because "learning is a *process* that leads to *change* in knowledge ... (which) unfolds over time" (Ambrose et al., 2010), and "...does not happen all at once"(White, 2016b).

To the best of the author's knowledge, almost no new longitudinal studies were reported in major search-as-learning literature in the last five years, that systematically studied students' information search behaviour and information-use over the long term, in their *in-situ* naturalistic environment and contexts, and linked those behaviours quantitatively to the students' learning outcomes and individual differences.

Higher education students are increasingly using the Internet as their main learning environment and source of information when studying. Yet, the short term nature of research in this domain creates significant gaps in our knowledge regarding how students' information search behaviour and information use develop over time, and how it affects their learning (Zlatkin-Troitschanskaia et al., 2021).

> *When research in this area "relies so heavily on (short-term) lab studies, can we realistically say we are comprehensively studying human-tech interactions – when many of those interactions take place over long periods of time in real-world contexts? ... An over-reliance on short studies risks inaccurate findings, potentially resulting in prematurely embracing or disregarding new concepts."*
>
> – (Koeman, 2020).

Current search engines and information retrieval systems "do not help us know what we want to know, ... do not help us know if what we've found is relevant or true; and they do not help us make sense of the retrieved information. All they do is quickly retrieve what other people on the internet have shared" (Ko, 2021). Unless we have more long-term understanding of the nature of knowledge gain during search, the limitations of current search systems will continue to persist. Increased knowledge and understanding of students', and more broadly searchers', information searching and learning behaviour over time will help us to overcome the limitations of current IR systems, and transform them into rich learning spaces where "search experiences and learning experiences are intertwined and even synergized" (Rieh, 2020). The internet and digital educational technologies offer great opportunities to transform learning and the education experience. Enabled by our increased comprehension of the longitudinal searching-as-learning process, improved and validated by empirical data, we can create a new wave of fundamentally transformative educational technologies and "e-learning ecologies, that will be more engaging for learners, more effective (than traditional classroom practices), more resource efficient, and more equitable in the face of learner diversity" (Cope & Kalantzis, 2017).

1.3 Purpose of this Dissertation Proposal

To address the gaps in our knowledge of how information searching influences students' learning process over time, this dissertation proposal proposes to conduct a semester-long longitudinal study (approx. 16 weeks) with university student participants. The overarching research aim is to identify how students' online searching behaviour correlate with their learning outcomes for a particular university course. Building upon principles from the Learning Sciences (Ambrose et al., 2010; National Research Council, 2000; Novak, 2010; Sawyer, 2005), and empirical evidences from the Information Sciences (Rieh et al., 2016; Vakkari, 2016; White, 2016a), this dissertation proposal aims to

- situate students as learners in their naturalistic contexts, and characterized by their individual differences,
- measure students' information search and information use behaviour over time, and
- correlate the information search behaviour with the learning outcomes for the university course.

Learning, or addressing a gap in one's knowledge, has been well established as an important motivator behind information-seeking activities (Section 1.1). Therefore, search systems that support rapid learning across a number of searchers, and a range of tasks, can be considered as more effective search systems (White, 2016a, p. 310). This dissertation proposal takes a step in this direction. "It opens great expectations for many-sided, great contribution to our knowledge on the relations between search process and learning outcomes" (Bhattacharya, 2021 anonymous reviewer).

1.4 Outline

This dissertation proposal document is structured as follows. First, principles of learning and relevant background from the domain of Educational Sciences are presented in Chapter 2. Next, relevant empirical evidences from the Information Searching Literature are discussed in Chapter 3. Chapter 4 presents the research questions, the overarching hypotheses, and discusses their rationale in the context of the existing research gaps. Chapter 5 describes the research methods, including the longitudinal study design, experimental procedures, data collection and analyses plans, anticipated limitations, and expected schedule to complete the dissertation.

2

Background: Knowledge and Learning

3

Background: Information Searching

4

Research Questions and Hypotheses

5

Methods: Longitudinal Study

5.1 Study Design

5.2 Apparatus

5.2.1 YASBIL Browsing Logger

5.2.2 Qualtrics Survey Software

5.2.3 Zoom Video-conferencing Software

5.3 Search Task Template

5.4 Procedure

Insert diagram and check how it looks

Reference it

	ENTRY SURVEY [SUR1]	INITIAL SESSION [SES1]	LONGITUDINAL TRACKING [SES2a, SES2b, SES2c SES2d]	MID-TERM SURVEY [SUR2]	FINAL SESSION [SES3]	EXIT SURVEY [SUR3]
Why	Record individual-differences	Establish baseline search behaviour and initial knowledge	Understand change in search behaviour and knowledge acquisition over time	Track changes in individual differences	Record "evolved" search behaviour, and "final" knowledge	Final state of individual differences
When	Week 1-2 of semester	Weeks 1-2 of semester; after SUR1	Four different points over the semester	Semester mid-point	After last day of classes	Anytime after SES3
Where	Asynchronous	**Synchronous**: Remote	Async	Async	**Sync:** Remote	Async
What	Only in SUR1: −Consent Form −Search Exp. & IT proficiency Repeated in SUR2 and SUR3: −Course Load −Note-taking strategies −Motivation −Self-regulation −Metacognition	**Two search tasks:** for each task, participants searched to find at least three unique, good quality online resources relevant to a given topic. • Pre-search self reporting: existing knowledge, interest, perceived difficulty • Post-search self reporting: perceived learning, perceived search success, interest and motivation, decision making **One website reliability assessment** from Stanford History Education Group (SHEG)	Participants recorded browsing activity when they worked on **final project assignment –** writing a research paper – at four different points in the semester. − SESa: Proposal − SES2b: Paper Outline − SES2c: Rough Draft − SES2d: Final Paper Participants also shared (anonymized) assignment submission	Similar to SUR1, with repeated components	**Two search tasks:** one task-topic repeated from SES1, one new; same format as SES1 **One website reliability assessment** from SHEG (topic different from SES1) **Semi-structured interview:** reflection on searching and learning experience.	Similar to SUR2 Participants self-reported scores and grades they received for different parts of the final project
Approx. Time Reqd.	10 - 15 mins	60 - 90 mins	No time limit for working on assignments. Sharing data with researchers took 1-5 minutes.	10 - 15 mins	60 - 90 mins	10 - 15 mins
Comp: (USD) $150	$5	$25	$5, $5, $10, $15 (total $35)	$10	$30	$15
	Bonus $30 paid in the end, if participant completed all parts of the study.					

Figure 5.1: Very very very very very very very very long caption.

5.4.1 SUR1: Entry Survey

5.4.2 SES1: Initial Session

5.4.3 SES2a - SES2d: Longitudinal Tracking Sessions

5.4.4 SUR2: Mid-Term Survey

5.4.5 SES3: Final Session

5.4.6 SUR3: Exit Survey

6 Data Analysis

Note about pronouns: all participants are referred to using gender-neutral they/them pronouns.

Final feedback: P022Pisa said > *It is great to be able to participate in the research this semester. Using the extension somehow brings me postive feedback and that helps me in study I303. So I wanna say thank you*
> - P022Pisa

6.1 Data Cleaning and Processing

6.2 Data Analysis Approach

see crescenzi thesis

6.3 URL Categorization

6.4 Latent Profile Analysis

- add WMC / memory span to features

- Use LIME / SHAP and counterfactual explanations to understand which components contribute to change in Profile Membership

6.5 Dwell Time Analysis

7 Results

- Also see Yung Sheng's Dissertation
- think hard about which data component has not been touched / analysed

7.1 RQ1: - search behaviours?

7.1.1 Q - query reformulation

- see Yung Sheng's Dissertation

7.1.2 L - source selection

7.1.3 I - interacting with sources

7.1.4 SHEG tasks - information evaluation

> We've confused young people's ability to operate digital devices with the sophistication they need to discern whether the information those devices yield is something that can be relied upon
>
> https://twitter.com/suzettelohmeyer/status/1617909351766757376 https://www.grid.news/story/misinformation/2023/01/23/will-information-literacy-in-schools-fix-our-misinformation-problem/

7.2 RQ2: mention here

7.3 RQ3: mention here

7.4 RQ4: mention here

8

Conclusions, Contributions, and Future Work

see Jacek's thesis

8.1 Research Summary

8.2 Summary of Results

8.3 Methodology

8.4 Contributions

8.5 Limitations

- No PDF
- N=16 to N=10
- Also check anticipated limitations section from proposal

8.6 Future Work

Appendices

A

Prior Work: Pilot Study

A.1 SES1: Initial Session

B

SUR1: Entry Survey

C

Questionnaires for Initial (SES1) and Final (SES3) Sessions

D

SUR2: Midterm Survey

E

SUR3: Exit Survey

F

Variables and Measures

G

Acknowledgements - The PhD Journey

Similar to David Maxwell's thesis.

This section will be fleshed out in more detail after the initial committee-submission on Feb 27, 2023. For now, I wish to thank the following people and organisations (in no particular order):

- Jacek Gwizdka
- Soo Young Rieh + Funding
- Committee Members
- HEB
- Finland people
- Slovenia People
- Germany People
  - Anke, Xiaofei, Michael, Hema, Himanshu, Ambika, Hardik...
- India People
- UK People
- USA People

- HCI4SouthAsia People
- ASIST People
- CHIIR People + Conferences
- UT Graduate School Funding
- SALPilot Study People
- I303 People
- DAAD
- ABB
- iSchool Doc Colleagues
- Labmates, Officemates
- LinkedIn people
- Twitter people
    - Jason Baldridge

References

Agosti, M., Fuhr, N., Toms, E., & Vakkari, P. (2014). Evaluation methodologies in information retrieval dagstuhl seminar 13441. *ACM SIGIR Forum*, *48*, 36–41.

Allan, J., Croft, B., Moffat, A., & Sanderson, M. (2012). Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in lorne. *ACM SIGIR Forum*, *46*, 2–32.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley & Sons.

Amina, T. (2017). Active knowledge making: Epistemic dimensions of e-learning. In *E-learning ecologies* (pp. 65–87). Routledge.

Bhattacharya, N. (2021). A longitudinal study to understand learning during search. *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, 363–366.

Borlund, P. (2013). Interactive Information Retrieval: An Introduction. *Journal of Information Science Theory and Practice*, *1*(3), 12–32. https://doi.org/10.1633/JISTAP.2013.1.3.2

Brookes, B. C. (1980). The foundations of information science. Part i. Philosophical aspects. *Journal of Information Science*, *2*(3-4), 125–133.

Collins-Thompson, K., Hansen, P., & Hauff, C. (2017). Search as learning (dagstuhl seminar 17092). *Dagstuhl Reports*, *7*.

Cope, B., & Kalantzis, M. (2017). *E-Learning Ecologies: Principles for New Learning and Assessment*. Taylor & Francis.

Cope, B., & Kalantzis, M. (2013). Towards a New Learning: The *Scholar* Social Knowledge Workspace, in Theory and Practice. *E-Learning and Digital Media*, *10*(4), 332–356. https://doi.org/10.2304/elea.2013.10.4.332

Eickhoff, C., Gwizdka, J., Hauff, C., & He, J. (2017). Introduction to the special issue on search as learning. *Information Retrieval Journal*, *20*(5), 399–402.

Freund, L., Gwizdka, J., Hansen, P., Kando, N., & Rieh, S. Y. (2013). From searching to learning. *Evaluation Methodologies in Information Retrieval. Dagstuhl Reports*, *13441*, 102–105.

Freund, L., He, J., Gwizdka, J., Kando, N., Hansen, P., & Rieh, S. Y. (2014). Searching as learning (SAL) workshop 2014. *Proceedings of the 5th Information Interaction in Context Symposium*, 7–7.

Gwizdka, J., Hansen, P., Hauff, C., He, J., & Kando, N. (2016). Search as learning (SAL) workshop 2016. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1249–1250.

Hansen, P., & Rieh, S. Y. (2016). Editorial: Recent advances on searching as learning: An introduction to the special issue. *Journal of Information Science*, *42*(1), 3–6. https://doi.org/10.1177/0165551515614473

Kalantzis, M., & Cope, B. (2012). *New Learning: Elements of a Science of Education*. Cambridge University Press.

Karapanos, E., Gerken, J., Kjeldskov, J., & Skov, M. B. (Eds.). (2021). *Advances in Longitudinal HCI Research*. Springer International Publishing. https://doi.org/10.1007/978-3-030-67322-2

Kelly, D. (2006a). Measuring online information seeking context, Part 1: Background and method. *Journal of the American Society for Information Science and Technology*, *57*(13), 1729–1739. https://doi.org/10.1002/asi.20483

Kelly, D. (2006b). Measuring online information seeking context, Part 2: Findings and discussion. *Journal of the American Society for Information Science and Technology*, *57*(14), 1862–1874. https://doi.org/10.1002/asi.20484

Kelly, D., Dumais, S., & Pedersen, J. O. (2009). Evaluation challenges and directions for information-seeking support systems. *IEEE Computer*, *42*(3).

Ko, A. J. (2021). Seeking information. In *Foundations of Information*. https://faculty.washington.edu/ajko/books/foundations-of-information/#/seeking

Koeman, L. (2020). *HCI/UX research: What methods do we use? – lisa koeman – blog*. https://lisakoeman.nl/blog/hci-ux-research-what-methods-do-we-use/.

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services* (Vol. 2). Libraries Unlimited Westport, CT.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press.

Marchionini, G. (2006). Toward human-computer information retrieval. *Bulletin of the American Society for Information Science and Technology*, *32*(5), 20–22.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. The National Academies Press. https://doi.org/10.17226/9853

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–92.

Novak, J. D. (2010). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations* (2nd ed). Routledge.

Rieh, S. Y. (2020). *Research area 1: Searching as learning*. https://rieh.ischool.utexas.edu/research.

Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H.-J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science*, *42*(1), 19–34. https://doi.org/10.1177/0165551515615841

Sawyer, R. K. (2005). *The Cambridge handbook of the learning sciences*. Cambridge University Press.

Vakkari, P. (2016). Searching as learning: A systematization based on literature. *Journal of Information Science*, *42*(1), 7–18. https://doi.org/10.1177/0165551515615833

Vakkari, P. (2001). Changes in search tactics and relevance judgements when preparing a research proposal a summary of the findings of a longitudinal study. *Information Retrieval*, *4*(3), 295–310.

Weber, H., Becker, D., & Hillmert, S. (2019). Information-seeking behaviour and academic success in higher education: Which search strategies matter for grade differences among university students and how does this relevance differ by field of study? *Higher Education*, *77*(4), 657–678. https://doi.org/10.1007/s10734-018-0296-4

Weber, H., Hillmert, S., & Rott, K. J. (2018). Can digital information literacy among undergraduates be improved? Evidence from an experimental study. *Teaching in Higher Education*, *23*(8), 909–926. https://doi.org/10.1080/13562517.2018.1449740

White, R. (2016a). *Interactions with search systems*. Cambridge University Press.

White, R. (2016b). Learning and use. In *Interactions with search systems* (pp. 231–

248). Cambridge University Press. https://doi.org/10.1017/CBO9781139525305.010

White, R., Dumais, S., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09*, 132. https://doi.org/10.1145/1498759.1498819

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, *55*(3), 246–258. https://doi.org/10.1002/asi.10367

Zlatkin-Troitschanskaia, O., Hartig, J., Goldhammer, F., & Krstev, J. (2021). Students' online information use and learning progress in higher education – A critical literature review. *Studies in Higher Education*, 1–26. https://doi.org/10.1080/03075079.2021.1953336